School of Molecular and Cell Biology

University of the Witwatersrand

Johannesburg

Characterisation of the replication-associated protein of *South African cassava mosaic virus* and elucidation of protein binding partners in cassava

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, in fulfilment of the requirements for the degree of Master of Science in the School of Molecular and Cell Biology

Frances Margaret Ayres

567123

Supervisor: Prof Chrissie Rey

Co-Supervisor: Dr Ikechukwu Achilonu

Johannesburg, 2018

Declaration

I, **Frances Margaret Ayres (567123)** am a student registered for the degree of MSc dissertation in the academic year 2015.

I hereby declare the following:

- • I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- • I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.
- • I have not submitted this work before for any other degree or examination at this or any other University.
- • The information used in the Thesis/Dissertation/Research Report **HAS NOT** been obtained by me while employed by, or working under the aegis of, any person or organisation other than the University.
- • I have followed the required conventions in referencing the thoughts and ideas of others.
- • I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signed  on the 31st of October 2018

Abstract

Geminiviruses are a large family of plant-infecting viruses that have severely affected a variety of economically important crops, such as cassava. Cassava is a perennial crop that plays a vital economic role in the lives of subsistence farmers of the developing world. Cassava is a salient crop in developing countries and is favoured due to its high starch content and drought tolerance. Africa has emerged as the leading producer of cassava; however, yields of cassava tuberous roots and leaves have been negatively impacted by the emergence of viral pathogens. Geminiviruses such as South African cassava mosaic virus (SACMV) introduce cassava mosaic disease which negatively impacts upon crop yields. The geminivirus-encoded replication-associated protein (Rep) is a highly conserved protein that is crucial for viral replication, and modulates a number of virus-host interactions. The aims of this study included the expression and characterisation of SACMV Rep and determining host protein binding partners of the viral protein. Expression of a soluble recombinant protein was successfully carried out. Structural analyses were undertaken by studying the secondary and tertiary structural features of the protein using circular dichroism (CD), and intrinsic fluorescence spectroscopy. Secondary structure studies were inconclusive, with intrinsic fluorescence revealing that adenosine triphosphate (ATP) binding did not induce significant protein conformational changes. Binding of 8-Anilino-1-naphthalenesulfonic acid (ANS) to SACMV Rep was observed, with the interaction not disrupted in the presence of ATP. This confirmed the lack of a persistent SACMV Rep-ATP interaction. Functional characterisation was undertaken using a DNA binding and cleavage assay. A comparison between results obtained in this study and previous geminivirus Rep research suggests cleavage of the nucleotide probe, and binding to the cleaved 5′-end. The formation of a protein-DNA complex confirmed that the expressed recombinant protein was functionally competent. Interaction of cassava proteins with SACMV Rep were probed using a yeast two-hybrid assay (YTH). Putative cassava protein binding partners of SACMV Rep were selected based on previous research and geminivirus related literature. The putative binding partners, histone H3, cyclin D3;2 and cyclin D4;2, were used as the prey proteins with bait protein SACMV Rep utilised as the bait. Results obtained from the X-gal filter lift assay suggested that there was no direct interaction of SACMV Rep with histone H3, cyclin D3;2 or cyclin D4;2. This study presents the first investigation into the elucidation of structural and functional features of SACMV Rep, and the first effort into identifying host protein binding partners of SACMV Rep.

This masters thesis is dedicated to my mother.

Diese Masterarbeit widme ich meiner Mama.

Ich hoffe du bist stolz auf mich.

Acknowledgements

I owe my sincere thanks and gratitude to Prof Chrissie Rey, who allowed me to delve into the viral protein world on her behalf. Thank you for your support and understanding over the course of this MSc. Many thanks to Dr Ikechukwu Achilonu, for his guidance and for expanding my protein knowledge.

A huge debt is owed to Alexander Zwolinski and Helen Walsh. Thank you, Alex, for pushing me to be better, for questioning everything I do, and for reigniting my passion in this project. I owe you so much. To Helen: Thank you for being a shoulder to cry on, and (often) the only sane person around. Your support throughout these few years has meant so much to me.

A big thank you to Dr Imanu Mwaba, without whose mentorship, guidance and knowledge I would not have even started this MSc. I will always be so grateful for all you have done for me.

Thank you to the Cassava Biotech Lab for all the laughs, and for keeping me busy.

Thank you to all of my friends, near and far for always asking me how I was doing, and when I would be done with the Masters.

To Mama, Dad, Steph, Jess, Jom and Luca: Although you never understood what I was doing, you never stopped encouraging me and telling me that everything was going to be alright. Your unwavering support (and of course all the jokes), have kept me going. I hope I've made you proud. I love you all.

Table of Contents

List of Figures

List of Tables

List of Equations

Abbreviations

aa	amino acids
AAV	Adeno-associated virus
ACMV	African cassava mosaic virus
AD	activation domain
ANS	8-Anilino-1-naphthalenesulfonic acid
ATP	Adenosine triphosphate
BGMV	Bean golden mosaic virus
BiFC	Bimolecular fluorescence complementation
BLAST	Basic local alignment services tool
bp	base pairs
BPV	Bovine papillomavirus
CBSD	Cassava brown streak disease
CD	Circular dichroism
CDK	Cyclin-dependant kinase
ChiLCV	Chilli leaf curl virus
CMD	Cassava mosaic disease
CP	Coat protein
CR	Common region
CRESS	Circular, Rep encoding single stranded
CYCD	Plant cyclin D family
DBD	DNA binding domain
DMSO	Dimethyl sulfoxide

ds	double-stranded
DTT	1,4-dithiothreitol (DTT),
EACMV	East African cassava mosaic virus
EDTA	Ethylenediaminetetraacetic acid
FPLC	Fast protein liquid chromatography
FRET	Fluorescence resonance energy transfer
GRACY	grand average of hydropathicity
GRIK	Geminivirus Rep interacting kinase
GRS	Geminivirus replication sequence
GST	Glutathione S-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-Tag	Oligo-Histidine Tag
HR	Hypersensitive response
HUB1	Histone monoubiquitination enzyme
IMAC	Immobilised metal ion affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRD	Iteron-related domain
ITC	Isothermal titration calorimetry
MALDI-TOF	Matrix-assisted laser desorption ionization time-of-flight
Mant-ATP	2′(3′)N-methylanthraniloyl-ATP
MBP	Maltose-binding protein
MCM	Minichromosome maintenance protein
MP	Movement protein

MS	Mass-spectrometry
MYMIV	Mungbean yellow mosaic India virus
NMR	Nuclear magnetic resonance
NSP	Nuclear shuttle protein
NTP	Nucleoside triphosphate
ORF	Open reading frame
ori	origin of replication
PAGE	Polyacrylamide gel electrophoresis
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PMSF	Phenylmethylsulfonyl fluoride
PPI	Protein-protein interaction
PSI-BLAST	Position specific iterate-basic local alignment services tool
PSSM	Position specific scoring matrix
PTGS	Post-transcriptional gene silencing
RBR	Retinoblastoma-related protein
RCR	Rolling circle replication
RDR	Recombinant dependant replication
Ren	Replication enhancer
Rep	Replication-associated protein
RFC	Replication factor C
RPA	Replication protein A
SACMV	South African cassava mosaic virus

SCE	SUMO conjugating enzyme
SDS	Sodium dodecyl sulfate
SF3	Super family 3
SNF	Sucrose non fermenting
SNRK	Geminivirus Rep interacting motor protein
SOB	Super optimum broth
SOC	Super optimum broth catabolite repression media
ss	single stranded
SV40 L-Tag	Simian virus 40 Large T-antigen
TF	Trigger factor
TGMV	Tomato golden mosaic virus
TGS	Transcriptional gene silencing
ToLCGuV	Tomato leaf curl Gujarat virus
ToLCV-Nde	Tomato leaf curl New Delhi virus
TrAP	Transcription activating protein
Tris	Tris(hydroxymethyl)aminomethane
TYLCSV	Tomato yellow leaf curl Sardinia virus
TYLCV	Tomato yellow leaf curl virus
UBC2	Ubiquitin conjugating enzyme
WDV	Wheat dwarf virus
X-gal	5-bromo-4-chloro-3-indolyl-β-d-galactopyranoside
YTH	Yeast two-hybrid

The IUPAC-IUBMB one- and three-letter abbreviations for the 20 standard amino acids were used.

Chapter I – Literature review

1.1 Cassava the crop and the threat to its production

Global food security has become one of the most important matters to be addressed by humanity in this century. Crop production and yield are under threat due to a changing global climate, plant pathogens and pests. These are some of the issues putting further strain on agricultural activity (Sundström et al., 2014). Approximately 30-40% of crops produced are lost on a yearly basis, emphasising the importance of improving plant health (Flood, 2010). Plant viruses are one of the causative agents of crop yield loss. Viral pathogens are difficult to remove and/or control due to their short generation time. Under natural selection pressure the high rate of genetic recombination and mutation enables the development of advantageous traits for the virus. These advantageous traits can be quickly distributed in the subsequent viral generations. To mitigate the damage caused by plant viruses, plant disease management strategies may be implemented, such as control of the insect vector, hygiene strategies and genetic engineering. Despite these various strategies, viruses evolve and as a result new viral species constantly emerge (Hanssen et al., 2010).

Cassava (*Manihot esculenta* Crantz), a perennial shrub, was first introduced from South America to the African continent in the 16th century. The crop subsequently spread throughout central and southern Africa in the 20th century (Carter et al., 1997). Its ability to withstand varying environmental conditions, such as drought, and its tolerance to acidic soils has promoted its cultivation in the developing regions of Africa and Asia (Legg et al., 2015). The cassava root can remain in the ground for long periods of time, and plant growth is not inhibited by high carbon dioxide concentrations. These characteristics support the classification of cassava as a food security crop, and an essential crop in times of climate change (Legg et al., 2014). On a global scale, cassava is rated as the 4th most important source of energy, sustaining more than 700 million people (Legg et al., 2015). In 2012, approximately 240 million tonnes of cassava root tubers were harvested, a yield that has increased significantly over the past 50 years (Legg et al., 2014). Of this yield, 70% is used for human consumption, the remaining 30% is utilized in the textile industry, as well as in paper production and animal feed (El-Sharkawy, 2004).

When comparing production yield of the cassava growing areas across the world, the highest production is found in Africa, where in excess of 50% is generated (Legg et al., 2015). In

addition to being an important source of starch, the cassava crop is a crucial form of income for subsistence farmers, particularly for women living in sub-Saharan Africa. The continent displays cassava production which outweighs those of maize and yams (Figure 1.1); however overall yield is largely impacted by pests and diseases.

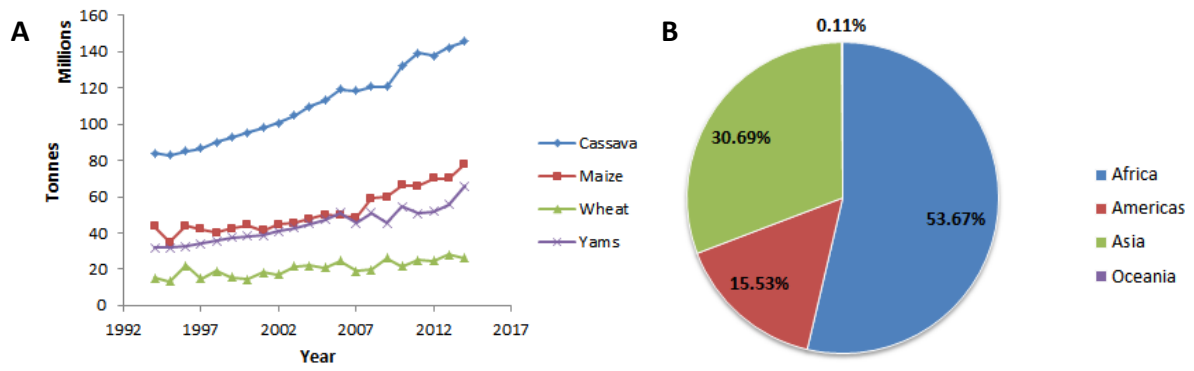

Figure 1.1 Cassava production on a global scale. A Graph showing production of crops in million metric tonnes in Africa. Cassava is the most produced crop across the continent, with the production of maize and yams being significantly less. The graph depicts the growth in cassava production over a period of 25 years. **B** Africa is the biggest producer, with a 54% production yield, ahead of Asia and the Americas (FAOSTAT, 2014).

Cassava mosaic disease (CMD) wreaks havoc on cassava yields in affected countries, as the disease can result in the more susceptible varieties of cassava displaying 100% loss of the crop (Fondong, 2017). Over the years from 2010 to 2014, an almost 20% reduction in cassava crop yield has been observed (Figure 1.2). As production of cassava in Africa increases, so does the threat of the two most important viral diseases of cassava; namely, Cassava Brown streak disease (CBSD) and CMD which both severely impact the yield of the crop (Figure 1.2). Due to the increasing economic impact of CMD and CBSD, they have been labelled as the most destructive to global cassava production (Legg et al., 2014).

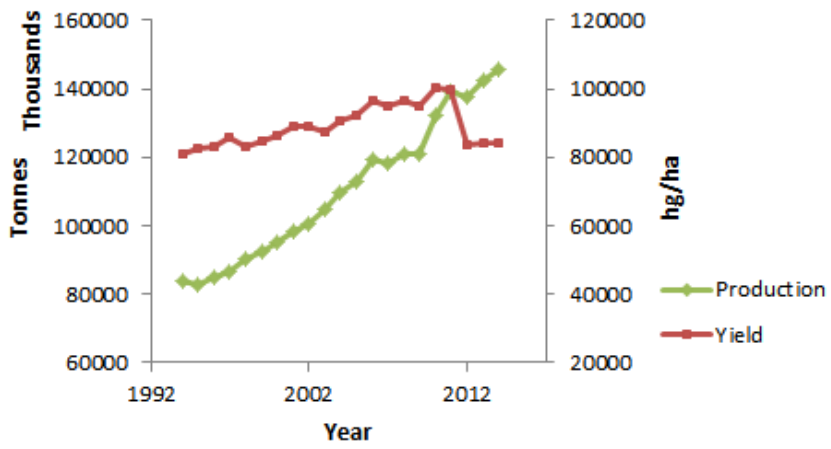

Figure 1.2 Production and yield of cassava in Africa.Despite a rise in production of this popular crop, the yield dropped significantly in 2012, and yield remains at pre-2000 levels (FAOSTAT, 2014).

1.2 Cassava mosaic disease and the causative agents belonging to the family *Geminiviridae*

CMD is caused by viruses of the family *Geminiviridae* which has nine genera assigned to it (Rey and Vanderschuren, 2017; Zerbini et al., 2017). This family of plant viruses infect many economically important crops such as tomato, tobacco, potato, beans and cassava (Czosnek et al., 2001; Hanley-Bowdoin et al., 2013). Geminiviruses are characterised by two incomplete icosahedra, attached at a waist segment, laying claim to the name "geminivirus". Each twinned particle encompasses a single circular single stranded DNA (ssDNA) molecule (Jeske, 2009). Disseminated by the whitefly insect vector *Bemisia tabaci*, begomoviruses form the largest genus of the geminivirus family and the only genus that includes bipartite viruses (Jeske, 2009). Other genera, such as mastre- and curtoviruses exhibit monopartite genomes only (Zerbini et al., 2017). The genome of bipartite begomovirus is made up of a DNA-A and a DNA-B component, which are both required for infection of the host (Jeske, 2009). DNA-A encodes six known proteins, and DNA-B encodes for two known proteins (Fondong, 2013). The DNA-A of the bipartite genome displays four open reading frames (ORFs) on the complementary strand, and two situated on the virion-sense strand. Two ORFs are present on DNA-B, one in complementary-sense and one in virion-sense (Figure 1.3) (Fondong, 2013). Contrastingly monopartite begomoviruses contain two virion-sense genes, and four complementary-sense genes, all encoded by DNA-A only (Fondong, 2013)

Cassava infecting begomoviruses are bipartite geminiviruses. Symptoms of cassava begomoviruses include leaf mosaic, stunting of the plant and reduced tuber size (Legg and Fauquet, 2004). Cassava begomoviruses endemic to Africa include African cassava mosaic virus (ACMV), East African cassava mosaic virus (EACMV) and South African cassava mosaic virus (SACMV) (Berry and Rey, 2001). SACMV arose through recombination events between EACMV, Tomato yellow leaf curl virus (TYLCV), and an unknown indigenous geminivirus (Berrie et al., 2001). DNA–A of SACMV (GenBank accession number: AF155806) (2800 base pairs (bp)) encodes six proteins from ORFs on the complementary strand, and two proteins from the viral-sense strand. DNA-B (AF155807), 2760 bp in length contains two ORFs, one on each strand (Figure 1.3).

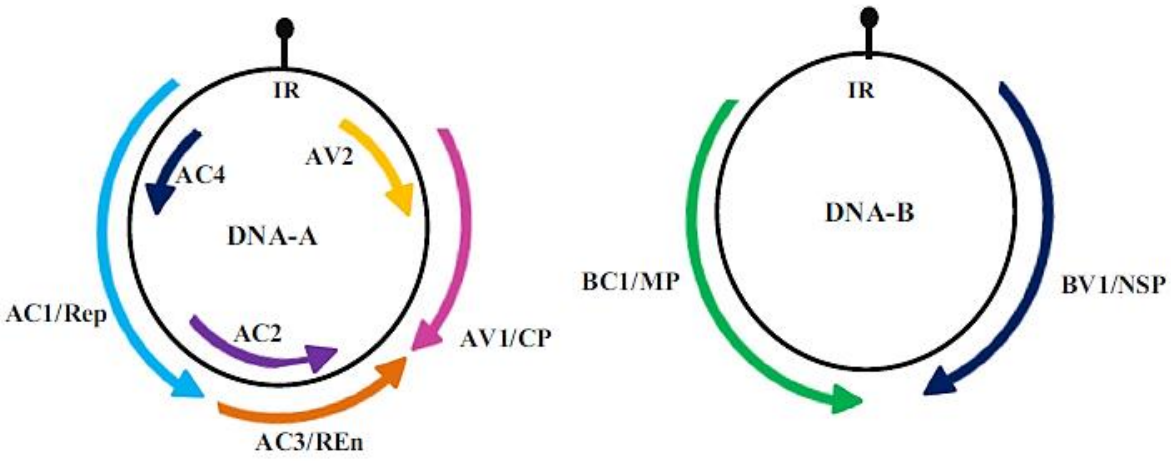

Figure 1.3 Schematic representation of bipartite begomovirus genome. Begomoviruse DNA A and DNA B genomic components shown as circular structures. The positions of the eight ORF are indicated on the viral genome in various colours with the gene/protein name also denoted (Rizvi et al., 2015).

DNA-A and DNA-B share a 200 bp common region (CR) which, in turn, contains an intergenic region (IR) (Legg and Fauquet, 2004) (Figure 1.3). The IR contains inverted repeats, which facilitate the formation of a stem-loop structure. Within the loop of the stem-loop structure a nonanucleotide sequence (TAATATTAC) has been identified. Due to the high level of conservation of this sequence amongst geminiviruses, it was established that the nonanucleotide contains the origin of replication (*ori)* (Fontes et al., 1992; Laufs et al., 1995a). Replication of geminiviruses occurs via rolling circle replication (RCR) and recombinant dependant replication (RDR). The *ori* and the presence of the replication-associated protein (AC1/Rep) are required for the initiation of RCR and RDR (Laufs et al., 1995a).

1.3 Open reading frames and translated proteins of bipartite begomoviruses

Rep is highly conserved in both sequence and function amongst the genus *Begomovirus*. It is the only protein product of the geminiviruses that is crucial for replication (Hanley-Bowdoin et al., 2004; Nash et al., 2011). The Rep ORF is located on the complementary strand of DNA-A and is expressed in the early stages of infection. The initiation and termination of RCR and RDR is dependent on the activity of Rep, as it acts as both an endonuclease and ligase (Laufs et al., 1995a). Rep facilitates the reprogramming of mature plant cells to re-enter the S-phase by altering the plant host cell cycle, thereby enabling the replication of the viral genome. Geminivirus Rep has also been shown to act as a helicase and shares motifs and structural characteristics with oncoproteins of animal ssDNA viruses (Campos-Olivas et al., 2002; George et al., 2014). Geminivirus Rep proteins of the nine genera range in size from 350-360 amino acids (aa) depending on the genus and species of geminivirus.

Other proteins encoded for on the complementary strand of DNA-A, include the transcription activator protein (TrAP/AC2), which serves to promote viral symptom progression and enables the transcription of the movement and coat protein encoded for on DNA-B. TrAP has also been identified as an RNA silencing suppressor (Fondong, 2013; Hanley-Bowdoin et al., 2013). TrAP works alongside the replication enhancer protein (REn/AC3) which aids in viral replication and promotes symptom development (Sunter et al., 1990). Its presence is not crucial for viral replication; however, it increases viral DNA accumulation (Sunter et al., 1990). The last gene on the complimentary strand is AC4. The ORF of AC4 is contained within AC1 but in a different frame, and is the least conserved of the geminivirus genes. For the *Mastrevirus* genus this gene is not present (Fondong, 2013). The function of AC4 differs between mono- and bipartite viruses, but has been shown to support viral movement and development of symptoms (Fondong, 2013). Similar to TrAP, AC4 was identified as a suppressor of plant RNA silencing, thereby increasing viral pathogenicity (Vanitharani et al., 2004).

On the sense (viral) strand, AV1 encodes for the coat protein (CP), the structural protein of geminiviruses that ensures systemic movement of the virus as well as cell to cell transport. It has been suggested that CP is a late product of viral infection cycle and participates in viral replication through the formation of a complex with ss- and double-stranded DNA (dsDNA). Binding of CP to newly synthesized ssDNA may intervene in viral replication (Liu et al., 1997; Saunders et al., 1991; Stenger et al., 1991). In addition, it ensures stable transport of the

virus between plant host and insect vector, as well as confirming vector specificity (Fondong, 2013; Jeske, 2009). The other protein encoded for on the viral strand, AV2 has not been assigned a function; however, this protein has been shown to act as a suppressor of RNA silencing (Chowda-Reddy et al., 2008). AV2 has only been identified in "Old World" bipartite begomoviruses (Fondong, 2013).

The ORFs on DNA-B encode for the nuclear shuttle protein (NSP/BV1) and the movement protein (MP/BC1). The coding sequences for NSP and MP are situated on the viral strand and complementary strand, respectively. NSP enables the export of ssDNA out of the nucleus into the cytoplasm by forming a complex with the viral DNA. For cell to cell movement, the NSP-ssDNA complex is bound by MP, and enables the transfer of viral DNA to and transport across the cell wall into neighbouring cells. Subsequent movement of the viral DNA into the nucleus is enabled by the NSP. The NSP has been shown to not be crucial for virus infectivity as the CP has demonstrated the ability to complete the functions of viral DNA nuclear import and export in the absence of the NSP (Fondong, 2013).

1.4 Domains, motifs and the evolutionary connection of Rep

1.4.1 Common ancestors of geminivirus Rep

Despite the crucial role that Rep plays during viral replication, it shares no sequence or functional homology with DNA polymerases. However, amino acid motifs and functional similarities have been identified between the replication protein of the different genera of geminiviruses, despite differences in host range and vector (Gutierrez, 1999; Hanley-Bowdoin et al., 1999).

Selected similarities exhibited amongst Rep proteins of geminiviruses are shared with replication initiator proteins of eubacterial plasmids as well as other ssDNA viruses, most notably the replication proteins of animal infecting viruses Simian virus 40 Large T-antigen (SV40 L-Tag) (Family *Polyomaviridae*), Adeno-associated virus (AAV) Rep (Family *Parvoviridae*), E1 from papillomavirus (Family *Papillomaviridae*) and Rep of the animal and plant infecting family *Circoviridae* (George et al., 2014; Jeske, 2009; Koonin and Ilyina, 1992). The mechanism of initiation and termination of RCR as well as helicase activity of the replication proteins across the viral families reveals their common ancestry. Furthermore, the RCR mechanism of these viral proteins is shared with plasmid replication in bacteria, algae

and archaea. It has been suggested that Rep is the only protein that has been present throughout the evolution of algae, bacteria, archaea as well as geminiviruses. This idea is supported by the discovery that Wheat dwarf virus (WDV) (*Mastreviridae*) and a Rep-like sequence of the red algae (*Porphyra pulchera*) share over 30% homology (Nawaz-ul-Rehman and Fauquet, 2009).

Three domains have been identified for Rep of the *Geminiviridae* family: DNA binding domain (also known as the catalytic domain and/or endonuclease domain) (Campos-Olivas et al., 2002; Orozco et al., 1997), the oligomerisation domain (Orozco et al., 1997, 2000) and the ATPase domain, also known as the helicase domain (Desbiez et al., 1995). It is interesting to note that the conserved nature of the endonuclease and helicase domain of Rep has been shown across a number of eukaryote infecting ssDNA virus families including *Geminiviridae*, *Circoviridae*, *Parvoviridae* and *Nanoviridae*, as well as bacteria and archaea infecting viruses, including *Microviridae* and *Spiraviridae* amongst others (Kazlauskas et al., 2018). The conserved nature of the replication proteins of eukaryotic circular, ssDNA viruses has even led to an unofficial grouping of these viruses, labelled the circular, Rep encoding single stranded (CRESS) viruses (Rosario et al., 2012).

1.4.2 Catalytic domain

The catalytic domain of Rep is situated on the N-terminus and is mapped approximately to the first 130 aa (Figure 1.4). This domain contains motifs that are fundamental for the initiation and termination of replication. The catalytic domain contains four conserved motifs, all indispensable for the primary function of Rep (Nash et al., 2011; Orozco and Hanley-Bowdoin, 1998; Orozco et al., 1997).

Motif I displays a conserved consensus aa sequence of FTLY and promotes DNA binding, thereby initiating the replication process (Orozco and Hanley-Bowdoin, 1998). To enable protein DNA binding, a metal coordination site in the form of motif II is present. The histidine residues within the conserved aa sequence of HUHUUU (U= large, hydrophobic residue) coordinate the Mg^{2+} or Mn^{2+} ions necessary for successful replication (Laufs et al., 1995a; Orozco and Hanley-Bowdoin, 1998). The third motif, with a consensus aa sequence of YXXKD/E/N (X=any residue) is essential for cleavage of viral DNA. The tyrosine residue acts as the catalytic residue, and carries out a nucleophillic attack on the phosphodiester bond of DNA. This chemical reaction has been shown to be the catalytic step of viral DNA

replication as cleavage of the viral DNA enables the initiation of RCR (Laufs et al., 1995b). The geminivirus replication sequence (GRS) is the fourth motif within the catalytic domain known to be indispensible for replication (Figure 1.4). Experiments analysing GRS mutants have shown that a Rep protein lacking this sequence cannot cleave the viral strand, thereby failing to carry out replication (Nash et al., 2011). This motif is highly conserved amongst geminiviruses; however, mutational studies have shown that the positioning of GRS is not conserved. Mutations to GRS do not affect DNA binding, but rather prohibit DNA cleavage activities. Due to this characteristic, GRS cannot be assigned one function and selected residues have been shown to be integral to the native, functional structure of Rep. It must be noted that these integral residues do not influence functionality of Rep, but rather ensure the structural integrity of the protein (Nash et al., 2011). Collectively, these four motifs carry out the initial and fundamental steps that are required for viral replication and creation of progeny virions.

1.4.3 Oligomerisation domain

The oligomerisation domain is found approximately between amino acid residues 121-181, overlapping with the catalytic domain, and the ATPase domain (Figure 1.4). It has been suggested that these surrounding domains aid in the oligomerisation of Rep (Orozco et al., 1997) with oligomerisation of Rep being a vital step in the replication process. Rep acts as a helicase only through its ability to form a homo-hexamer, with ATPase activity also being dependant on oligomerisation (Choudhury et al., 2006; Clérot and Bernardi, 2006). Rep has been shown to form higher order oligomers, such as hexamers, dodecamers, and even 24-mers, as seen for Tomato leaf curl Gujarat virus (ToLCGuV), Tomato yellow leaf curl Sardinia virus (TYLCSV) and Mungbean yellow mosaic India virus (MYMIV), respectively (Choudhury et al., 2006; Clérot and Bernardi, 2006; George et al., 2014). Binding to the plant-related retinoblastoma related protein (RBR), situated around residues 132-156 (as shown for Tomato golden mosaic virus (TGMV)), is also only possible if Rep forms homo-oligomers. Within the oligomerization domain two additional host protein binding sites have been identified. The REn and proliferating cell nuclear antigen (PCNA) binding sites have been mapped to the central domain, more specifically to residues 134-181 on MYMIV Rep protein (Bagewadi et al., 2004; Settlage et al., 2001) (Figure 1.4).

Alongside geminiviruses' Rep proteins, the replication proteins of the virus families *Polyomaviridae* (e.g. SV40 L-Tag), *Papillomaviridae* (e.g. Bovine papillomavirus (BPV) E1), and *Parvoviridae* (e.g. AAV E1A) were identified as helicases of superfamily 3 (SF3) (Gorbalenya and Koonin, 1993). Oligomerization of the SF3 helicase is crucial for a functioning enzyme, and when found in its monomeric form, the enzyme is inactive (Clérot and Bernardi, 2006). In order for oligomerization to take place, SF3 helicases require a neighbouring oligomerization domain for proper functioning (Clérot and Bernardi, 2006). Like other helicases of SF3, geminivirus Rep requires the 3′ hydroxyl group on ssDNA to unwind DNA in the 3′-to-5′ direction. Geminivirus Rep does not share all characteristics of SF3 helicases. A difference between Rep and other SF3 helicases, is that SF3 helicases predominantly form hexamers (Clérot and Bernardi, 2006). Geminivirus replication proteins, however, have been shown to form high order oligomers, including oligomers of 24 subunits (Choudhury et al., 2006; George et al., 2014).

1.4.4 NTP-binding domain

Members of the SF3 helicase superfamily were assigned to the large family of AAA+ATPases, a family of enzymes that encompass a P-loop nucleoside triphosphate (NTP) domain and are known to function as helicases (Iyer et al., 2004). Geminivirus Rep are in possession of the NTPase domain and display ATPase activity which enables helicase activity (Clérot and Bernardi, 2006). This ATPase activity has been shown to improve endonuclease activity of Rep; however, the cleavage and ligation activities of Rep are not dependent on the ATPase activity (Clérot and Bernardi, 2006; Desbiez et al., 1995; Pant et al., 2001).

The C-terminal domain of Rep is labelled the NTP-binding domain (Hanley-Bowdoin et al., 1999). The presence of an NTP-binding domain was first established in geminiviruses when high levels of sequence similarity were discovered between the NTP-binding domains of positive strand RNA viruses and small DNA-viruses (Gorbalenya et al., 1990). The motifs identified within the NTP binding domain in the study completed by Gorbalenya et al. (1990) included the Walker A and Walker B motifs as well as motif C (Gorbalenya et al., 1990) (Figure 1.4). Walker A is denoted by the conserved sequence GXXXGK(T/S) (X=any residue), and binds to adenosine triphosphate (ATP) (Clérot and Bernardi, 2006). Walker B is involved in ATP hydrolysis and displays a conserved sequence of DXXD or XXXDD. Motif

C consists of a conserved asparagine residue that binds to and interacts with the gamma phosphate of ATP and the so-called “apical” water molecule (George et al., 2014). The more recently identified B′ motif is located between Walker B and motif C and is denoted by an amino acid motif of (K/R)X3-4GX7-8K (Figure 1.4). This motif is predicted to be involved in ATP and DNA binding. ATP binding to Rep and hydrolysis thereof induces a conformational change in the protein and allows for DNA binding to occur (George et al., 2014). The integrity of the NTP-binding loop is essential for ATPase activity, as was first shown with TYLCV (Desbiez et al., 1995).

The AAA+ATPase domain of geminivirus Rep is smaller in length than ATPase domains of other SF3 helicases (Clérot and Bernardi, 2006). As shown for ToLCGuV, the SF3 conventional arginine finger motif is absent and has not been identified in Rep proteins of *Geminiviridae* (Choudhury et al., 2006).

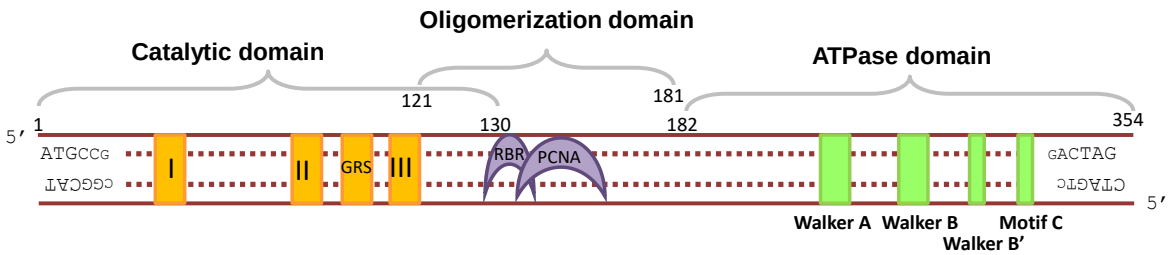

Figure 1.4 Schematic representation of the domains of Rep. The three domains are displayed, as well as the identified motifs and host protein binding sites based on functional and structural studies of geminivirus Rep (adapted from Rizvi et al., (2015)).

1.5 The role of Rep in the geminivirus life cycle

The geminiviral lifecycle begins when the whitefly vector transfers the virus into the phloem, where the virus then travels through to the phloem to companion and phloem-parenchyma cells (Jeske, 2009). Within the cytoplasm the ssDNA is released from the capsid, and is converted to dsDNA in the nucleus. This process is supported by the availability of plant host DNA polymerases and the synthesis of RNA primers complementary to the intergenic region (Saunders et al., 1992). For phloem-limited geminiviruses, this process is restricted to nucleus containing cells such as companion and phloem parenchyma cells (Jeske, 2009). Minichromosomes are formed by the binding of dsDNA to histones, with chromatin modifications affecting downstream gene regulation activities (Pilartz and Jeske, 2003). The plant host RNA polymerase transcribes viral genes from the double-stranded intermediate, producing Rep as well as the other viral protein products. This is performed in a bidirectional

manner, as bidirectional promoters are present within the intergenic region. Rep and viral proteins encoded for on the complementary strand are transcribed first, followed by the ORFs situated on the viral strand (Hanley-Bowdoin et al., 1999).

Once translated, Rep binds to the Rep binding site located within the CR where the transcription start site and TATA box have been mapped and cognate sequences determine the specificity of Rep binding to the viral genome (Fontes et al., 1992, 1994a) (Figure 1.5). The cognate sequence is a directly repeated sequence found upstream of the conserved stem-loop and nonanucleotide sequence, and it is specific for Rep of each virus species (Arguello-Astorga et al., 1994; Fontes et al., 1992, 1994a).

Binding specificity is further determined by iterons contained within the viral genome. The iterons are 8-12 nucleotide long repeats within the CR that differ between different virus species, but are common throughout geminiviruses that infect dicotyledons (Argüello-Astorga et al., 1994). In addition to the nonanucleotide sequence, iteron related domains (IRDs) further influence binding specificity of Rep to the CR. IRD are contained within the amino acid sequence and further control the sequence specific binding specificity of Rep to the viral genome (Arguello-Astorga and Ruiz-Medrano, 2001). The Rep binding site has also been mapped to the stem-loop and 60 bp upstream of the loop (Fontes et al., 1994a). An iteron downstream of the stem-loop cleavage site has been reported, an essential component for replication of MYMIV (Arguello-Astorga et al., 1994; Singh et al., 2008).

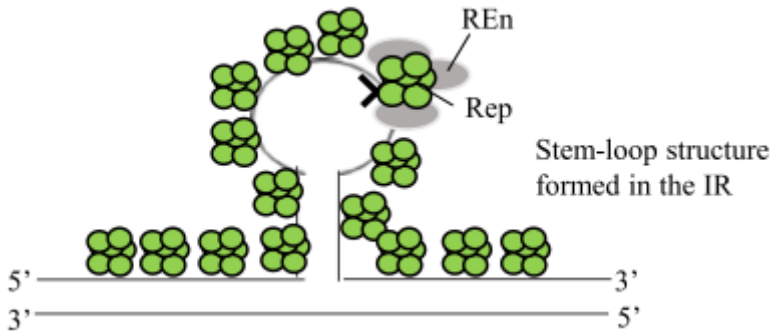

Figure 1.5 A schematic representation showing Rep binding to the viral origin. Rep binds to the cognate sequence in a cooperative manner. Oligomeric Rep binds to and cleaves within the nonanucleotide on the viral strand, in turn initiating replication, with REn involved in the replication process (Rizvi et al., 2015).

Viral replication initiation requires the recruitment of host replication associated proteins by Rep, including the replication protein A (RPA) (Singh et al., 2007), PCNA (Bagewadi et al.,

2004; Castillo et al., 2003) and Replication Factor C (RFC) (Luque et al., 2002) (Table 1.1). MYMIV Rep was shown to bind to the large subunit of RPA, which induced the down-regulation of nicking and ligating activity. Complex formation resulted in increased ATPase activity, thereby increasing viral DNA replication (Singh et al., 2007).

The binding of TYLCSV Rep, REn and plant host PCNA has been demonstrated, and has been shown to aid in geminiviral replication. This interaction may allow for the proteins involved in the replisome to be situated close to the viral origin. The inclusion of REn in this binding complex ensures an increase in the resultant viral DNA (Castillo et al., 2003). PCNA also acts as the processivity factor for the host polymerase δ and is known to bind to a number of other host proteins involved in replication of DNA (Hanley-Bowdoin et al., 2013). The large subunit of RFC acts as the PCNA loading clamp and binding of WDV Rep to RFC suggests that geminiviral Rep is involved in the beginning stages of the elongation complex (Luque et al., 2002).

The binding of Rep to the *ori* on the viral genome enables the formation of the replisome. Following this, Rep carries out endonuclease activity by cleaving within the nonanucleotide sequence (TAATATTAC), situated on the stem-loop (Laufs et al., 1995a; Stanley, 1995) (Figure 1.5). Rep cleaves between the thymine and adenine, the adenine being the penultimate nucleotide of the conserved sequence. The 3′ end of the cleaved strand is primed for the synthesis of the viral strand, and binds the host DNA polymerase. Using the complementary strand as the template, the DNA polymerase extends the 3′- end of the viral strand (Laufs et al., 1995a). The resulting 5′ end of the cleaved viral strand binds to Rep and is displaced during replication due to the helicase activity of Rep (Clérot and Bernardi, 2006). Helicase activity is facilitated by Rep binding to ATP, and resultant ATPase activity (Clérot and Bernardi, 2006).

The binding of MYMIV Rep to minichromosome maintenance 2 (MCM2) has also been shown, suggesting that this interaction aids in the unwinding of the replication fork as well as the assembly, stabilisation and progression of the replication fork (Bagewadi et al., 2004; Suyal et al., 2013) (Table 1.1). After one round of replication, Rep binds to and cleaves the newly generated *ori* on the viral strand to release the displaced strand. The displaced strand is circularised by the ligation activity of Rep, whereby the released 3′ hydroxyl group and 5′ end of ssDNA are ligated to yield a new ssDNA viral genome (Laufs et al., 1995a). This single stranded genome can be packaged into viral coat protein to form new virions. Alternatively

the ssDNA can be replicated to form dsDNA, and re-enter the replication cycle (Pooggin, 2013).

Besides its role as initiator of viral replication, Rep has been implicated as a regulator of replication and transcription. Histone H3 is a core histone required for establishment of the viral genome minichromosome, and also in the regulation of viral genome replication and transcription. The interaction of histone H3 with TGMV Rep has been confirmed (Table 1.1). During viral replication, as Rep is bound to the 5′ intergenic region, the Rep-H3 complex has been suggested to play a role in the displacement of the nucleosome, ensuring the transcription of host replication related proteins (Kong and Hanley-Bowdoin, 2002). Due to the involvement of histone H3 in replication and transcription, this protein acts as a gene regulator (Ascencio-Ibáñez et al., 2008). By binding to RFC and PCNA, Rep negatively regulates its transcription; however, this has been suggested to only occur during the later stages of the virus lifecycle. The formation of this regulatory complex prevents Rep binding to the viral origin and initiating viral replication (Bagewadi et al., 2004).

Table 1.1 Plant host proteins that interact with geminivirus Rep. The species of geminivirus Rep is indicated, as well as the reference.

Plant host proteins	Geminivirus species	Reference
RPA	MYMIV	Singh et al. (2007)
RFC	WDV	Luque et al. (2002)
PCNA	TYLCSV; MYMIV	Castillo et al. (2003); Bagewadi et al. (2004)
Histone H3	TGMV; CaLCuV	Kong et al. (2002)
RBR	TGMV	Ach et al. (1997); Kong et al. (2000)
GRIMP	TGMV; CaLCuV	Kong et al. (2002)
GRIK	TGMV; CaLCuV	Kong et al. (2002)
NbUBC2	ChiLCV	Kushwaha et al. (2017)
MCM2	MYMIV	Suyal et al. (2013)
SEC1	TGMV	Sanchez-Duran et al. (2011)

ChiLCV- Chilli leaf curl virus; CaLCuV- Cabbage leaf curl virus.

The question of how the newly synthesised viral ssDNA is packaged into the coat protein for transport through the cell, and for distribution to the insect vector, remains unanswered. Binding of Rep and CP, as shown for MYMIV, may explain how post-RCR packaging of viral ssDNA is carried out (Malik et al., 2005).

1.6 Function of Rep in cell cycle manipulation, cell-signalling, silencing defence pathways and transcriptional regulation

Rep is a multifunctional protein and has been shown to bind to host proteins during infection, or indirectly interfere with host molecular functions (Reyes et al., 2013). Geminiviruses are known to replicate in differentiated cells where the plant host replication machinery would not be present (Egelkrout et al., 2001; Nagar et al., 1995, 2002). In order to create a cellular environment that promotes replication in the early stages of viral infection, Rep ensures the availability of the proteins that make up the replisome by manipulating the plant cell cycle (Castillo et al., 2003; Hanley-Bowdoin et al., 1999). In normal dividing cells, expression of replisome proteins is induced by the phosphorylation of RBR, which is bound to the E2F transcription factor (Gutzat et al., 2012). The transcription factor, when not in a complex with RBR, induces the expression of replication machinery components, such as PCNA. During geminivirus infection the interaction of Rep with RBR occurs in lieu of the phosphorylation event, enabling E2F to facilitate the expression of host cell proteins. The expression of E2F coupled genes drives the infected host cell into the S-phase of the plant cell cycle (Egelkrout et al., 2001; Nagar et al., 1995, 2002).

Rep has been shown to down-regulate the expression *methyltransferase 1* and c*hromomethylase 3* in the host plant *Nicotiana benthamiana*, thereby proving its involvement in the transcriptional gene silencing (TGS) pathway (Rodríguez-Negrete et al., 2013). AC4, a virus suppressor of the post transcriptional gene silencing (PTGS) pathway, enhances down-regulation of TGS related genes. Despite this, in order for Rep to carry out its transcriptional silencing functions, the presence of AC4 is not paramount, and Rep can act independently (Rodríguez-Negrete et al., 2013).

An additional binding partner includes the geminivirus Rep-interacting kinase 1 (GRIK1) and GRIK2, which are assumed to be involved in plant growth pathways, as well as in virus infection progression (Kong and Hanley-Bowdoin, 2002; Shen et al., 2009). Expression experiments have shown that expression of GRIK 1 and 2 are up-regulated in plants infected with geminiviruses (Ascencio-Ibáñez et al., 2008; Kong and Hanley-Bowdoin, 2002). Furthermore, both proteins have been localised to the nucleus, allowing for an interaction with Rep to take place (Kong and Hanley-Bowdoin, 2002). Sucrose non-fermenting 1 (SNF1)-related protein kinase (SNRK1) regulates the activities of GRIK1 and 2, and is predicted to play a role in plant metabolism and pathways linked to biotic and abiotic stress

responses. Signalling cascades involving SNF1 and GRIK may influence viral replication by supplying an energy and nutrient source for efficient replication (Hanley-Bowdoin et al., 2013). The result of Rep-SNF1/GRIK complexes has not been elucidated as yet. However, it has been suggested that Rep binds to GRIK1/2 to shield it from degradation. This would allow kinases to recruit further host proteins but exact functions have not been clarified as yet (Kong and Hanley-Bowdoin, 2002).

Recent research has looked into how Rep counters the hypersensitive response (HR) that AC2/C2 has been shown to elicit. In *Agrobacterium*-mediated transient expression of TYLCSV C2, the hypersensitive response was elicited though Rep was co-expressed, HR was circumvented, and a chlorotic reaction was stimulated, suggesting the role of Rep in counteracting the HR response (Matic et al., 2016).

In *N. benthamiana* interactions between ChiLCV Rep and ubiquitin conjugating enzyme (NbUBC2) was confirmed both in the cytoplasm and nucleus. Furthermore, using fluorescence resonance energy transfer (FRET) microscopy and yeast two-hybrid assays binding of Rep to histone monoubiquitination enzyme (NbHUB1) was verified (Kushwaha et al., 2017). Despite the interactions of Rep with the monoubiquitination machinery of the host plant, the protein itself is not ubiquitinated. Previous studies have shown that geminiviral infection results in a transcriptionally active and non-active form of minichromosome. The transcriptionally active form of minichromosome is characterised by the presence of active markers on histone H3 and histone 2B, methylation and ubiquitination, respectively (H3-K4me3 and H2B-ub) (Ceniceros-Ojeda et al., 2016). Research conducted by Kushwaha et al. (2017) acquired similar results, as they discovered two different viral minichromsomes present in the host plant cell. The transcriptionally active form was enriched in the active markers, H3-K4me3 and H2B-ub, whereas the transcriptionally repressed form was characterised by methylation of histone H3 (H3-K27me3). The later form of minichromsome was hypothesised to be utilised for processes not related to transcription (Ceniceros-Ojeda et al., 2016). The research group hypothesised that Rep binds to UBC2 in the cytoplasm and transports it into the nucleus, and once there, it forms a monoubiquitination complex with HUB1. Rep binds to the nucleosomes of the viral genome, stimulating a complex between the monoubiquitination machinery and the viral nucleosome. This interaction promotes ubiquitination of the histone H2B, which consequently induces trimethylation of H3-K4. It is thought that the modification to the viral nucleosome brings about the bidirectional transcription of the viral genome (Kushwaha et al., 2017). Research on the E1A protein of

human adenovirus confirms this theory of geminiviral Rep activity. E1A has been shown to stimulate viral replication as well as early transcription of viral genome (Fonseca et al., 2013). The newly discovered role of geminivirus Rep in transcription activation is another example of the diverse functions that Rep fulfils in order to promote viral proliferation within the host plant (Kushwaha et al., 2017).

A study conducted by Castillo et al., (2004) first revealed the interaction of TGMV Rep with the *N. benthamiana* SUMO-conjugating enzyme (SCE1) (Castillo et al., 2004). SCE1 directs the post-translational modification of target proteins through the addition of an ubiquitin-like polypeptide labelled SUMO. By interacting with SCE1, Rep can regulate the addition of SUMO to a number of plant host proteins, thereby further regulating the host environment and ensuring its suitability for viral replication (Sánchez-Durán et al., 2011).

1.7 Structural features of Rep and their functions

The nuclear magnetic resonance (NMR) structure of the catalytic domain of TYLCSV has been solved and remains the only structure of geminivirus Rep to date (Campos-Olivas et al., 2002) (Figure 1.6). Multiple models of the C-terminal end of geminivirus Rep have been generated using structural data from other oncoproteins (George et al., 2014). Furthermore, secondary and tertiary structural features of many Rep proteins of geminiviruses have been predicted using bioinformatics tools, but the N-terminus of TYLCSV remains the solitary model based on obtained experimental data.

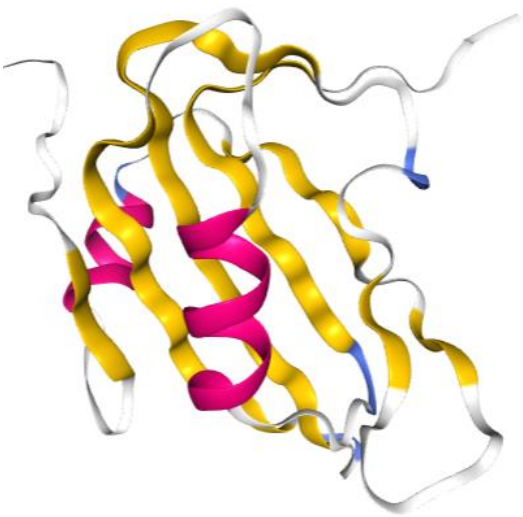

Figure 1.6 Three-dimensional structure of TYLCSV Rep$_{4\text{-}121}$ (PDB ID 1L2M). Ribbon representation for the catalytic domain of tomato yellow leaf curl Sardinia virus (TYLCSV) Rep, with the α-helices coloured in pink, β-strands depicted in yellow and small extension sheets in blue and loops are depicted as grey (Campos-Olivas et al., 2002). The image was generated using NGL viewer (Rose and Hildebrand, 2015; Rose et al., 2016).

The NMR structure solved by Campos-Olivas et al. (2002) revealed that the catalytic domain (aa 4-121) is made up of nine β-sheets and two α-helices. Residues FLTYP of motif I were identified on the structure and thus functionality could be linked to the placement of these residues. The amino acids of motif I are situated so as to form close connections with the histidine residue of motif II. Campos-Olivas et al. (2002) noted the structural similarities between TYLCSV Rep and DNA binding domain (DBD) of ssDNA viruses as well as other RCR initiator proteins.

Structural studies of the helicase domain were carried out by George et al (2014). The experimentally determined structure of BPV type 1a E1 protein was used as a structural template to assemble the helicase domain of ToLCGuV (genus *Begomovirus*) (George et al., 2014). By introducing mutations into the conserved AAA+ATPase family motifs and comparing the resultant activity, the roles that the individual amino acids play in helicase activity were hypothesised. Functional and structural information of SF3 helicase motifs aided in completing the structural model.

Full length replication proteins, as well as partial and mutated proteins have been expressed in a few organisms. Although a number of functional assays have been completed on wild type, mutant and partial proteins, limited structural characterisation has been carried out.

Elucidating the tertiary and quaternary structure of Rep using various biochemical methods will allow for a more complete molecular model to be developed.

1.8 Rationale for study

1.8.1 Novelty

Cassava is an economically important food crop in the developing world as it serves as the primary source of nutrition and income for many subsistence farmers. It is favoured because of its drought-tolerability and ability to grow in nutrient poor soils, making it an inexpensive staple food crop to grow and maintain (Burns et al., 2010). On the sub-Saharan African continent, cassava starch has been recognised as a valuable bio-resource for many industry-based processes (Baguma et al., 2017). In South Africa, cassava starch for industrial and food applications are being explored. Cassava starch can also be used for biofuels and for animal feed supplementation, and a recent cassava market study was funded by the Technical Innovation Agency (TIA)-funded project through Wits Enterprise. Due to its potential usage in different industries and its drought tolerating properties, it can be designated an important 21st century food security crop and, therefore, should be further invested in (Legg et al., 2014). Geminivirus disease in Africa has had debilitating effects on cassava yields as large amounts of cassava crops succumb to viral disease (Legg et al., 2014). Forming part of a small and unique group of plant viruses, the mode of action of ssDNA geminiviruses has not yet been fully elucidated. Research on Rep of begomoviruses and other geminiviruses has been extensive; however, no structural or functional research has been carried out on SACMV Rep. Structural and functional characterisation of Rep will aid in understanding the molecular mechanisms of Rep, and the roles it plays in viral infection and pathogenicity. Due to the structurally and functionally conserved nature of Rep across the virus family, and the similarities shared with other helicases and RCR initiation proteins, any structural or functional information revealed through SACMV Rep will be of importance.

1.8.2 Virus-Host interactome network expansion

Studying the protein-binding mechanisms of Rep, as well as potential host-interacting proteins would contribute greatly to the overall knowledge of this geminivirus protein. The virus-host interactome could be greatly expanded by revealing further host proteins that interact with Rep, in turn shedding further light on the multifarious role that this protein plays

in virus infection. Determining these host-protein interaction networks may potentially aid in establishing strategies of interfering with viral replication, a step toward pathogen control. Uncovering information on Rep may also help in directing further research into other plant and animal viruses, which may direct future research in the management of ssDNA viruses.

Chapter II – Expression as well as Functional and Structural Characterisation of SACMV Rep

2.1 Abstract

Recombinant protein expression enables the overexpression of a protein of interest using a modified expression host. The process of soluble protein overexpression and purification has been improved by the availability of expression vectors, chaperone vectors and purification tags. Geminivirus Rep acts as the viral replication initiator through binding and cleavage of the conserved nonanucleotide sequence contained within the origin of replication on the viral genome. Structural information of geminivirus Rep proteins is minimal, with the only structural (3D) information originating from NMR analysis of a truncated Rep protein. In order to study the structural and functional characteristics of SACMV Rep, the recombinant protein was overexpressed in an *Escherichia coli* host in the soluble fraction. Immobilised ion affinity chromatography (IMAC) was successfully carried out exploiting the co-expressed Histidine-Tag. Bioinformatics analysis revealed secondary structural features comparable to structurally characterised geminivirus Rep proteins. Binding of ATP to Rep did not yield substantial structural changes, as shown by intrinsic fluorescence, suggesting transient binding. Using extrinsic fluorescence, binding affinity of the hydrophobic probe, 8-Anilino-1-naphthalenesulfonic acid (ANS) to recombinant SACMV Rep was established, confirming the presence of binding pockets on the protein surface. Recombinant SACMV Rep displayed binding and cleavage of the nucleotide probe containing the conserved nonanucleotide sequence, confirming functional integrity of the recombinant viral protein.

2.2 Introduction

2.2.1 Expression, purification and characterisation of recombinant proteins

Recombinant protein expression describes the process of overexpressing a large amount of a protein of interest in a genetically modified organism for commercial purposes, as well as for use in studies of molecular and cell biology (Palomares et al., 2004).

Expression hosts, such as bacteria, yeast and mammalian cells, can be used to express recombinant protein, and hosts are chosen based on the particular expression requirements. The expression of smaller proteins (<~60 kDa) can be carried out using the expression system of *Escherichia coli* (Gräslund et al., 2008). *E. coli* is advantageous as it is a well-studied, low cost and fast growing expression host. There are a number of expression vectors designed for efficient expression of even toxic proteins in this host (Baneyx, 1999). *E. coli* is the preferred host for expression of industrial enzymes which are required in large quantities (Demain and Vaishnav, 2009). The disadvantages of recombinant protein expression include production of recombinant proteins within insoluble inclusion bodies, as inactive or truncated variants of the protein. Refolding of these insoluble proteins can prove difficult (Kyratsous et al., 2009). Another disadvantage of solubilising and re-folding recombinant protein is the possibility of the re-folded structure not adopting its true (native) conformation (Rudolph and Lilie, 1996). A number of different strains have been developed for successful recombinant protein production. *E. coli* BL21 (DE3) has proven to facilitate high levels of expression (Rosano and Ceccarelli, 2014). It is also a preferred strain due to the lack of *lon* and *ompT* proteases which can degrade the recombinant protein, and can be used with expression vectors that include a T7 RNA polymerase promoter region. The benefits of the DE3 strain include the induction of T7 RNA polymerase expression only in the presence of Isopropyl β-D-1-thiogalactopyranoside (IPTG), with bacterial cellular machinery directed predominantly toward recombinant protein expression (Gräslund et al., 2008). Other expression hosts, such as yeast and mammalian systems enable the expression of larger proteins, and proteins that require post-translational modifications, with the most well characterised and most used yeast strain being *Saccharomyces cerevisiae*. Mammalian expression systems are most advantageous for proteins requiring specific mammalian post-translational modifications including the addition of fatty-acid chains (Demain and Vaishnav, 2009).

Aggregated, insoluble protein expression is manifested through the production of inclusion bodies. Inclusion body formation can be curtailed through a number of methods. Decreasing the cultivation temperature can result in the expression of the recombinant protein in the soluble fraction. Additionally, overexpression of a chaperone can prevent insoluble protein expression (Baneyx, 1999). *E. coli* produces its own chaperone proteins, with the more commonly expressed chaperones including DnaK-DnaJ-GrpE, as well as GroEL-GroES. Overexpression of a recombinant protein may negatively affect the ability of the chaperones to assist in expression of high levels of soluble protein, resulting in aggregation of the protein of interest (Kyratsous et al., 2009; Nishihara et al., 1998). The overexpression of these chaperones using chaperone-expressing plasmids has been shown to induce correct folding and expression of an aggregation-prone recombinant protein in the soluble fraction (Baneyx, 1999). This system has been made available through chaperone plasmid sets (Takara Bio Inc., Japan) which enable the expression of the chaperone set DnaK-DnaJ-GrpE and GroEL-GroES independently of the recombinant protein (Nishihara et al., 1998).

Expression vectors can simplify purification of the recombinant protein through the addition of a fusion tag. The addition of Oligo-histidine (His-), glutathione S-transferase (GST-) and maltose-binding protein (MBP-) tags allow for a streamlined purification protocol to be used. Commercially available chromatography columns, such as the nickel ion (Ni^{2+}) or cobalt ion (Co^{2+}) charged, glutathione or cross-linked amylose matrices facilitate the extraction of the recombinant protein from bacterial cell lysate (Terpe, 2003). Fusion tags may also facilitate the over expression of stable, correctly folded recombinant proteins. MBP-fusion tags have been reported to improve the expression of eukaryotic proteins in the soluble fraction (Terpe, 2003). Trigger factor (TF), a 50S ribosomal subunit linked protein, is another fusion tag that induces soluble protein expression through its association with chaperones of the expression host (Baneyx, 1999).

Biophysical characterisation of recombinant protein structures is carried out by analysing secondary, tertiary and quaternary structure, as well as undertaking ligand binding assays. Functional studies may be carried out by determining the enzymatic abilities of the recombinant expressed protein (Achilonu et al., 2014; Zininga et al., 2015). Secondary structure can be evaluated using far-UV circular dichroism (CD), a technique using polarised light that measures the uneven absorbance of that polarised light by the protein. The resulting spectrum reveals the secondary structural features that make up the recombinant protein (Kelly et al., 2005). Tertiary structural insights are obtained through the examination of

protein intrinsic fluorescence. Intrinsic fluorescence is the analysis of the environment of tryptophan and tyrosine residues and a shift in the protein conformation that may result from binding of the protein to ligands (DNA or ATP) (Ghisaidoobe and Chung, 2014; Lakowicz, 2006). Quaternary structure can be studied through the use of size exclusion analysis, a technique that provides an insight into the molecular weight of the protein of interest (Kostanski et al., 2004). Functional characterisation of a recombinant protein can be undertaken by analysing enzymatic activity of the protein, such as ATPase activity or substrate cleavage (Achilonu et al., 2014; Naicker et al., 2013; Zininga et al., 2015).

2.2.2 DNA binding and cleavage activity of Rep

The crucial step of rolling circle replication initiated by SACMV Rep is the binding of the viral protein to the origin of replication and cleavage of the viral genome. Studies carried out with TGMV Rep initially showed that Rep bound to a region within the common region, a region conserved between DNA-A and DNA-B (Fontes et al., 1992). The protein bound to both viral genomes with no difference in binding affinities. Binding, however, was proven to be sequence specific as Rep of one viral strain would not recognise the binding site of a heterologous viral genome (Chatterji et al., 2000a; Fontes et al., 1994a).

The sequences that define virus specificity for binding activity were determined to be directly repeated sequence motifs, also labelled iterons. These motifs, between 12-13 bp in length and situated upstream of the stem-loop, were shown to act as cognate binding sites for Rep (Fontes et al., 1992, 1994a, 1994b). Further research proved that requirements for optimal viral replication included the recognition and binding of Rep to these iteron sequences. These iterons confer sequence specific cleavage activity, as is required for replication of the positive viral strand (Chatterji et al., 2000b; Laufs et al., 1995a). Competitive assays carried out with MYMIV, Tomato leaf curl New Delhi virus (ToLCV-Nde), TGMV and Bean golden mosaic virus (BGMV) Rep proteins has confirmed this is a requirement across different virus species (Chatterji et al., 2000a; Fontes et al., 1994a; Pant et al., 2001). The presence of a fourth iteron on the right hand side of the stem-loop was identified in MYMIV replication. This finding also showed that each iteron contributed differently toward virus specificity and binding (Singh et al., 2008).

The presence of inverted repeats within the CR of geminiviruses is conserved throughout the virus family, and has the capacity to form stem-loop or cruciform structures (Lazarowitz,

1987; Lazarowitz et al., 1992). *In vivo* studies showed that the presence of the conserved stem-loop is essential for Rep nicking activity as shown with TGMV Rep (Orozco and Hanley-Bowdoin, 1996). The requirement of a stem-loop for cleavage activity confirmed that Rep requires a single stranded substrate for cleavage (Laufs et al., 1995b, 1995a; Singh et al., 2008). Furthermore, the single stranded nature of the stem-loop is also essential for binding of host replication factors (Singh et al., 2008). This confirms the importance of DNA secondary structure, as stem-loop formation provides a single stranded nonanucleotide for positive strand replication. Secondary structure of origins of replication have also been confirmed to be crucial for parvovirus replication (Orozco and Hanley-Bowdoin, 1996).

Research carried out by Singh et al. (2008) has suggested a role for Rep in the formation of the stem-loop. The research suggested that after initial binding of oligomerised Rep to iterons, co-operative binding of further Rep oligomers around the nonanucleotide region would commence, which may result in the formation of the stem-loop structure. The stem-loop features a single stranded nonanucleotide which can act as the cleavage substrate for Rep, initiating replication (Singh et al., 2008) (Figure 2.1).

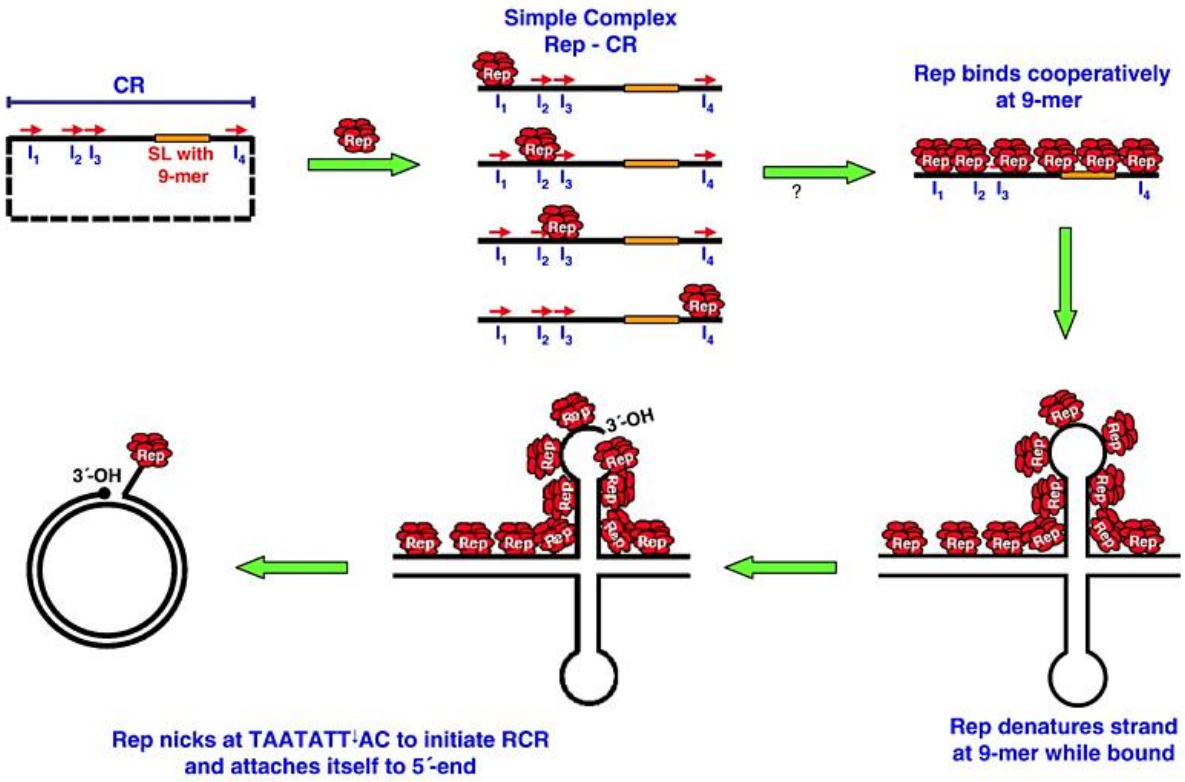

Figure 2.1 Model showing the predicted steps of Rep involvement in viral replication. Intial steps predict binding of geminivirus Rep to iterons with further cooperative binding of oligomeric proteins. Rep proteins are predicted to bind at the conserved nonanucleotide, unwind the double-stranded DNA intermediate to generate a cruciform structure containing the single standed DNA stem-loop, and carry out site specific cleavage on the viral sense strand. Rep remains bound to the 5′ end of cleaved strand (Singh et al. 2008).

In vitro studies of Rep binding and cleavage have revealed that the presence of iterons is not required for DNA binding and cleavage, nor is the formation of a stem-loop (Laufs et al., 1995a). A number of studies have used short nucleotide probes to confirm the DNA binding and cleavage activity of Rep. These probes contain the nonanucleotide cleavage site, and a few nucleotides downstream of the cognate sequence. Binding and cleavage was confirmed for TYLCV, TYLCSV, and ACMV Rep proteins, showing that Rep can bind to and cleave the single stranded nonanucleotide sequence, and form a protein-DNA adduct with the 5′ end of the cleaved probe (Campos-Olivas et al., 2002; Hipp et al., 2014; Kittelmann et al., 2009; Laufs et al., 1995a).

2.2.3 Structural knowledge of Rep and reasons for study

Limited research has been carried out on the molecular structure of Rep proteins of geminiviruses. To date, only the catalytic domain of TYLCSV has been resolved by NMR

analysis (Campos-Olivas et al., 2002). The 3D structure of TYLCSV Rep revealed structural similarities between this Rep and the DNA binding domains of SV40 L-Tag of the family *Polyomaviridae*, and E1 and E2 proteins of Papillomaviruses (Campos-Olivas et al., 2002).

The catalytic domain revealed the central 5-stranded antiparallel β-sheet, consisting of β_2, β_3, β_4, β_8 and β_9 with two smaller β-strands around the side of the central sheet (β_5 and β_9) making up a small β-sheet. An additional two β-strands (β_6 and β_7) form a β-stem-loop. The one side of the central β-sheet is buried by the presence of one of the α-helices (α_1), and the two-stranded β-sheet. The other face of the β-sheet is more exposed, only facing the other α-helix (α_2) of the catalytic domain. The catalytic tyrosine is situated on α_2, directed at the exposed side of the central β-sheet, suggesting that this surface is directly involved in DNA binding and catalytic activity (Campos-Olivas et al., 2002).

Motif I is situated on β_2, with lysine and tyrosine aiding in the folding and packing of the hydrophobic core of the protein. Phenylalanine and threonine in contrast are situated on the outside of the beta sheet, exposed to the solvent and form close connections with motif II histidine residues, situated on β_2. The negative side chain of glutamic acid in close proximity to the histidines of motif II may act as a metal ion coordination site. Binding of either Mg^{2+} or Mn^{2+} is essential; however, the importance of this specific glutamic acid has not been verified as yet (Campos-Olivas et al., 2002; Laufs et al., 1995a).

It was suggested that positively charged residues situated on the extended β-sheet containing β_1 and β_5 interact with dsDNA. Separation of dsDNA allows for *ori* ssDNA to come into contact with the exposed central β-sheet. The catalytic tyrosine, situated on the exposed side of the beta sheet, is primed for DNA binding and endonuclease activity. DNA binding proteins, often involved in viral replication, interact with the nucleotides through this exposed hydrophobic residue (Campos-Olivas et al., 2002). Neighbouring aromatic and hydrophobic side chains aid in DNA binding. To determine which amino acid residues take part in DNA binding, comparative studies were carried out, looking at SV40 L-Tag and papillomavirus E1, as these proteins share functional similarities as well as structural likeness with geminivirus Rep. It was suggested that the small beta sheet extending from the large beta sheet, involving β_1 and β_5, and the positively charged residues contained within this sheet enable DNA origin recognition and binding (Campos-Olivas et al., 2002).

Secondary structural modelling studies carried out on ToLCGuV RepC (aa 122-361) revealed an α-helical spectrum. Using the K2D algorithm (Andrade et al., 1993), the c-terminal

fragment was predicted to consist of 50% random coil, approximately 37% α-helix and 13% β-sheet (George et al., 2014). To study structural features of the protein, the research group modelled RepC on the solved structure of BPV E1 protein (PDB ID 2GXA), which also forms part of the SF3 helicases and AAA+ATPase family (Enemark and Joshua-Tor, 2006). The model revealed that the Walker A and B motifs are solvent-exposed, as they are situated at subunit interfaces primed for ATP binding (George et al., 2014). The crystal structures of other SF3 helicases (AAV-Rep40) display similar architecture of the secondary structures, with some differences observed (James et al., 2003).

The research team further observed that certain amino acids proved crucial in the proper functioning of the helicase domain. The study showed that an aspartic acid in the Walker B motif aids in proficient ATP hydrolysis, proving this negatively charged amino acid is indispensible for successful replication (Choudhury et al., 2006; George et al., 2014). Additional analysis indicated that two lysine residues, that form part of the B′ motif in ToLCGuV, play a role in DNA binding and helicase activity. The conformational changes induced through mutations showed that the positively charged lysine situated in the B′ motif is located within a β-hairpin structure within the channel of the protein, and aids in binding to ssDNA prior to unwinding activity. The importance of the β-hairpin loop has also been confirmed in SV40 L-Tag, AAV2 Rep and BVP E1 (George et al., 2014; Shen et al., 2005). The lysine residue situated in the B′ motif has been shown to be crucial in inducing structural changes that occur after binding ATP. These allosteric modifications allow the positively charged residue on the β-hairpin loop contained within the central pore of the helicase domain to bind to ssDNA and commence unwinding of viral DNA (George et al., 2014).

With very little structural information available, the functions of geminivirus Rep proteins have largely been elucidated by mutagenesis with subsequent alterations of selected amino acids. The effects of these mutations on mutant Rep enzymatic activity have been observed. Domain studies have been carried out by analysing the functional activity of partial protein (George et al., 2014; Orozco et al., 2000). The studies have shown that changes to the primary structure of Rep can affect functional activities, as it is known that protein structure affects function (Walsh, 2002). Using helicases of SF3 or RCR initiator proteins as models, the structure of geminivirus Rep proteins can be further elucidated (George et al., 2014). In vitro protein studies provide a more reliable assessment of protein structure and function than in silico analysis which utilises homology modelling.

By elucidating the structure of Rep, the interactions it forms with host proteins can be examined which in turn lends further information to virus-host interaction studies. Structural information on Rep would contribute greatly to the overall knowledge of this geminivirus protein, and may potentially aid in determining strategies of interfering with viral replication, a step toward pathogen control. Furthermore, as Rep is an SF3 helicase, any structural information on geminivirus Rep would help in expanding the knowledgebase of this large group of helicases.

2.2.4 Aim

The aim of this study is to carry out a functional and structural characterisation of SACMV Rep.

2.2.5 Specific objectives

- To predict structural and biophysical characteristics of SACMV Rep from bioinformatics-derived data
- To perform co-transformation of Rep with chaperone vector, pG-KJE8 into *Escherichia coli* BL21 (DE3)
- To overexpress soluble recombinant Rep with *E. coli* chaperones (DnaK, DnaJ, GrpE, GroEL and GroES) expressed from pG-KJE8 and purify using IMAC
- To confirm functionality of recombinant Rep through an *in vitro* DNA cleavage assay
- To use far-UV CD to elucidate the secondary structural features of recombinant Rep
- To study intrinsic and extrinsic fluorescence of Rep using ANS and ATP

2.3 Methods and Materials

2.3.1 Bioinformatics study of SACMV Rep

The lack of knowledge of SACMV Rep prompted an initial bioinformatics study of the protein. In order to determine some physicochemical characteristics of the protein, the amino acid sequence (Berrie et al., 2001) was entered into the online tool ExPASy ProtParam (Gasteiger et al., 2005)

Secondary structure prediction was completed using Phyre2 (Kelley and Sternberg, 2009) a web portal that enables access to a collection of prediction tools. The secondary structure was predicted and analysed using PSIPRED (Jones, 1999). PSIPRED uses an algorithm that makes use of two neural networks that use both position-specific scoring matrix (PSSM) and position specific iterate-basic local alignment tool (PSI-BLAST) to predict the presence and location of secondary structures using only the amino acid structure as input (Altschul et al., 1997).

Using a multiple sequence alignment tool, Clustal Omega (Sievers et al., 2011) the sequences of a number of begomoviral Rep proteins were aligned, and the conserved motifs were identified.

2.3.2 Cloning of SACMV *Rep* into pET-28a and sequence confirmation

The AC1 ORF encoding the SACMV replicase protein (NP_620665.1) (Figure 2.2), also labelled the replication initiator protein, was previously cloned into pJET1.2 from an infectious clone, pC8A (Mwaba, 2011). This was carried out by polymerase chain reaction (PCR) amplifying AC1 from a plasmid construct, pC8A (containing the ORFS of SACMV DNA-A) with primers encompassing restriction sites at the 5′ end. A *Bam*HI and *Not*I restriction site were added to the forward and reverse primer, respectively (Table 2.1). PCR amplified AC1 was cloned into pJET1.2 and transformed into *E. coli* DH5α. Sequencing was performed on extracted plasmid (pJET1.2/AC1) to confirm correct nucleotide sequence of SACMV AC1 (Inqaba Biotech, Pretoria).

A 5′- ATGCCGAGGGCTGGTCGTTTTAGCATAAAAGCCAAAAATTATTTCCTCACGTATCCGAAATGCACTCTCT
CGAAAGAAGCGGCATTAGATCAACTCCGACAACTCCAAACCCCAACAAATAAATTGTTCATCAAGATCTG
CAGAGAACTCCATGAAAATGGGGAACCTCATTTGCATGCCCTCATTCAGTTCGAGGGCAAGTACAATTGT
ACCAACCAACGATTCTTCGACCTCATATCCCCTTCCAGGTCAACACATTTCCATCCAAACATTCAGGGAG
CTAAATCCAGTTCTGACGTCAAGTCCTATTTGGACAAGGACGGAGACACCATCCAATGGGGCGAGTTTCA
GATCGACGGACGATCTGCTCGCGGCGGACAACAATCCGCCAATGACGCTTACGCCAAGGCTCTTAACGCA
GCAAGTAAAACAGAGGCTCTTAATGTAATCCGGGAACTAGCCCCAAAGGATTTTGTTTTACAGTTTCATA
ATTTAAATAGCAATTTAGATAGGATTTTTCAGGAGCCTCCGATTCCTTATATTTCTCCCTTTCTTTCTTC
TTCTTTCACTCATGTTCCTGAGGAACTTGAAGACTGGGTTTCCGAGAACGTGATGGGTTTCGCTGCGCGG
CCATGGAGACCGAGTAGTATCGTCATCGAGGGCGATAGTAGGACAGGGAAGACGATGTGGGCCCGATCTC
TGGGACCACACAACTACTTATGTGGACATTTGGATCTCAGTCCAAAGGTTTACAGCAACGACGCATGGTA
CAACGTCATTGATGACGTCGACCCCCATTACCTCAAGCACTTCAAAGAATTCATGGGGGCCCAAAGGGAC
TGGCAAAGCAATACCAAGTACGGGAAGCCGATTCAAATTAAAGGCGGCATTCCCACTATCTTCCTATGCA
ATCCAGGACCGACATCATCATATAAAGAGTTTCTGGACGAGGAAAAGAACCAGTCCCTTAAAGCCTGGGC
TTTAAAGAATGCAACCTTCATCACCCTCCACGAGCCATTGTTCTCAAGTGCCCATCAAAGTCCAACACCG
CACCGCGAAGACTAG -3′

B

```
1    MPRAGRFSIKAKNYFLTYPKCTLSKEAALDQLRQLQTPTNKLFIKICREL
51   HENGEPHLHALIQFEGKYNCTNQRFFDLISPSRSTHFHPNIQGAKSSSDV
101  KSYLDKDGDTIQWGEFQIDGRSARGGQQSANDAYAKALNAASKTEALNVI
151  RELAPKDFVLQFHNLNSNLDRIFQEPPIPYISPFLSSSFTHVPEELEDWV
201  SENVMGFAARPWRPSSIVIEGDSRTGKTMWARSLGPHNYLCGHLDLSPKV
251  YSNDAWYNVIDDVDPHYLKHFKEFMGAQRDWQSNTKYGKPIQIKGGIPTI
301  FLCNPGPTSSYKEFLDEEKNQSLKAWALKNATFITLHEPLFSSAHQSPTP
351  HRED
```

Figure 2.2 Nucleotide and amino acid sequence of SACMV Rep. Nucleotide (A) and amino acid sequence (NP_620665.1) (B) of South African cassava mosaic virus (SACMV) Rep protein.

Table 2.1 Primers for amplification of SACMV AC1. Primers designed to amplify the South African cassava mosaic virus (SACMV) AC1 open reading frame (ORF) from infectious clone pC8A. Restriction sites were added to the 5′ end of forward and reverse primers and are shown in bold. The names of the primers are shown.

Primer Name	Nucleotide sequence	Added restriction site
RepEXPF16	5′- TAAT**GGATCC**ATGCCGAGGGCTGGTCGTTT	*Bam*HI
RepEXPRA16	5′- TAAT**GCGGCCGC**GTCTTCGCGGTGCGGTGT TGGACTT	*Not*I

It is important to note that the infectious clone utilised in the present study contains a single nucleotide mutation in the AC1 ORF at nucleotide position 620. This mutation arose

spontaneously and resulted in an amino acid mutation at position 207 from a phenylalanine in the originally-deposited sequence (NP_620665.1) to a serine in the infectious clone used in this study (Figure 2.8). This amino acid mutation is located outside the functional motifs of Rep and is also present in the Zimbabwean variant of SACMV (CAE01424.1).

An expression vector was kindly provided by Dr Ikechukwu Achilonu (Protein Structure Function Research Unit, School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg). Sub-cloning of AC1 from pJET1.2 into pET-28a was undertaken by performing a double restriction digest on pJET1.2/AC1 and pET-28a using Fast digest™ *Bam*HI and *Not*I (Thermo Fisher Scientific, Waltham, USA). Expression of recombinant protein from pET-28a incorporates an N- and C-terminus His-Tag onto the protein product, thereby enabling purification using IMAC. The digested products were run on a 1% agarose gel, and gel extraction was carried out on digested pET-28a/AC1 using the GeneJet Gel Extraction Kit (Thermo Fisher Scientific, Waltham, USA). The resulting products were quantified using a NanoDrop™ 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, USA). The vector and insert were mixed at a 1:3 ratio. The ligation reaction was carried out with the components and quantities as shown in Table 2.2. Ligation reactions were carried out to ligate the gene of interest, AC1, into the destination vector, pET-28a. The reaction was carried out in a final volume of 20 µL. 5× ligation buffer and T4 DNA ligase are products of Thermo Fisher Scientific (Waltham, USA).

Table 2.2 Ligation reaction components for ligation of AC1 into pET-28a.

Reaction component	**Volume (µl)**
H_2O	1.6
5× ligation buffer	4
Vector (linearised pET-28a)	8
Insert (AC1)	4.4
T4 DNA ligase	2
Total reaction volume	20

The ligation reaction was incubated at room temperature overnight then transformed into competent *E. coli* DH5α using the heat-shock method. This method involves the addition of plasmid to a vial of competent cells and incubating at 42° C for 60 s, followed by the addition of super optimum broth (SOB) (see Appendix) with catabolyte repression media (SOC) (see

Appendix). The cells were grown at 37° C for one hour, before being plated onto selective media and incubated overnight at 37° C. From the resulting colonies a colony PCR was performed using primers (Table 2.1). Positive colonies were grown overnight in liquid SOB media at 37° C. Plasmid extraction was carried out on cultures using the GeneJET plasmid miniprep kit (Thermo Fisher Scientific, Waltham, USA). After quantification, a restriction digest using Fast digest™ *Not*I and *BamH*I (Thermo Fisher Scientific, Waltham, USA) was performed to confirm the presence of the plasmid and insert of interest. Extracted plasmid was sent for sequencing using the T7 forward primer and T7 terminal reverse primer (Table 2.3) (Inqaba Biotech, Pretoria).

Table 2.3 Primers used for amplification of AC1 in pET-28a for sequencing.

Primer Name	**Nucleotide sequence**
T7 forward primer	5′- TAATACGACTCACTATAGGG
T7 terminal reverse primer	5′- GCTAGTTATTGCTCAGCGG

The gene of interest was extracted from the sequencing data (see Appendix) and translated using an online program, Translate (Gasteiger et al., 2005). The translated amino acid sequence was compared to the protein sequence of SACMV Rep (NP_620665.1) using basic local alignment services tool (BLAST) (Altschul et al., 1990). The recombinant expression vector, pET-28a/AC1 (Figure 2.4) was confirmed to contain SACMV AC1 encoding SACMV Rep.

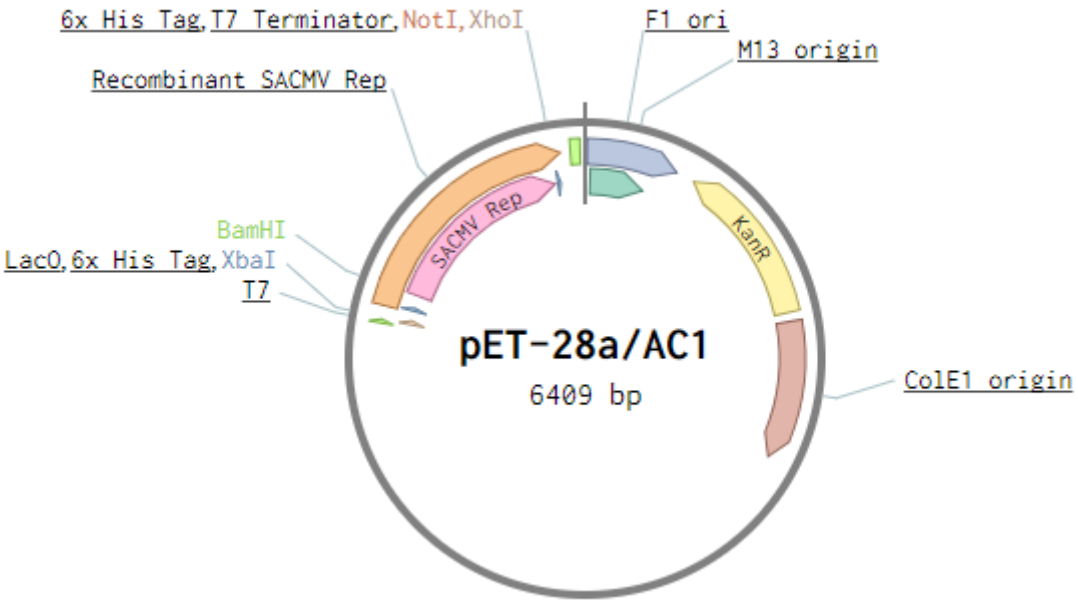

Figure 2.3 Schematic representation of expression vector pET-28a/AC1. AC1 encodes for the South African cassava mosaic virus (SACMV) Rep protein. Recombinant AC1 encoding for recombinant Rep is shown in orange. The locations of the N- and C-terminal tags are indicated as 6x His-Tag. The T7 promoter and terminator regions are indicated on the plasmid map. Restriction sites *Bam*HI, *Not*I and *Nco*I sites are shown in colour. Kanamycin resistance gene is labelled in yellow as KanR, with the ColE1 origin shown in brown. F1 ori is labelled in blue, with the M13 origin labelled alongside it in turquoise. T7 promoter and terminator are labelled in green. Lac operon is shown in light pink.

2.3.3 Transformation of *Escherichia coli* with chaperone plasmid, pG-KJE8, and pET-28a/AC1

The chaperone plasmid pG-KJE8 (Takara, Clontech, USA) was kindly provided by Dr Ikechukwu Achilonu (Protein Structure Function Research Unit, School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg). *E. coli* JM109 containing chaperone plasmid pG-KJE8 was cultured overnight in SOB media. The plasmid was extracted using the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific, Waltham, USA) and digested using Fast digest^{TM} *Eco*RI (Thermo Fisher Scientific, Waltham, USA) to confirm its presence. The plasmid was transformed into competent *E. coli* BL21 (DE3) using the heat shock method and cultured overnight in liquid broth. *E. coli* BL21 (DE3) pG-KJE8 cells were made competent using the Inoue method as per Inoue et al. (1990). Newly competent cells containing pG-KJE8 were transformed with pET-28a/AC1 and plated onto SOB media with 20 $\mu g.mL^{-1}$ chloramphenicol. Colonies from the transformation were cultured in liquid media. Cell cultures were used to make 2:1 glycerol stocks (cell culture: 50% glycerol) and for plasmid extraction using the GeneJET Plasmid Miniprep

Kit (Thermo Fisher Scientific, Waltham, USA). Restriction digest was carried out on pET-28a/AC1 and pG-KJE8 with Fast digest^{TM} *Nco*I (Thermo Fisher Scientific, Waltham, USA), as this restriction enzyme is known to cut both pET-28a/AC1 and pG-KJE8 twice.

2.3.4 Expression trials

Expression trials of the recombinant *E. coli* were performed in SOB, using a range of L-arabinose and tetracycline concentrations to induce expression of the chaperones from pG-KJE8. A range of IPTG was used to induce expression of Rep from pET-28a. Expression of DnaK-DnaJ-GrpE was induced through L-arabinose. The addition of tetracycline induced the expression of GroES-GroEL.

Glycerol stocks of recombinant *E. coli* BL21 (DE3) as prepared in Section 2.3.3 were used to inoculate 50 mL SOB overnight culture, supplemented with 50 μg.mL^{-1} kanamycin and 20 μg.mL^{-1} chloramphenicol, incubated at 37° C and shaking at 200 rpm. The culture was harvested by centrifugation at 5000 ×*g* for 15 min. A 40-fold dilution was made into fresh SOB, supplemented with 50 μg.mL^{-1} kanamycin, 20 μg.mL^{-1} chloramphenicol, and concentrations of L-arabinose ranging from 0.5-2 mg.mL^{-1} as well as 5 ng/mL^{-1} tetracycline. Cultures were incubated at 37° C until OD_{600} was between 0.4 and 0.6. Protein expression was induced using combinations of IPTG (0.1, 0.3, 0.5 mM) and L-arabinose (0.5-3 mg.mL^{-1}) at 18 ° C for 18 h to evaluate conditions that enabled expression of Rep in the soluble fraction. These conditions were selected due to their success in leading to soluble expression of recombinant SACMV Rep in another pET expression system.

Samples (1 mL) were collected after the 18 h induction period of recombinant Rep and chaperones with varying concentrations of inducers. Samples were spun down at 5000 ×*g* for 15 min, and the supernatant discarded. The pellets were resuspended in buffer A [25 mM Na_2HPO_4/NaH_2PO_4, pH 7.0, 10% (v/v) glycerol, 0.05% (v/v) β-mercaptoethanol, 0.5 M NaCl, 20 mM imidazole, 0.02% (w/v) NaN_3] and 0.5 mg.mL^{-1} lysozyme. Glycerol was added to stabilise the soluble, recombinant protein (Yamaguchi and Miyazaki, 2014). Cell lysate was sonicated at 15 MHz using the Misonix Microson ultrasonic cell disruptor (Misonix Inc, New York, USA) for 10 s bursts 5 times in order to lyse the cells. Cell lysate was fractionated by centrifugation at 20000 ×*g* for 20 min at 4° C. All subsequent purification steps were carried out at 4° C. Soluble fraction was removed and the insoluble pellet was resuspended in 1% sodium dodecyl sulfate (SDS). Samples (100 μL) were taken from both fractions and

separately combined with 2× sample buffer [50 mM tris(hydroxymethyl)aminomethane (Tris)-HCl, 20% (v/v) glycerol, 10% (w/v) SDS, 100 mM β-mercaptoethanol, 0.05% bromophenol blue, pH 6.8] at a 1:1 ratio and resolved on a 12% TGX Stain-Free™ FastCast™ acrylamide gel (Bio-Rad, Hercules, CA, USA).

2.3.5 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis

SDS, an anionic detergent, is used for SDS-PAGE as it confers a uniform negative charge across proteins and ensures that the proteins are converted into linear molecules. The shape and charge conferred by SDS guarantees that the individual polypeptides are separated based on individual polypeptide molecular weight, not by charge or shape (Pitt-Rivers and Impiombato, 1968).

SDS-polyacrylamide gel electrophoresis (PAGE) was carried out using the TGX Stain-Free™ FastCast™ Acrylamide solutions (12% (w/v)) (Bio-Rad, Hercules, USA). The TGX Stain-Free™ FastCast™ Acrylamide kit uses the Laemmli SDS-PAGE method, with a trihalose compound included in the gels, which allows for easy visualisation with the Gel Doc™ EZ system (Bio-Rad). Tryptophan (Trp) residues within the protein sample react with the trihalose compound, releasing a fluorescent signal when visualised.

Gels were set up as described in the kit. Protein samples were combined with 2× sample buffer and heated at 90° C for 5 min. Sample volumes of 15 µL were loaded onto the gel alongside the Precision Plus Protein™ Unstained Standards (Bio-Rad, Hercules, USA), which contained reference bands from 10 – 250 kilo Daltons (kDa). Electrophoresis was carried out using Tris/Glycine/SDS buffer (Bio-Rad, Hercules, USA) (25 mM Tris, 192 mM Glycine, 0.1% SDS, pH 8.3) and run using the PowerPac™ Basic Bio-Rad electrophoresis system (Bio-Rad, Hercules, USA) at 300 V for 20 min. Gels were visualised using the Gel Doc™ EZ imager (Bio-Rad, Hercules, USA).

2.3.6 Large scale expression and purification of Rep

A 50 mL overnight culture, containing 50 $\mu g.mL^{-1}$ kanamycin and 20 $\mu g.mL^{-1}$ chloramphenicol, was inoculated from glycerol stock and cultured at 37° C for 18 h, shaking at 200 rpm. Larger volumes of SOB supplemented with 50 $\mu g.mL^{-1}$ kanamycin, 20 $\mu g.mL^{-1}$ chloramphenicol, 2 $mg.mL^{-1}$ L-arabinose and 5 $ng.uL^{-1}$ tetracycline were

prepared. The overnight culture was used to inoculate the expression broth and the overnight culture volume used was 1:25 of the expression broth volume. The appropriate amount of overnight culture was spun down at 5000 ×*g* for 15 min and the pellet was resuspended in 1 mL of fresh broth, and used to inoculate the large volume of expression broth. The culture was grown at 37° C until the OD_{600} measurements were observed to lie between 0.4-0.6. The culture was cooled down to approximately 15° C. Expression was carried out at 18° C for 18 h after induction with 0.3 mM IPTG. Cell cultures were centrifuged to collect cell material at 7000 ×g for 20 min. Cell pellets were stored at -80 ° C.

At a ratio of 5:1, 5 mL of buffer A [0.5 $mg.mL^{-1}$ of lysozyme, 1 mM phenylmethylsulfonyl fluoride (PMSF)] was added per gram of wet cell pellet. Resuspension was aided by 30 min of gentle agitation on ice. Sonication was carried out to lyse the cells, and release all intracellular material. This was performed in five 30 s bursts, with a 30 s rest interval between bursts. The sonic output was recorded as 15 MHz using the Misonix Microson ultrasonic cell disruptor (Misonix Inc, New York, USA). The lysed cell content was fractionated by centrifugation at 25000 ×g for 25 min at 4° C. The soluble fraction was aspirated and filter-sterilised using a 0.45 µm syringe filter.

IMAC was used to purify recombinant Rep protein. N- and C-terminal His-Tags that are co-expressed with Rep allow for this technique to be applied to recombinant protein. Chelating agents, which are bound to a solid chromatography support, attract nickel (Ni^{2+}) or cobalt ions (Co^{2+}), which in turn exhibit an affinity for His-Tagged recombinant proteins. The imidazole side chain of histidine displays an affinity for chelated metal ions, enabling the binding of the His-Tag to the metal ions. Other cell lysate proteins may display an affinity for the Ni^{2+}- charged chromatography support, but are prevented from binding to the column by the presence of NaCl and low concentrations of imidazole in Buffer A. Multiple wash steps may be carried out to elute protein or DNA bound non-specifically to the column. The recombinant protein is eluted using a specific concentration of imidazole, which outcompetes the His-Tag to bind to the metal ions (Arnau et al., 2006).

A 5 ml nickel ion (Ni^{2+}) -charged FF HisTrap column (GE Healthcare, USA) was attached to ÄKTA fast protein liquid chromatography (FPLC) purification system (GE Healthcare, USA) coupled to a computer with Primeview 1.0 software (GE Healthcare, USA). The column was equilibrated using Buffer A. Soluble protein was loaded onto the column at an injection flow rate of 1 $mL.min^{-1}$ and the column was washed with Buffer A. The flow through was

collected for further analysis. The column was washed with ten column volumes of Buffer A to ensure that unbound proteins had been removed. Protein was eluted using a gradient elution with buffer A and elution buffer [25 mM Na_2HPO_4/NaH_2PO_4, 10% (v/v) glycerol, 0.05% (v/v) β-mercaptoethanol, 0.5 M NaCl, 700 mM imidazole, 0.02% (w/v) NaN_3]. As contaminating proteins were removed from the column, so the ratio of elution buffer to buffer A increased, thereby subjecting the column to an increasing percentage of elution buffer and in turn increasing concentrations of imidazole. All fractions were collected for further analysis. Presence of eluted Rep was confirmed by running all collected fractions, including the cell lysate, soluble and insoluble fractions obtained previously, on a 12% TGX Stain-Free™ FastCast™ SDS-PAGE gel (Bio-Rad, Hercules, USA). A western blot of the SDS-PAGE gel was carried out to verify the presence of His-Tagged recombinant protein using an anti-His primary antibody (Santa Cruz Biotechnology, USA).

The Western blot was carried out by transferring the protein bands from the SDS-PAGE gel to a Trans-Blot Turbo™ PVDF membrane (Bio-Rad, Hercules, USA) using the Trans-Blot Turbo Transfer System (Bio-Rad, Hercules, USA). The membrane was incubated in 5% blocking buffer for 30 min while agitating. Blocking buffer was made up of Blotting-Grade blocker (Bio-Rad, Hercules, USA) in TBSTT (10 mM Tris-HCl, 150 mM NaCl, 0.1% Tween 20, pH 7.5). The membrane was then washed with TBSTT 3x 5 min. The membrane was incubated with primary His-probe (H3) monoclonal antibody (Santa Cruz Biotechnology, USA), diluted 1:5000 in blocking buffer, for 1 h with agitation. The membrane was washed in TBSTT 3x 5 min. Secondary antibody, Goat anti-mouse HRP conjugated antibody (Bio-Rad, Hercules, USA), diluted 1: 15000 in blocking buffer, and 1 µL Streptactin-HRP (Bio-Rad, Hercules, USA) was applied to the membrane and incubated for 30 min with agitation. The Streptactin-HRP was added to the secondary antibody mixture to detect the Precision Plus Protein™ Unstained Protein Standards (Bio-Rad, Hercules, USA). The membrane was washed to remove non-specifically bound secondary antibody 3x 10 min. The blot was visualised using Clarity™ Western ECL Substrate (Bio-Rad, Hercules, USA) and the Gel Doc™ EZ imager (Bio-Rad, Hercules, USA).

2.3.7 Quantification of pure recombinant Rep protein

Eluted protein fractions were separated on a 12% TGX Stain-Free™ FastCast™ SDS-PAGE gel (Bio-Rad, Hercules, USA) and imaged using the Gel Doc™ EZ Imager (Bio-Rad,

Hercules, USA). Fractions of purified Rep were pooled together and a serial dilution was performed. Buffer was filter-sterilised using a 0.2 μm filter. Full spectrum readings (240-350 nm) were taken of the dilution series (consisting of 6 fractions) using a UV-Vis spectrophotometer (Jasco, V630 Spectrophotometer, Jasco Inc., Easton, USA). Absorbance readings (A_{280}) were corrected by subtracting buffer absorbance as well as the A_{340} reading, which refers to the absorbance of aggregated protein. Corrected absorbance of six diluted protein samples and corresponding concentration factors were fitted to a linear regression line. The slope value from the resultant straight line equation was used to calculate the concentration of protein present ($mg.mL^{-1}$) using Equation 1.

$$\frac{mg}{ml} = \frac{slopexM_r}{\varepsilon} \qquad \text{Eq. 1}$$

Where M_r represents the molecular weight and ε represents the molar extinction coefficient.

Eq. 1 is used to work out the $mg.mL^{-1}$ of protein. Using Equation 2, the Beer-Lambert Law, the concentration (in M) of Rep can be worked out. The slope acquired from the standard curve would be substituted as A_{280}.

$$A = \varepsilon cl \qquad \text{Eq. 2}$$

Where A represents the absorbance, l the light path length (cm), c is the molar concentration, and ε the predicted molar extinction coefficient ($M^{-1}cm^{-1}$).

The molar extinction coefficient was predicted using ProtParam (Gasteiger et al., 2005) and was calculated using Equation 3A. The calculation uses the primary structure and the extinction coefficient of three amino acids, tyrosine, tryptophan and cystine (Gill and von Hippel, 1989) (Eq. 3B). Cystine residues contribute to the molar extinction coefficient; however, due to the presence of a reducing agent in protein elution and storage buffers in this study, cystines were converted to cysteines. Cysteines do not contribute to the calculation of the molar extinction coefficient due to their inability to absorb light at wavelengths above 260 nm (Pace et al., 1995), and do therefore not play a role in Eq. 3A and Eq. 3B.

$$\varepsilon(prot) = \varepsilon(tyr) \times n(tyr) + \varepsilon(trp) \times n(trp) + \varepsilon(cys) + n(cys) \qquad \text{Eq. 3A}$$

$$\varepsilon(280)(M^{-1}cm^{-1}) = 1490 \times \varepsilon(tyr) + 5500 \times n(trp) + 125 \times n(cys) \qquad \text{Eq. 3B}$$

Where ε refers to the molar extinction coefficient of the individual amino acid, and n refers to the number of each particular amino acid. ε (prot) refers to the predicted molar extinction coefficient.

2.3.8 Structural and functional characterisation of recombinant Rep

2.3.8.1 Intrinsic fluorescence of Rep in the presence of ATP

Fluorescence is a phenomenon where a molecule can absorb light at a certain wavelength. The molecules of the fluorophore absorb the light energy, resulting in the electrons moving to higher energy level. As light is emitted so the electrons move back to the ground state energy level (Figure 2.4). The wavelength of the emitted light is longer than the wavelength of the light used for excitation of the fluorophore (Lakowicz, 2006).

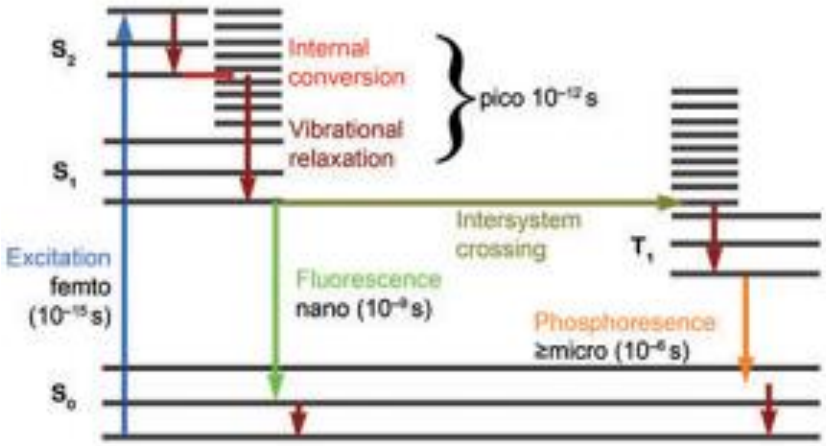

Figure 2.4 Jablonski diagram showing the fluorescence phenomenon. Diagram shows the energy states and the time interval taken for fluorescence or phosphorescence to occur. Excitation is shown by the blue line, with internal conversion and vibrational relaxation depicted by the maroon arrows. Fluorescence is shown by the green arrow and takes nanoseconds. Phosphorescence and intersystem are shown in orange and khaki (Lichtman and Conchello, 2005).

Proteins contain amino acids that can act as fluorophores, such as tryptophan, tyrosine, phenylalanine and cystine. The latter two amino acids demonstrate a very weak fluorescence signal and are therefore not used when studying protein fluorescence. Recombinant SACMV Rep contains seven tryptophan (Trp) and 12 tyrosine (Tyr) residues (Gasteiger et al., 2005). Tryptophan is labelled a strong fluorophore as it displays a good quality signal and quenches tyrosine quantum yields. Only tryptophan can be excited at 295 nm, whereas an excitation wavelength of 280 nm excites both tryptophan and tyrosine. Both fluorophores can be used to

monitor fluctuations in the protein tertiary structure. A hypsochromic shift (blue shift) is indicative of a decreased polarity of the fluorophore's environment. This can be attributed to burying of the tryptophan, thereby shielding it from the polar solvent environment. Contrastingly, the bathochromic shift (red shift) indicates the exposure of the fluorophore to a more polar environment, such as the polar solvent (Lakowicz, 2006).

Pooled and quantified protein fractions were dialysed into storage buffer containing 50 mM Tris, 0.5 M NaCl, 200 mM L-Arginine, 5 mM 1,4-dithiothreitol (DTT), 10% (v/v) glycerol, and 0.02% (w/v) NaN_3. Protein was quantified as described in Section 2.3.7. Rep protein at a concentration of 1 μM was diluted in MilliQ water and immediately subjected to excitation at 290 nm using the Jasco FP-6300 Spectrofluorometer (Jasco Inc., Easton, USA). Emission spectra were recorded from 300–500 nm. The scanning speed was set to 500 $nm.min^{-1}$, and both the excitation and emission band widths were set to 5 nm. Three accumulations were recorded and the emission spectrum was corrected for buffer emission. Intrinsic fluorescence was performed on Rep to monitor intrinsic fluorescence of tryptophan residues, and to observe whether the addition of ATP would result in protein conformational changes which would subsequently affect the fluorescence emission of the tryptophan residues (Lakowicz, 2006).

2.3.8.2 Extrinsic fluorescence using ANS

ANS is a fluorescent probe used to study protein unfolding and ligand binding. Crystal studies have shown that ANS binds to hydrophobic cavities through the aniline and naphthalene moieties (Hawe et al., 2008; Matulis and Lovrien, 1998; Stryer, 1965).

A change in fluorescence of ANS is indicative of binding to hydrophobic pockets of the protein. In a polar environment such as a solvent, the probe exhibits almost negligible fluorescence emission. When bound to a hydrophobic pocket which is an environment that is less polar than the polar solvent, a blue shift as well as an increase in quantum yield of the emission maximum (470 nm) is noted (Cattoni et al., 2009). Interactions between ANS and the protein occur through positively charged amino acids, such as lysine, histidine and arginine, located within the hydrophobic pocket (Matulis and Lovrien, 1998).

A 20 mM stock of ANS was prepared by combining a measured amount with ANS buffer, which consisted of 100 mM sodium phosphate (pH 6.5), 1 mM ethylenediaminetetraacetic acid (EDTA) and 0.02% (w/v) NaN_3. The solution was protected from light and agitated for

one hour. The exact concentration was determined by performing a dilution series on the stock, and measuring absorbance of each dilution at 350 nm using the Jasco V-630 Spectrophotometer (Jasco Inc., Easton, USA). The absorbance of each dilution was plotted against the respective concentration factor and the concentration for ANS was determined using Eq. 1.

The emission spectrum of 100 μM ANS in storage buffer and 100 μM ANS with 5 μM Rep was recorded after excitation at 390 nm using the Jasco FP-3600 Spectrofluorometer (Jasco Inc., Easton, USA). Emission spectrum for 100 μM ANS, 5 μM Rep and 8 M urea was recorded after excitation at 390 nm. Urea was used to unfold the protein, resulting in denaturation thereby exposing ANS to the solvent. Three accumulations were recorded for each reading, and an excitation and emission bandwidth of 5 mm was used. The spectra were corrected for buffer emission.

A competitive binding study was carried out to observe if ATP would displace ANS and bind to Rep. Fluorescence emission spectra of 100 μM ANS, and 5 μM Rep with 100 μM ANS excited at 390 nm were recorded. One hundred μM ATP was added to the Rep and ANS reaction and, after excitation at 390 nm, the emission spectrum was recorded. As before, three accumulations were recorded for each reading and an excitation and emission bandwidth of 5 mm was used. The spectra were corrected for buffer emission.

2.3.8.3 Far-UV circular dichroism

CD is a spectroscopic technique used to determine the secondary structural elements present in a protein of interest. The technique is described by the uneven absorbance of right and left hand polarised light by a compound. The chirality displayed by the α-carbon of amino acids contributes the unequal absorbance of polarised light, resulting in a CD spectrum. Different secondary structural features display specific CD spectra (Kelly et al., 2005).

CD readings were carried out by diluting protein into 0.1 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES). CD was carried out on Jasco J-810 spectropolarimeter (Jasco Inc., Easton, USA), using a 2 mm cuvette and 2 μM of protein. A bandwidth of 2 nm was used, and three accumulations were recorded.

The CD data was entered into the online server, Dichroweb which provides the user access to a number of software packages for analysing the CD data. The algorithm CONTIN and reference set SP175 was used to determine the percentage of secondary structural features. A

reconstructed spectrum was supplied and plotted alongside the original experimental data and an NRMSD was supplied (Lees et al., 2006; Provencher and Gloeckner, 1981; Van Stokkum et al., 1990).

2.3.8.4 In vitro cleavage reaction of Rep with nonanucleotide probe

In vitro cleavage reaction was carried out to determine the DNA binding and cleavage activity of recombinant Rep. A single-stranded nucleotide probe (SACMVN26V), as described in Campos-Olivas et al. (2002), was designed based on the SACMV DNA-A genome, and encompassed the conserved nonanucleotide sequence:

- SACMVN26V - 5′ - TCTAATATTACCGGATGGCCGCGCCC - 3′

To make a double-stranded probe, the probe was annealed with the complementary sequence (SACMVN26C) using an annealing buffer made of 10 mM Tris (pH 7.0), 1 mM EDTA and 50 mM NaCl. The complementary strands were incubated at 95 ° C for 5 min, and cooled at room temperature for up to 60 min. A negative control of pBlueScript (Novagen, Merck Milipore, USA) (see Appendix) was included to confirm that DNA binding was specific. A DNA-only and protein-only control was included. The type of probe and concentration of DNA and protein used in each reaction is shown in Table 2.4. The cleavage buffer was adapted from Kittelmann et al. (2009) and was composed of 25 mM Tris (pH 7.0), 300 mM NaCl, 5 mM $MgCl_2$ and 0.5 mM EDTA. Comparing the cleavage buffer used in this study, to that used by Kittelmann et al. (2009), a higher NaCl concentration was employed in this study so as to maintain the ionic strength present in the storage buffer of recombinant Rep. The reaction was terminated by adding 2× sample buffer and boiling at 90° C for 5 min. Samples were run on a 12% TGX FastCast™ SDS-PAGE gel (Bio-Rad, Hercules, USA) and stained using the Fairbanks method and visualised using the Gel Doc™ EZ imager (Bio-Rad, Hercules, USA). The Fairbanks method (Fairbanks et al., 1971) was used to visualise the bands (see Appendix). Western blotting was carried out on 12% SDS-PAGE gels using methods as in Section 2.3.6.

Table 2.4 *In vitro* DNA binding and cleavage assay of Rep and nucleotide probe.Different DNA probe types are shown, as well as concentrations of DNA probe and immobilised metal ion affinity chromatography (IMAC)-purified South African cassava mosaic virus (SACMV) Rep protein.

Reaction	Nucleotide type	Probe	DNA (µg)	Rep protein (µg)
1	ssDNA, linear	SACMVN26V	1	N/A
2	N/A	N/A	N/A	1
3	ssDNA, linear	SACMVN26V	1	1
4	ssDNA, linear	SACMVN26V	1	2
5	ssDNA, linear	SACMVN26V	1	5
6	dsDNA, linear	SACMVN26V	1	1
7	dsDNA, circular	pBlueScript	1	1

Key: N/A, not applicable; Reactions are labelled 1 – 7, with the first two reactions acting as negative controls; Reactions 3 – 5 contained the ssDNA 26-nt probe, and 1 µg of DNA; Reaction 6 contained the dsDNA 26-nt probe and reaction 7 combined Rep with pBlueScript vector.

2.4 Results

2.4.1 Bioinformatics analysis of recombinant Rep

Bioinformatics tools were utilised to predict certain physiochemical characteristics, and structural features which aid in the primary characterisation of SACMV Rep.

The bioinformatics tool ExPASy ProtParam (Gasteiger et al., 2005) uses an input sequence and assigns a *pI* value, a predicted molecular weight (MW), amino acid composition, predicted extinction coefficient, grand average of hydropathicity (GRAVY) and an instability and aliphatic index. The sequence of recombinant Rep (see Appendix) was computed to yield a protein with a predicted MW of 45053.4 daltons, and a predicted *pI* of 7.81. The recombinant protein was predicted to be stable with an instability index value of 39.73. The aliphatic index computes the relative volume of the protein that is contributed by the aliphatic side chains such as leucine, valine, alanine and isoleucine. The aliphatic index is used to predict the thermostability of proteins based on studies conducted on thermophillic bacterial proteins (Ikai, 1980). The computed GRAVY score of -0.627 indicates that recombinant SACMV Rep was hydrophilic (Gasteiger et al., 2005; Kyte and Doolittle, 1982).

Secondary structure and the degree of disorder were predicted using Phyre2, a web-based portal (Kelley and Sternberg, 2009). The bioinformatics tools available through Phyre2 computed the overall percentage of each secondary structural feature, with α-helices contributing 31%, β-strands 21%. As much as 27% of the protein is predicted to be disordered (Kelley and Sternberg, 2009) (Figure 2.5).

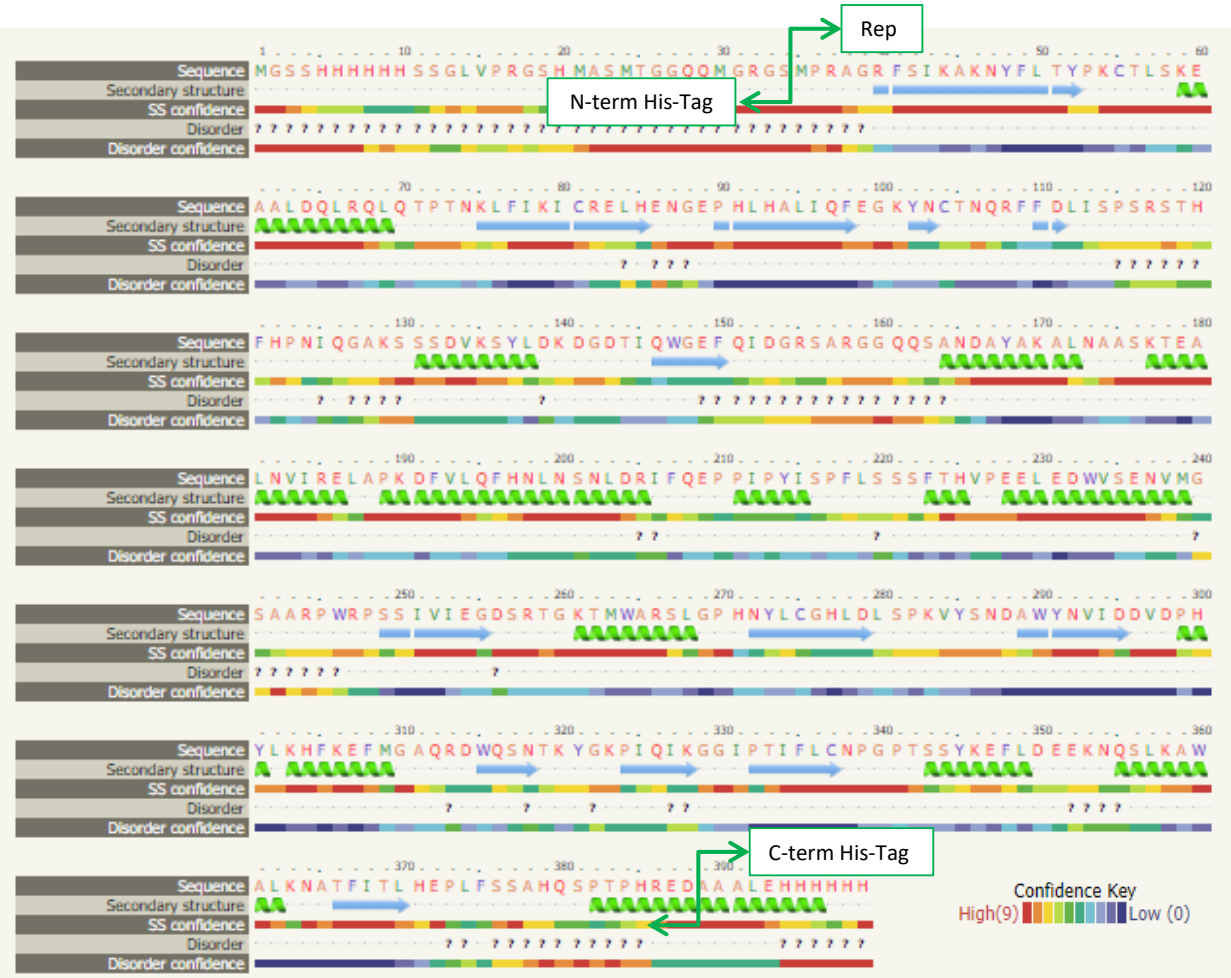

Figure 2.5 Web-portal output of secondary structure of SACMV Rep. PSI-PRED output showing secondary structure and disorder prediction with the degree of confidence of recombinant South African cassava mosaic virus (SACMV) Rep. The blue arrows indicate β-strands, and the green coils are representative of α-helices. The confidence score in secondary structure prediction is shown by the colour bar. The degree of disorder is shown through the use of question marks, with the colour bar below this, signifying the confidence in disorder prediction. The confidence key is shown (Jones, 1999). Location of N- and C-terminal His-Tags are shown by arrows, with the starting aa of SACMV Rep denoted.

Multiple sequence alignment of SACMV Rep with other begomovirus species was carried out to compare the presence and degree of conservation of the known motifs across the genus (Figure 2.6) (Sievers et al., 2011). Motifs contributing to DNA binding and cleavage catalysis (Motif I, II, III and GRS motif) are highly conserved across Rep proteins of begomoviruses. Walker A, Walker B, motif C and B′ are also shown to be conserved across Rep proteins of the genus *begomovirus*.

```
TGMV      MPSHPKRFQINAKNYFLTYPQCSLSKEESLSQLQALNTPINKKFIKICRELHEDGQPHLH    60
MBYMV     -MPRLGRFAINAKNYFLTYPRCPLRKEDALEELLALSTPVNKKFIRVCRELHEDGEPHLH    59
ToLCGuV   -MAAPNRFKINAKNYFLTYPKCSLTKEEALSQLLNLQTPTFKKFIRICRELHDDGTPHLH    59
SACMV     -MPRAGRFSIKAKNYFLTYPKCTLSKEAALDQLRQLQTPTNKLFIKICRELHENGEPHLH    59
TYLCSV    -MPRSGRFSIKAKNYFLTYPKCDLTKENALSQITNLQTPTNKLFIKICRELHENGEPHLH    59
                 ** *:*****:* * ** :*.::  *.**  * **::*****::* ****

TGMV      VLIQFEGKYCCQNQRFFDLVSPTRSAHFHPNIQRAKSSSDVKTYIDKDGDTLVWGEFQVD   120
MBYMV     VLLQFEGKFQTKNERFFDLVSSTRSAHYHPNVQAAKSASDVKSYMDKDGDVVDHGSFQVD   119
ToLCGuV   VLIQFEGKFQCKNNRFFDLTSPTRSAHFHPNIQGAKSSSDVKSYMEKDGDVIDHGVFQVD   119
SACMV     ALIQFEGKYNCTNQRFFDLISPSRSTHFHPNIQGAKSSSDVKSYLDKDGDTIQWGEFQID   119
TYLCSV    ILIQFEGKYNCTNQRFFDLVSPTRSAHFHPNIQGAKSSSDVKSYIDKDGDVLEWGTFQID   119
           *:*****:   *:***** * :**.*:***:* *** **** :*.:***:*.:  * **:*

TGMV      GRSARGGCQTSNDAAAEALNASSKEEALQIIREKIPEKYLFQFHNLNSNLDRIFDKTPEP   180
MBYMV     GRSARGGKQSANDAYAEALNSGSKLQALNILREKAPKDYILQFHNLNCNLSRIFSDEVPL   179
ToLCGuV   GRSARGGCQSANDAYAEAINSGSKAQALNILREKAPKDFVLQFHNLNSNLDRIFTPPMDV   179
SACMV     GRSARGGQQSANDAYAKALNAASKTEALNVIRELAPKDFVLQFHNLNSNLDRIFQEPPIP   179
TYLCSV    GRSARGGQQTANDAYAKAINAGSKSQALDVIKELAPRDYVLHFHNINSNLDKVFQVPPAP   179
          ******* *:.*** *:*:*:.** :**::::*  *..::::***:*.**.::*

TGMV      WLPPFHVSSFTNVPDEMRQWAENYFGKSSAARPERPISIIIGGDSRTGKTMWARSLGPHN   240
MBYMV     YVSPYSLSAFDKVPSYISSWASENVR-HPA-APERPISIVIGGDSRTGKTMWARALGPHN   237
ToLCGuV   YVSPFISSSFDQVPEELEAWAAENVC-SPAARPLRPISIVIGGDSRTGKTMWARSLGPHN   238
SACMV     YISPFLSSSFTHVPEELEDWVSENVM-GSAARPWRPSSIVIGGDSRTGKTMWARSLGPHN   238
TYLCSV    YVSPFLSSSFDQVPDELEHWVSENVM-DAAARPWRPVSIVIGGDSRTGKTTWARSLGPHN   238
          :: *:  *:* :**. :  *. : .    *  * ** **:*********:***:*****

TGMV      YLSGHLDLNSRVYSNKVEYNVIDDVTPQYLKLKHWKELIGAQRDWQTNCKYGKPVQIKGG   300
MBYMV     YLCGHLDLNSKIYSNDAWYNVIDDVDPHYL--KHFKEFMGAQRDWQSNVKYGKPTHIKGG   295
ToLCGuV   YLCGHLDLSPKVYSNDAWYNVIDDVDPHYL--KHFKEFMGAQRDWQSNTKYGKPVQIKGG   296
SACMV     YLCGHLDLSPKVYSNDAWYNVIDDVDPHYL--KHFKEFMGAQRDWQSNTKYGKPIQIKGG   296
TYLCSV    YLCGHLDLSQKVYSNNAWYNVIDDVDPHYL--KHFKEFMGAQRDWQSNTKYGKPIQIKGG   296
          **.*****. ::***.. ****** *:**  **:**::*******:* **:*** :****

TGMV      IPSIVLCNPGEGASYKVFLDKEENTPLKNWTFHNAKFVFLNSPLYQSSTQSS--------   352
MBYMV     IPTIFLCNPGPKSSYKEYLDEADNTALKLWASKNAEFYTLKEPLFSSVDQSATQGCQEAS   355
ToLCGuV   IPTTFLCNPGPNSSYKEYLDEEKNSALKNWALKNATFVTLEGPLYSSTNQSTAQASQEGD   356
SACMV     IPTIFLCNPGPTSSYKEFLDEEKNQSLKAWALKNATFITLHEPLFSSAHQSPTPHRED--   354
TYLCSV    IPTIFLCNPGPQSSFKEYLDEEKNQALKNWATKNAIFVTIHQPLFADTNQNTTSHRQEEA   356
          **  .*****  :*:* :**: .*  ** *: :** *  :. **: .  *.

TGMV      ------     352
MBYMV     NSTLSN     361
ToLCGuV   QTSTS-     361
SACMV     ------     354
TYLCSV    SEA---     359
```

Figure 2.6 Multiple sequence alignment of begomovirus Rep protein sequences using Clustal Omega (Sievers et al., 2011). Rep protein sequence of tomato golden mosaic virus (TGMV), Mung bean yellow mosaic India virus (MYMIV), Tomato leaf curl Gujarati virus (ToLCGuV), South African cassava mosaic virus (SACMV) and Tomato yellow leaf curl Sardinia virus (TYLCSV) were used. Motifs I, II, Geminivirus replication sequence (GRS) and motif III are framed in orange, respectively. Walker A, Walker B, B′ and motif C are enclosed in green, respectively.

2.4.2 Cloning SACMV AC1 ORF into pET-28a and co-transformation of *Escherichia coli* with pG-KJE8

SACMV AC1 was previously cloned into pJET1.2. To extract the AC1 gene of interest, the vector was digested using *Bam*HI and *Not*I. The destination vector, pET-28a was also linearised using *Not*I and *Bam*HI and both digestion reactions were run on a 1% agarose gel for gel extraction. Following gel extraction, the purified linearised pET-28a and AC1 were run on a 1% agarose gel to confirm their purity (Figure 2.7).

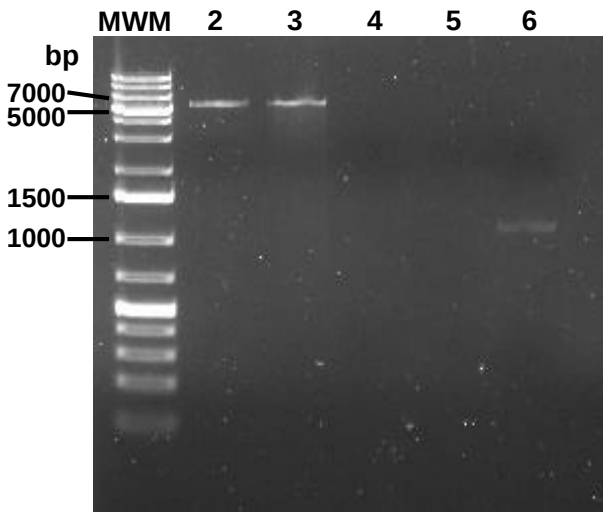

Figure 2.7 Linearised expression vector pET-28a and SACMV AC1 insert. Lane two and three show the linearised expression vector, pET-28a (5369 bp) following a restriction digest with Fast digest™ *Bam*HI and *Not*I (Thermo-Fisher Scientific, Waltham, USA) and a gel extraction procedure. Purified and concentrated insert (AC1) is shown in the sixth lane at 1065 bp. DNA standards are shown in lane one with the DNA molecular weight marker GeneRuler™ 1 kB Plus (Thermo Fisher Scientific, Waltham, USA). Lanes four and five are empty.

The recombinant vector was sequenced to verify the presence of AC1 and to confirm that no mutations had occurred. This was carried out using BLASTX on NCBI (Altschul et al., 1990). The query sequence displayed a 99% match to SACMV replicase (NP_620665.1) (Figure 2.8). As mentioned previously, the mutation F207S of AC1 is present in the infectious clone, pC8A, as well as in the SACMV-ZW variant and many other SACMV Rep variants.

```
Query  122   MPRAGRFSIKAKNYFLTYPKCTLSKEAALDQLRQLQTPTNKLFIKICRELHENGEPHLHA  301
             MPRAGRFSIKAKNYFLTYPKCTLSKEAALDQLRQLQTPTNKLFIKICRELHENGEPHLHA
Sbjct  1     MPRAGRFSIKAKNYFLTYPKCTLSKEAALDQLRQLQTPTNKLFIKICRELHENGEPHLHA  60

Query  302   LIQFEGKYNCTNQRFFDLISPSRSTHFHPNIQGAKSSSDVKSYLDKDGDTIQWGEFQIDG  481
             LIQFEGKYNCTNQRFFDLISPSRSTHFHPNIQGAKSSSDVKSYLDKDGDTIQWGEFQIDG
Sbjct  61    LIQFEGKYNCTNQRFFDLISPSRSTHFHPNIQGAKSSSDVKSYLDKDGDTIQWGEFQIDG  120

Query  482   RSARGGQQsandayakalnaasktealnvirelapkdfvlqfhnlnsnldrifqeppipy  661
             RSARGGQQSANDAYAKALNAASKTEALNVIRELAPKDFVLQFHNLNSNLDRIFQEPPIPY
Sbjct  121   RSARGGQQSANDAYAKALNAASKTEALNVIRELAPKDFVLQFHNLNSNLDRIFQEPPIPY  180

Query  662   ISPFLSSSFTHVPEELEDWVSENVMGSAARPWRPSSIVIEGDSRTGKTMWARSLGPHNYL  841
             ISPFLSSSFTHVPEELEDWVSENVMG AARPWRPSSIVIEGDSRTGKTMWARSLGPHNYL
Sbjct  181   ISPFLSSSFTHVPEELEDWVSENVMGFAARPWRPSSIVIEGDSRTGKTMWARSLGPHNYL  240

Query  842   CGHLDLSPKVYSNDAWYNVIDDVDPHYLKHFKEFMGAQRDWQSNTKYGKPIQIKGGIPTI  1021
             CGHLDLSPKVYSNDAWYNVIDDVDPHYLKHFKEFMGAQRDWQSNTKYGKPIQIKGGIPTI
Sbjct  241   CGHLDLSPKVYSNDAWYNVIDDVDPHYLKHFKEFMGAQRDWQSNTKYGKPIQIKGGIPTI  300

Query  1022  FLCNPGPTSSYKEFLDEEKNQSLKAWALKNATFITLHEPLFSSAHQSPTPHRED  1183
             FLCNPGPTSSYKEFLDEEKNQSLKAWALKNATFITLHEPLFSSAHQSPTPHRED
Sbjct  301   FLCNPGPTSSYKEFLDEEKNQSLKAWALKNATFITLHEPLFSSAHQSPTPHRED  354
```

Figure 2.8 Alignment of Rep and SACMV replicase amino acid sequences. BLASTX results of sequenced Rep in expression vector pET-28a, aligned with South African cassava mosaic virus (SACMV) replicase (accession number NP 620665.1) (Altschul et al., 1990).

After two consecutive transformation reactions, a restriction digest was carried out to ensure the presence of both vectors using the restriction enzyme Fast digest™ *Nco*I (Thermo Fisher Scientific, Waltham, USA). The vector pET-28a contains two *Nco*I cut sites (Figure 2.3), and pG-KJE8 contains three *Nco*I cut sites. Restriction digest of pET-28a/AC1 gave the expected banding pattern, with fragments of 1072 bp and 5337 bp. Restriction digest carried out on pG-KJE8 using *Nco*I resulted in three bands, 1236 bp, 1464 bp and 8495 bp. Restriction digestion carried out using both plasmids depicted the correct banding pattern (Figure 2.9).

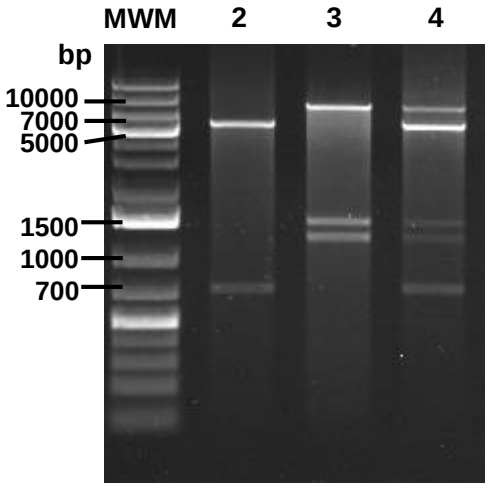

Figure 2.9 Restriction digest of pET-28/AC1, pGKJE8, and both pET-28a/AC1 and pG-KJE8. Lane two, three and four depict restriction digests carried out using Fast digest™ *Nco*I (Thermo-Fisher Scientific, Waltham, USA) of pET-28a/AC1, pG-KJE8, and pG-KJE8-pET-28a/AC1, respectively. DNA standards are shown in the first lane with the DNA molecular weight marker GeneRuler™ 1 kB Plus (Thermo Fisher Scientific, Waltham, USA).

2.4.3 Expression trials and large scale IMAC purification

Following confirmation of the presence of both recombinant expression and chaperone vectors, expression trials were carried out. Expression trials confirmed that expression of soluble recombinant Rep from pET-28a, alongside expression of chaperones, was achieved at 18° C for a period of 18 h. Recombinant Rep expression from pET-28a was carried out in the absence of chaperones as a control (Figure 2.10).

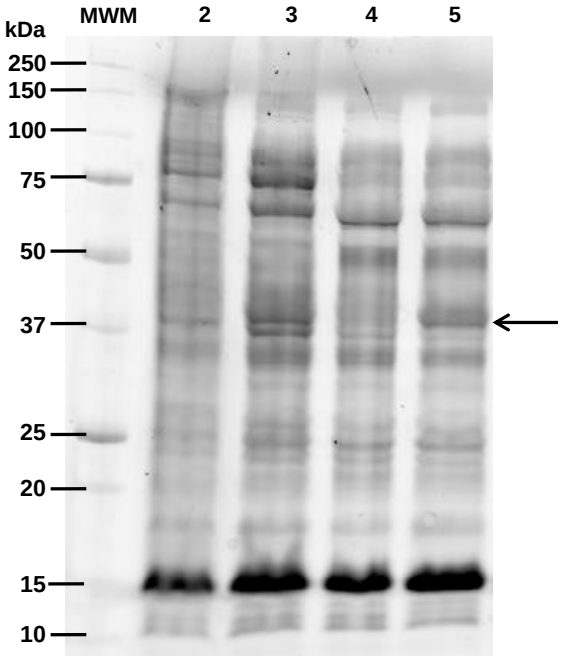

Figure 2.10 Expression trials of recombinant SACMV Rep. Expression trials were conducted at 18° C for 18 h. A sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) shows the different samples, run under reducing conditions. Lanes two and three show the protein expression profile of *Escherichia coli* BL21 (DE3) pET-28a/AC1 un-induced and induced lysate, respectively. Lane four and five represent the protein expression profile of *E. coli* BL21 (DE3) pET-28a/AC1 with the chaperone vector pG-KJE8, un-induced, and the sample induced with 0.3 mM IPTG, respectively. The arrow indicates the presence of recombinant SACMV Rep in the induced sample. The first lane shows the protein standards, Precision Plus Protein™ Unstained Protein Standards (Bio-Rad, Hercules, USA), with the sizes of the standards alongside the bands.

Following large-scale expression of recombinant Rep after induction with 0.3 mM IPTG, IMAC purification was carried out to obtain a pure fraction of the protein. The ÄKTA (GE Healthcare, USA) fast protein liquid chromatography (FPLC) purification system and Ni^{2+} charged FF HisTrap column (GE Healthcare, USA) facilitates purification on a large scale, enabling the processing of significant volumes of soluble fraction.

Step-wise elution was carried out by applying an imidazole concentration of 0-1M to the column. The first peak as shown on Figure 2.11 A is representative of the flow through (FT), with the second peak signifying the wash fraction, a fraction obtained after the column was exposed to 5-10 mM imidazole. The third distinctive peak and the shoulder of the peak, E1 and E2, respectively (Figure 2.11 B) were identified as a combination of the protein of interest and a multitude of other proteins of different molecular weights (Figure 2.12). Increasing the percentage of elution buffer and exposing the column to imidazole

concentration between 550 – 700 mM, resulted in the elution of the protein of interest (Figure 2.11 B and Figure 2.12).

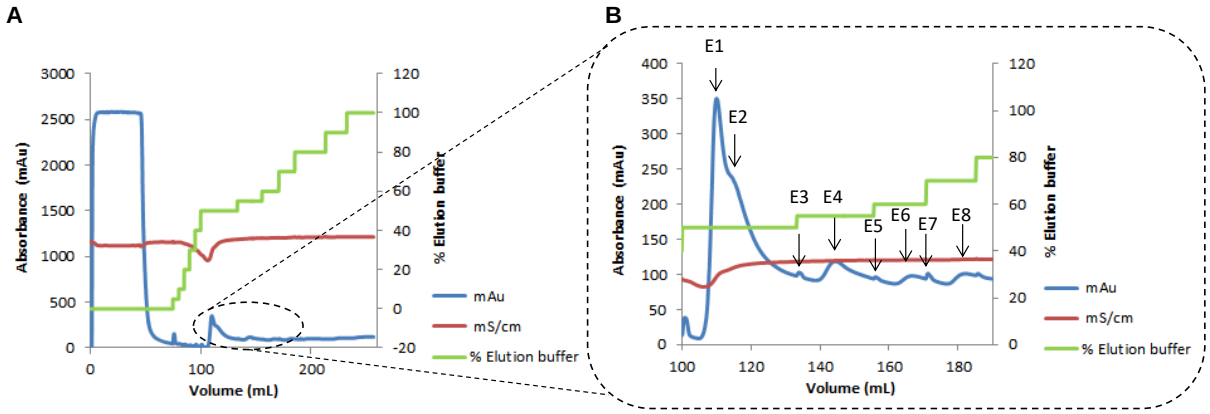

Figure 2.11 Purification profile of recombinant expressed Rep. Profile was obtained from ÄKTA and PrimeView 1.0 software (GE Healthcare, USA) following a nickel ion (Ni^{2+}) immobilised metal ion affinity chromatography (IMAC) purification of recombinant expressed Rep. **A** Purification profile showing absorbance readings in blue and corresponding conductivity readings in red. The green line represents % of elution buffer, which is an indication of the increasing concentration of imidazole being applied to the column. **B** Enlarged view of dashed circle in **A**. Elution fractions E1-E8 are indicated on the purification profile.

Cell lysate, a soluble fraction sample as well as purification samples were collected and electrophoresed on a 12% SDS-PAGE gel (Figure 2.12). Purification samples included fractions from the flow through, the wash step, and eight fractions collected that correlate with the peaks observed during elution of the protein of interest (Figure 2.11 B).

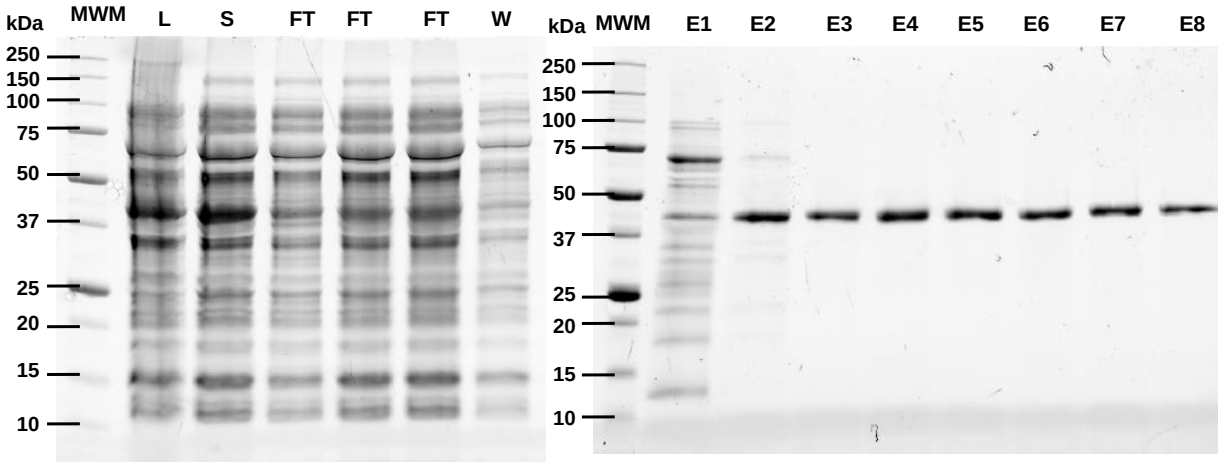

Figure 2.12 SDS-PAGE showing fractions collected during IMAC purification. 12% Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) showing the fractions collected from the nickel ion (Ni^{2+}) immobilised metal ion affinity chromatography (IMAC) purification (GE Healthcare, USA) run under reducing conditions. L represent the lysate, S represents the soluble fraction, FT represents the flow through and W is the wash fraction. Fractions labelled 'E' depicts the eluted fractions collected over the imidazole concentration range of 550 – 700 mM (E2-E8). The encircled region shown in Figure 2.11 B is represented by the bands as shown in E1-E8. The first lane of both gels shows the molecular weight marker (MWM), Precision Plus Protein™ Unstained Protein Standards (Bio-Rad, Hercules, USA), with the sizes of the standards alongside the bands.

The purified samples were analysed to determine purity and calculate MW. The gel showed that the purified protein sample did not contain contaminants, and the MW was calculated to be 44.9 kDa (Figure 2.13). The calculated and predicted molecular weight of recombinant Rep differs by 0.1 kDa.

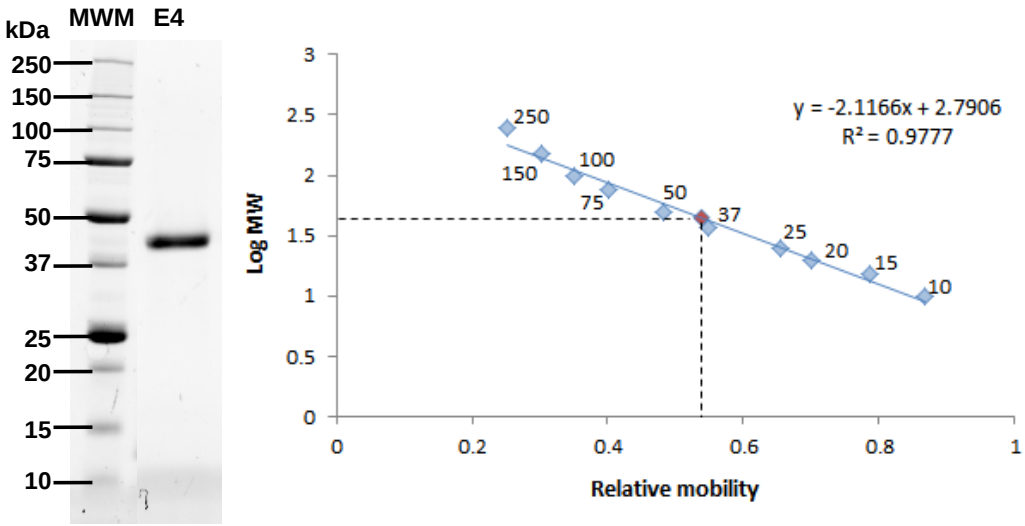

Figure 2.13 12% SDS-PAGE gel showing pure fraction of recombinant SACMV Rep, and the standard curve used to calculate the molecular weight of SACMV recombinant Rep. Purified Rep (E4 from Figure 2.12 B) is shown alongside the molecular weight marker (MWM) (Precision Plus Protein™ Unstained Protein Standards (Bio-Rad, Hercules, USA)). Using the known standards from the MWM, a standard curve was constructed. Standard curve equation was used to calculate molecular weight of Rep (red diamond) using a relative mobility of 0.538. Molecular weight standards are shown on the standard curve and the SDS-PAGE gel was run under reducing conditions.

2.4.4 Concentration determination of pure SACMV Rep protein

The concentration of purified Rep protein sample was determined by measuring absorbance at 280 nm for a range of diluted samples using a Jasco V-630 spectrophotometer (Jasco Inc., Easton, USA). The diluted sample spectra were used to generate a standard curve (Figure 2.14 A.). Using this and Eq. 1, the concentration of Rep was determined. A full spectrum of undiluted Rep was also carried out (Figure 2.14 B). The A_{280}/A_{260} ratio was calculated to be 1.68 suggesting negligible DNA contamination.

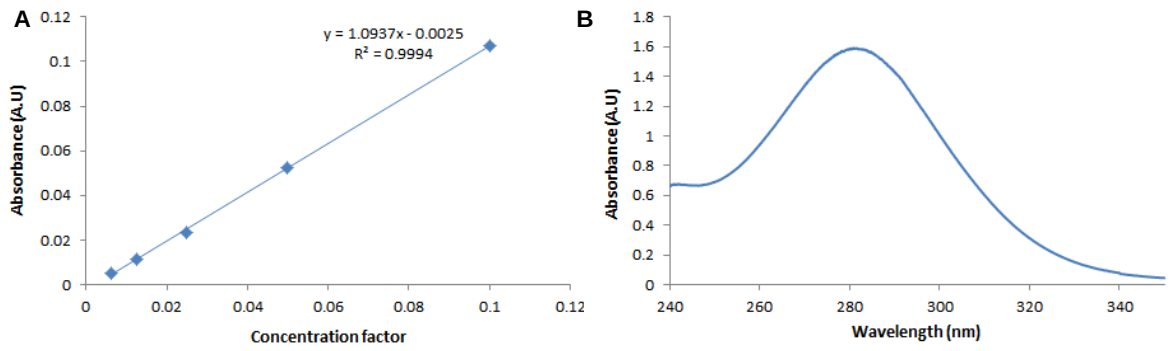

Figure 2.14 Standard curve of absorbance of diluted fractions, and UV/Vis spectrum of recombinant SACMV Rep. A standard curve shows absorbance of respective diluted fractions. The straight line function assists in ascertaining the concentration of Rep. **B** UV/Vis spectrum of Rep shows high absorbance at 280 nm. Some absorbance is evident at 340 nm, indicating presence of aggregates.

Expression of a His-Tagged recombinant Rep was confirmed by a western blot using an His-probe (H3) monoclonal antibody (Santa Cruz Biotechnology, USA) (Figure 2.15 B). The anti-His antibody binds to the His-Tag located on the N- and C-terminus of recombinant Rep, and suggests the presence of recombinant His-Tag Rep.

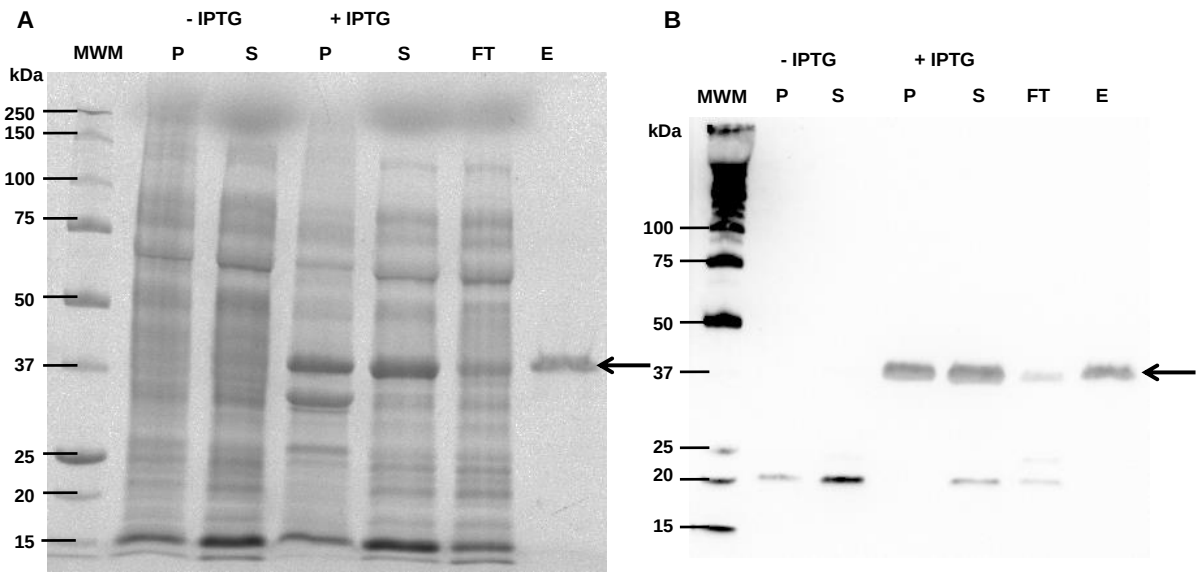

Figure 2.15 12% SDS-PAGE and western blot showing expression profile of recombinant Rep. **A** Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) gel shows the expression profile of *Escherichia coli* BL21 (DE3) pG-KJE8 pET-28a/AC1, and fractions following purification on nickel ion (Ni^{2+}) charged immobilised metal ion affinity chromatography (IMAC). The second and third lane represents the un-induced (-IPTG) pellet (P) and soluble (S) fraction, with lanes four and five displaying the P and S fraction induced with Isopropyl β-D-1-thiogalactopyranoside (+IPTG). Lanes six and seven represent the flow through (FT) of a purification carried out on the soluble induced fraction, as well as an eluent (E), showing a pure fraction of Rep. The SDS-PAGE was run under reducing conditions. **B** Western blot performed on 12% SDS-PAGE gel (**A**), using His-probe (H3) monoclonal antibody (Santa Cruz Biotechnology, USA). Presence of Rep was detected in pellet and soluble induced fractions, as well as in FT and E. The first lanes of **A** and **B** show the protein standards [Precision Plus Protein™ Unstained Protein Standards (Bio-Rad, Hercules, USA)], with the sizes of the standards alongside the bands.

2.4.5 Structural and functional characterisation of recombinant Rep

Structural characterisation of Rep was carried out using spectroscopic techniques. These included intrinsic fluorescence, extrinsic fluorescence and CD. Functional characterisation was completed to confirm functional activity of the protein. The activity assay investigated the cleavage activity of Rep on a conserved nucleotide fragment of the SACMV viral genome.

2.4.5.1 Intrinsic fluorescence of Rep

The intrinsic tryptophan fluorescence of Rep was analysed by exciting the residues at 290 nm. Excitation wavelength of 290 nm excited both tyrosine and tryptophan residues, with fluorescence of predominantly tryptophan noted at approximately 350 nm and tyrosine contributing weak fluorescence at 350 nm (Lakowicz, 2006). Addition of 100 μM ATP did

not lead to a hypso- or bathochromic shift, instead a small increase in fluorescence quantum yield was noted (Figure 2.16).

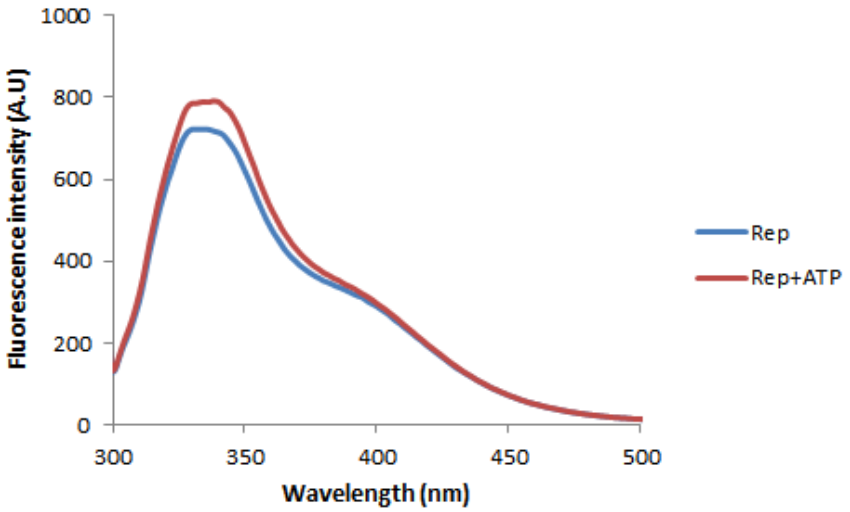

Figure 2.16 Intrinsic tryptophan fluorescence of SACMV Rep in the absence and presence of ATP.The intrinsic fluorescence of the recombinant South African cassava mosaic virus (SACMV) Rep was measured in the absence and presence of Adenosine triphosphate (ATP). An excitation wavelength of 290 nm was used and spectra were corrected for buffer emission.

2.4.5.2 Extrinsic fluorescence of Rep using ANS

ANS was used to show the presence of hydrophobic pockets on the surface of Rep. ANS is excited at 390 nm, and shows weak fluorescence in solvent. Binding of ANS to a hydrophobic pocket on Rep resulted in a hypsochromic shift (blue shift) and an increase in the quantum yield. Addition of urea, a denaturant, resulted in a fluorescence peak of approximately 510 nm. This was indicative of free (unbound) ANS. A strong decrease in fluorescence intensity was also observed (Figure 2.17).

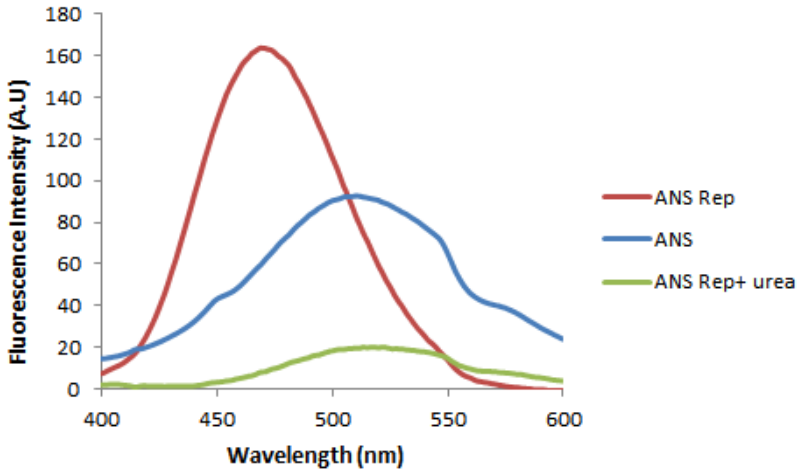

Figure 2.17 Extrinsic fluorescence spectrum of recombinant SACMV Rep in the absence and presence of ANS. The blue line shows the emission spectrum of 8-Anilino-1-naphthalenesulfonic acid (ANS) after excitation at 390 nm. The red line shows emission peak of Rep in the presence of ANS after excitation at 390 nm. The green line shows the emission spectrum of recombinant SACMV Rep in the presence of ANS and 8 M urea following excitation of ANS at 390 nm.

A competitive binding assay was carried out to determine if ANS binds to Rep at the predicted ATP binding site. Free ANS, as seen in Figure 2.17, was weakly fluorescent when excited at 390 nm. In the presence of Rep, the fluorescent signal exhibited a blue shift and an emission maximum at approximately 470 nm. After the addition of ATP, the signal did not change significantly, although a small red shift and increase in quantum yield of the emission peak was noted (Figure 2.18).

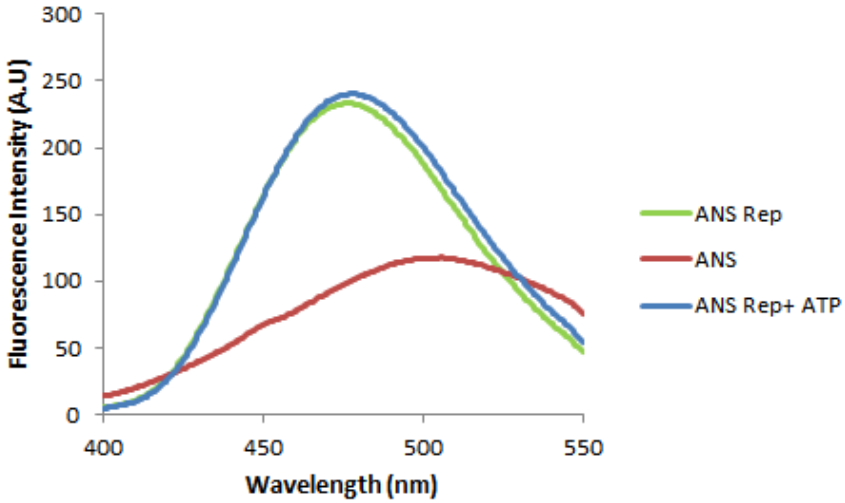

Figure 2.18 Extrinsic fluorescence of ANS binding to SACMV Rep in the presence and absence of ATP. Competitive binding assay of recombinant South African cassava mosaic virus (SACMV) Rep with 8-Anilino-1-naphthalenesulfonic acid (ANS) and Adenosine triphosphate (ATP). The red line is indicative of ANS after excitation at 390 nm. The blue and green line represents ANS Rep and the binding competitor ATP, and ANS Rep, respectively after excitation at 390 nm.

2.4.5.3 Far-UV circular dichroism

Far-UV CD was used to determine the secondary structural features of the protein of interest (Figure 19, red line). The spectrum most resembled a type II β-turn spectrum with a minor trough seen at 235 nm and a distinctive peak present at approximately 205 nm (Brahms et al., 1977; Perczel and Fasman, 1992; Perczel and Hollosi, 1996). Using the online analysis tool, DichroWeb and the CONTIN algorithm with reference set SP175, the percentage of secondary structure features was calculated. The reconstructed spectrum displayed a distinctive peak at around 203 nm, and to minor peaks present in the region of 191 nm and 223 nm. A trough is noted at approximately 231 nm. A normalised root means square displacement was calculated to be 0.356 and a reconstructed CD spectrum was computed (Figure 2.19, blue line).

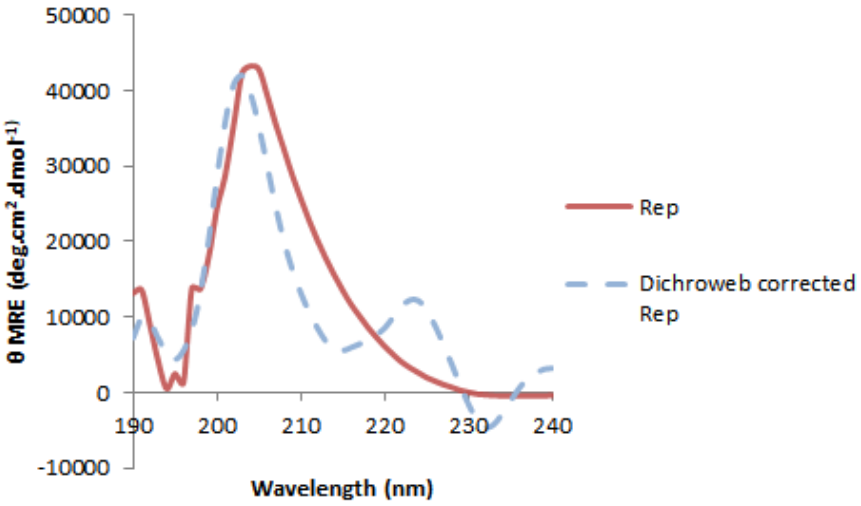

Figure 2.19 Far-UV CD spectrum of corrected and experimental recombinant SACMV Rep. Far-UV circular dichroism spectrum of recombinant South African cassava mosaic virus (SACMV) Rep, experimental and reconstructed. Far-UV CD spectrum of Rep, red line, does not resemble common far-UV CD spectra. Dichroweb reconstructed data results in a spectrum shown by the blue dotted line (Lees et al., 2006; Provencher and Gloeckner, 1981; van Stokkum et al., 1990).

2.4.5.4 In vitro cleavage assay

Geminivirus Rep is known to bind to the conserved nonanucleotide sequence situated within the IR of the viral genome and initiate viral replication through site-specific cleavage activity. The cleavage assay was carried out to confirm the DNA-binding and cleavage activity of recombinant Rep. DNA and protein only standards were included. Cleavage reactions were carried out with cleavage substrate (SACMVN26V) and increasing amounts of recombinant expressed Rep (Table 2.4). The probe encompassed the nonanucleotide with the conserved

cleavage site between nucleotide positions nine and ten. The probe was calculated to have a molecular weight of ~8 kDa. The fragment resulting from cleavage of the probe at the conserved cleavage site had a molecular weight of 5.2 kDa (Stothard, 2000). Rep-ssDNA reactions resolved on a 12% SDS-PAGE gel showed a protein-DNA adduct, calculated to be 48.4 kDa using the standard curve (Figure 2.20 A and C). No binding of Rep was shown with the negative control, pBlueScript, and with ds SACMVN26V. Western blots confirmed the presence of His-Tagged recombinant Rep as a free protein, and as bound to the cleavage product (Figure 2.20 B).

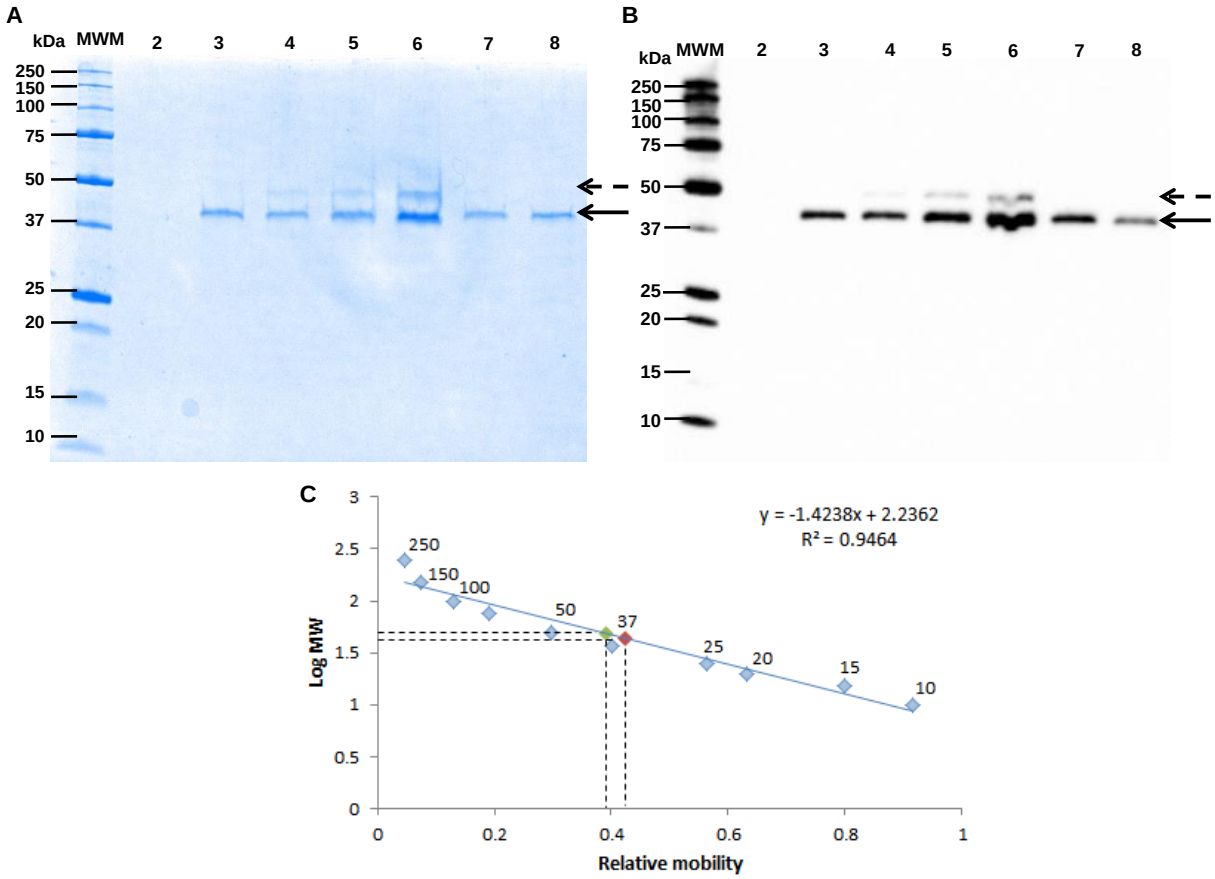

Figure 2.20 ***In vitro*** **cleavage assay of SACMV Rep with conserved nonanucleotide probe.** A Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) gel, run under reducing conditions, showing cleavage assay reaction using nickel ion (Ni^{2+}) charged immobilised metal ion affinity chromatography (IMAC)-purified recombinant South African cassava mosaic virus (SACMV) Rep and nucleotide probe SACMVN26V (**A**) and corresponding western blot (**B**). Cleavage reaction resolved on 12% TGX SDS-PAGE (Bio-Rad, Hercules, USA) gel (**A**) and western blot of PAGE gel (**B**) using His-probe (H3) monoclonal antibody (Santa Cruz Biotechnology, USA). Lanes two and three depict the DNA and protein only controls. Lanes four-six show the cleavage reaction of Rep with 26-mer ssDNA, with protein concentration of 1 µg, 2µg and 5 µg and DNA concentration of 1 µg. Lanes 7 and 8 show reaction of 1 µg Rep with 1 µg 26-mer dsDNA cleavage substrate, and 1 µg of negative control pBlueScript plasmid, respectively. Free Rep is indicated by the filled arrow, Rep-DNA adduct is indicated by the dashed arrow. Western blot shows binding of primary antibody to free protein, and protein-DNA adduct. **C** Standard curve of protein markers is used to calculate molecular weight of Rep (red diamond), and Rep bound to cleavage product (green diamond). Free protein (lane 6 in **A**) is shown at relative mobility of 0.42 and protein-DNA adduct of 0.39, with a calculated MW of 43.5 kDa and 48.4 kDa, respectively. Molecular weight of protein standards is shown on the graph (**C**) and on the gels (**A**) and (**B**). MWM in the first lane of **A** and **B** is Precision Plus Protein™ Unstained Protein Standards (Bio-Rad, Hercules, USA).

2.5 Discussion

The aim of this study, to express SACMV Rep and carry out structural and functional characterisation of the recombinant protein, was successfully accomplished. Expression of SACMV Rep in the soluble fraction, and purification using IMAC was carried out to completion. Some tertiary structural features were established, and preliminary results suggest DNA-binding activities of SACMV Rep. A number of begomovirus Rep proteins have been expressed, purified and studied to understand key functional characteristics (Campos-Olivas et al., 2002; Choudhury et al., 2006; Desbiez et al., 1995; George et al., 2014; Orozco and Hanley-Bowdoin, 1998; Singh et al., 2008). The structure and function of SACMV Rep; however, had not been investigated prior to this study.

Preliminary steps of the study included a bioinformatics analysis of recombinant Rep. The bioinformatics analysis was carried out on the recombinant expressed Rep sequence, as expression from pET-28a results in N- and C-terminal His-Tags being added to the native protein. The recombinant protein contains 45 additional amino acids (residues of tags and spacer), which was computed to add approximately 4.84 kDa to native SACMV Rep (40.2 kDa), with a final MW of 45.04 kDa (Gasteiger et al., 2005). The predicted *pI* and molecular weight (MW) of recombinant Rep were used to assist in designing experiments and in identifying the overexpressed protein on SDS-PAGE gels. The computed instability index of recombinant Rep predicted a stable protein, in contrast to the predicted instability of native Rep (not shown).

The amino acid sequence alignment of SACMV Rep with studied geminivirus Rep proteins was carried out to confirm the presence of the conserved motifs across the different viral species. Motifs involved in DNA binding and cleavage activity, as well as helicase and ATPase-linked motifs were shown to be highly conserved across the aligned proteins (Figure 2.6). This gives an indication of the conserved nature of the motifs involved in Rep-related activities across the genus *Begomovirus* (Hanley-Bowdoin et al., 2004). More specifically, the catalytic tyrosine residue found in motif III is highly conserved. Walker-A, the NTP binding motif, is also highly conserved across the genus (Figure 2.6). The conserved nature of Rep protein motifs suggests that SACMV Rep would display functional activities and structural characteristics as observed in other begomovirus Rep proteins. This study confirmed some of these aspects.

Cloning of the sequence of interest was performed successfully, as restriction digests showed the correct banding pattern of pET-28a/AC1 and pG-KJE8 (Figure 2.9). Sequencing of the recombinant pET-28a confirmed that no other mutations had been introduced into the nucleotide sequence (Figure 2.8).

Protein expression of recombinant SACMV Rep was carried out in *E. coli* which is the preferred expression host due its economic advantages, as well as its fast growth. Furthermore multiple expression vectors have been designed to facilitate expression of recombinant proteins in this host (Chen and Texada, 2006). Eukaryotic expression hosts, such as yeast, are more suitable for the expression of proteins that require post-translational modifications (Rosano and Ceccarelli, 2014). A multitude of geminivirus Rep proteins have been expressed in *E. coli*, an expression host lacking post-translational mechanisms, to carry out protein binding studies, as well as functional and structural studies (Bagewadi et al., 2004; Campos-Olivas et al., 2002; Castillo et al., 2004; Choudhury et al., 2006; Desbiez et al., 1995; Kushwaha et al., 2017; Laufs et al., 1995a; Pant et al., 2001).

The expression of a soluble, recombinant protein was made possible through the addition of the chaperone plasmid, pG-KJE8. Chaperones DnaK, DnaJ, GroEL, GroES and GrpE were expressed under the inducers L-arabinose and tetracycline. The addition of 2 $mg.mL^{-1}$ L-arabinose, 5 $ng.\mu L^{-1}$ tetracycline and protein expression at 18° C enabled expression of the protein of interest in the soluble fraction. Expression of recombinant Rep in other pET expression systems showed that optimum expression conditions required employing chaperones and expression at 18° C. In the absence of chaperone expression in this study, the recombinant protein could not be identified in the soluble fraction, despite expression at lower temperatures and low IPTG concentrations (Figure 2.10). Due to expression at low temperatures only low concentrations of chaperones were required for soluble expression, with previous studies reporting that GroEL-GroES plays a more important role in the prevention of aggregation compared to DnaK-DnaJ-GrpE (Nishihara et al., 1998). In this study, both sets of chaperones were employed for expression of moderate levels of recombinant SACMV Rep protein in the soluble fraction (Figure 2.10). Future experiments could include confirming the presence of chaperones when expression is induced through the use of Western blots.

The pET expression vectors have previously been utilised for the expression of recombinant geminivirus Rep proteins. In this study, the addition of the His-Tag enabled the purification

of recombinant SACMV Rep. Using a nickel charged chromatography column, the His-tagged recombinant protein was immobilised on the matrix and was eluted from the column through the addition of imidazole (Figure 2.11 and Figure 2.12). The flow through, however, also contained recombinant protein of interest (Figure 2.12 and Figure 2.15). Residual recombinant SACMV Rep could potentially be recovered from the flow through fraction by running the collected fraction over the IMAC column again.

The presence of an expressed and purified His-Tagged protein was confirmed through a western blot using His probe (H3) primary antibody (Figure 2.15 B), predicted to be recombinant SACMV Rep. The MW of the recombinant protein was calculated to be 44.9 kDa (Figure 2.13), further suggesting the identity of the protein to be recombinant SACMV Rep. The calculated and predicted molecular weight of recombinant SACMV Rep differed by 0.1 kDa which may be attributed to the minor calculation errors incurred during the determination of relative mobility or due to discrepancies that can arise when calculating protein molecular weight from SDS-PAGE gels.

A definitive method of identification of the His-Tagged protein identified on the SDS-PAGE gel and on the Western Blot would be to carry out mass spectrometry on the purified protein band present in the SDS-PAGE gel (Wilm et al., 1996). Analysis of the sequence would either confirm or deny the identity of the protein as recombinant SACMV Rep. Future work could include raising an antibody against SACMV Rep, as per Kittelman et al. (2009) and Campos-Olivas et al. (2002), with the subsequent western blot proving the presence of recombinant SACMV Rep in expression profiles.

The His-Tag was not cleaved in previous geminivirus Rep protein studies as the tag is negligible in size. The His-Tag is predicted to not interfere with SACMV Rep activity or structure, as previous studies on His-Tag fusion geminivirus Rep proteins suggest tag interference is negligible, as structural and functional assays have been carried out on geminivirus Rep fusion proteins (Bagewadi et al., 2004; Choudhury et al., 2006; Kushwaha et al., 2017; Pant et al., 2001). Furthermore, numerous His-Tag protein crystal structures can be found in the protein database (PDB) (Terpe, 2003), which may suggest that the addition of an expression tag did not significantly alter the protein structure. The His-Tag is the preferred affinity tag in recombinant protein expression (Arnau et al., 2006), and can be used in a number of expression hosts, including yeast and mammalian cells (Terpe, 2003). Its contribution to solubilisation of recombinant proteins is minimal (Waugh, 2005). The

addition of a solubilising tag, such as MBP or TF, has been confirmed to contribute to correct folding *in vivo* during recombinant protein expression (Arnau et al., 2006). TF has been incorporated into the pCOLD^{TM}TF vector (Takara Bio Inc., Japan), where it facilitates correct folding of fusion proteins during expression (Baneyx, 1999).

Full UV-Vis spectrum analysis of recombinant Rep revealed negligible DNA contamination as well as the presence of a small amount of aggregates (Figure 2.14). This may have been due to the predisposition of the protein to aggregate in this study. The purified protein was very prone to aggregation and any further dialysis or concentrating efforts were not successful. Proteins are more likely to aggregate when at higher concentrations, especially at concentration levels required for structural analysis (Bondos and Bicknell, 2003). Difficulties in concentrating purified begomovirus Rep is not uncommon. Obtaining concentrations of ToLCGuV Rep required for crystallography trials have proven impossible, with the protein being very aggregation-prone (pers. comm. Prof Gourinath, Jawaharlal Nehru University, New Delhi, India). In a study analysing the helicase activity of TYLCSV Rep, the expression of full length Rep proved difficult due to the susceptibility of the protein to aggregate. Deletion of the N-terminal domain of TYLCSV Rep prevented aggregation and enabled functional studies of TYLCSV RepC to be completed (Clérot and Bernardi, 2006).

Future work will include looking at expression of SACMV RepC using the expression vector pCOLD^{TM} TF (Takara Bio Inc., Japan). The sequence will be codon optimised for increased protein expression and to ensure the correct amino acids are incorporated (Chen and Texada, 2006). Rare codons in a nucleotide sequence are defined as codons which have a low abundance of tRNA in a particular expression host. Their presence in sequences coding for recombinant-expressed proteins can result in reduced expression, truncation, or incorrect amino acids incorporated into the recombinant protein (Chen and Texada, 2006). The codon optimisation of the sequence will result in higher recombinant protein concentration yields in the soluble fraction (Chen and Texada, 2006). Using pCOLD^{TM}TF vector for expression may increase concentration of soluble and more stably-expressed recombinant protein and enable further analytical studies that require high concentrations of protein. Eliminating the N-terminal portion of SACMV Rep, to express a truncated recombinant protein may also contribute to a more stable protein. This truncated protein would likely not be aggregation-prone, as observed in studies of begomovirus RepC (Clérot and Bernardi, 2006; George et al., 2014).

The intrinsic fluorescence of SACMV Rep showed the excitation of tryptophan and tyrosine residues at 290 nm, and the emission maxima at approximately 350 nm (Figure 2.16). The addition of ATP was expected to alter this fluorescence activity, due to its ability to bind to Rep. ATP binding would have been indicated by a hypsochromic shift in fluorescence maximum and an increase in quantum yield due to binding of ATP to the NTP binding loop on the ATPase domain of Rep (Gorbalenya et al., 1990; Lakowicz, 2006). The lack of a shift in fluorescence upon interaction of Rep and ATP may be explained by the transient nature of the interaction (pers. comm. Prof Gourinath, Jawaharlal Nehru University, New Delhi, India). To confirm Rep-ATP binding, a large conformational change would have occurred, and been observed through intrinsic fluorescence. The binding of ATP to the protein may not have induced significant structural changes (Laconte et al., 2017), resulting in an intrinsic fluorescence spectrum as seen in Figure 2.16. The interaction of geminivirus Rep with ATP and ATPase activity has been confirmed with recombinant expressed TYLCV and ToLCGuV Rep proteins (Desbiez et al., 1995; George et al., 2014). ATPase assays monitor the release of free radiolabelled P_i, and not the binding interaction of Rep with ATP (Choudhury et al., 2006; Desbiez et al., 1995; George et al., 2014). The use of a non-hydrolysable ATP was used to confirm a binding interaction of ATP with TYLCV Rep (Desbiez et al., 1995).

Future work to confirm the binding of SACMV Rep to ATP could be done through the use of a non-hydrolysable ATP, such as 2′(3′)N-methylanthraniloyl-ATP (MANT-ATP). Binding of MANT-ATP to Rep would be analysed through the intrinsic fluorescence of Rep in the presence and absence of MANT-ATP. After excitation at 280 nm, free Rep would display the characteristic fluorescence emission of 350 nm due to the fluorescence of tryptophan and tyrosine residues (Lakowicz, 2006). The fluorescence spectrum of free MANT-ATP after excitation at 280 nm would produce a negligible emission maximum at 450 nm (Ni et al., 2000). Binding of MANT-ATP to Rep would give rise to the phenomenon of FRET. This is manifested through the transfer of energy from the excited tryptophan residues on the protein to the fluorophore of MANT-ATP, bound at the ATP binding site. FRET would result in excitation of MANT-ATP, and it displaying a fluorescence maximum at approximately 450 nm (Moore and Lohman, 1994). Other methods of monitoring ATP binding to Rep would include isothermal titration calorimetry (ITC). This method would analyse the energetics of Rep-ATP binding (Freire et al., 1990).

Future work could also include the structural analysis of Rep binding to DNA. This interaction would result in a fluorescence shift upon DNA binding to Rep after excitation of

tyrosine residues at 280 nm. The catalytic tyrosine, situated in a DNA binding pocket, binds to and cleaves the viral genome at the nonanucleotide sequence (Laufs, Schumacher, et al. 1995). Addition of protein to this DNA binding pocket would display a hypsochromic shift in the fluorescence spectrum and an increase in quantum intensity, thereby confirming DNA binding to Rep (Lakowicz, 2006).

The binding of the fluorescent probe ANS to SACMV Rep was carried out to characterise the potential binding pockets of Rep, as ANS is often used to probe for hydrophobic pockets on the protein surface (Gasymov and Glasgow, 2007). Free ANS displays almost negligible fluorescence intensity, and an emission peak at approximately 515 nm. The blue shift and increase in quantum yield of the fluorescence spectrum confirms the binding of Rep to ANS (Figure 2.17). ANS fluorescence is sensitive to its environment. Upon binding to the protein binding pocket, the environment of ANS shifted from a polar to a less polar environment, resulting in a shift of fluorescence maximum to approximately 470 nm and an increase in quantum yield (Figure 2.17) (Hawe et al., 2008; Stryer, 1965). The addition of urea to the reaction resulted in an emission spectrum mimicking that of free ANS. Urea acted as a denaturant and unfolded Rep, exposing ANS to the polar solvent. The fluorescence spectrum displayed a bathochromic shift, and decrease in fluorescence intensity (Figure 2.17). These results indicate the presence of hydrophobic cavities of SACMV Rep that could serve as binding sites for DNA, ATP or other plant host proteins (Matulis and Lovrien, 1998). The differing fluorescence spectra (Figure 2.17) confirm a conformational change upon ANS binding to SACMV Rep (Baumann et al., 1998).

The inclusion of ATP in the ANS-Rep reaction was carried out in this study to observe possible displacement of ANS by ATP. Binding of ATP was expected to displace bound ANS and disrupt binding, resulting in a red shift and decrease in quantum intensity of ANS emission. However, ANS bound to Rep was not displaced (Figure 2.18). As binding of Rep to ATP (Figure 2.16) was not evident, this may explain the lack of competitive binding that ATP exhibits. Future attempts to displace ANS present in the binding pocket of Rep by ATP would include the use of a non-hydrolysable ATP containing a fluorophore, such as MANT-ATP. Addition of MANT-ATP to Rep-ANS and excitation of ANS at 390 nm should result in the displacement of ANS from the binding pocket with non-transient ATP binding. In addition, a fluorescence spectrum of emission at 515 nm, and decreased quantum yield would be observed (Hawe et al., 2008; Stryer, 1965).

Elements of the secondary structure of SACMV Rep were predicted using a bioinformatics tool, PSI-PRED (Figure 2.5), and determined experimentally using far-UV CD (Figure 2.19). CD data was analysed through DichroWeb, using CONTIN, a secondary structure determination algorithm, and the reference set SP175. PSI-PRED predicted the recombinant protein to contain mostly α-helices, but also many disordered regions. The secondary structure prediction did not assign any structure to the N-terminal His-Tag, with isolated disordered regions being identified throughout the protein. (Campos-Olivas et al., 2002; Jones, 1999)

The far-UV CD spectrum (Fig 2.19, red line) most resembles the CD spectrum of a type II β-turn. Type II β-turns display a CD spectrum featuring a slight trough at 216 nm, and a peak at 196 nm, although other spectra of type II β turns do not display the trough, instead only displaying a peak at approximately 200 nm (Brahms et al., 1977; Perczel and Fasman, 1992; Perczel et al., 1991). The lack of type II β turn models may be due to the fact that β-turn models have been shown to adopt more than one conformation, making it difficult to obtain one model that would accurately represent β-turns (Perczel et al., 1991). The CD spectrum of Rep most resembles these spectra, as Figure 2.19 shows, with a slight trough between 230 and 240 nm, and a peak at approximately 205 nm.

Dichroweb analysis, using the CONTIN algorithm and SP175 reference set, identified a reference spectrum (Lees et al., 2006; Provencher and Glockner, 1981; Van Stokkum et al., 1990). The input far-UV CD data was reconstructed to best fit the reference spectrum (Figure 2.19, blue dotted line), generating a normalised root-mean-square deviation (NRMSD) of 0.35. Analysis of the CD data was carried out using the other available algorithms and reference sets, but the combination of CONTINLL and reference set SP175 yielded the lowest NRMSD. However, this value suggests that an error has occurred during the analysis (Kelly et al., 2005). The inability of DichroWeb to accurately predict the secondary structure of recombinant Rep could be attributed to the lack of reference spectra in the given databases, due to their unique nature or the poor quality of the input data (Whitmore and Wallace, 2008).

Another reason for the unexpected CD spectrum obtained in this study could be the dynamic nature of the protein. Viral proteins are very dynamic and disordered molecules (Chen et al., 2008) suggesting that the dynamic nature of Rep could potentially distort the far-UV CD spectrum. Far-UV CD analysis could be repeated using higher concentrations of protein, as

the low concentration of protein used may have resulted in a distorted spectrum. Using a higher concentration of protein, or avoiding high absorbing buffer additives, would result in more reliable spectral data (Kelly et al., 2005). However, due to the difficulties in concentrating or dialysing recombinant Rep, using a higher concentration of recombinant Rep may not be possible. Using difference expression systems, such as pCold^{TM}TF, may result in more stable protein that can be concentrated and used in far-UV CD analysis.

The experimental CD of SACMV Rep suggested the presence of a type II β-turn; however, the lack of these identified by bioinformatics tools did not enable comparisons to be made. The PSI-PRED predicted results may be a more reliable account of secondary structure features, due to the accuracy of the prediction when compared to secondary structures of the resolved 3D structure of TYLCSV Rep (Campos-Olivas et al., 2002). The secondary structures predicted for the first 180 amino acids of TYLCSV Rep share great similarity, in position and length, with the predicted secondary structure of SACMV Rep bar two β-sheets (Jones, 1999). The positioning of functionally crucial motifs of the catalytic domain on either β-sheets or α-helices is observed in SACMV Rep predicted secondary structure, as it is seen on TYLCSV Rep (Jones 1999; Campos-Olivas et al. 2002). A comparison of the predicted secondary structure of Rep, with the secondary structures of AAV Rep40 and SV40 L-Tag (determined through crystal structures) (James et al., 2003; Li et al., 2003) confirms a prevalence of β-strands. The β-strands of SF3 helicases form a β-sheet making up the ATPase core domain (Clérot and Bernardi, 2006). It must be noted that protein structure databases contain only one geminivirus Rep protein structure (PDB 1L2M) (Campos-Olivas et al., 2002), with no other homologues available in these databases. Therefore, structural features predicted using bioinformatics approaches are based on a limited amount of deposited protein structural data (pers. comm. Alexander Zwolinksi).

Functional activity of recombinant Rep was investigated by carrying out an in vitro cleavage assay. Binding of Rep to the probe was expected to result in a cleaved product, and the formation of a protein-DNA adduct (Campos-Olivas et al., 2002; Hipp et al., 2014; Kittelmann et al., 2009; Laufs et al., 1995a) (Figure 2.21). Preliminary results suggest SACMV recombinant Rep bound to and cleaved a 26-mer single- stranded DNA probe containing the conserved nonanucleotide. Further confirmation is required to confirm the cleavage event.

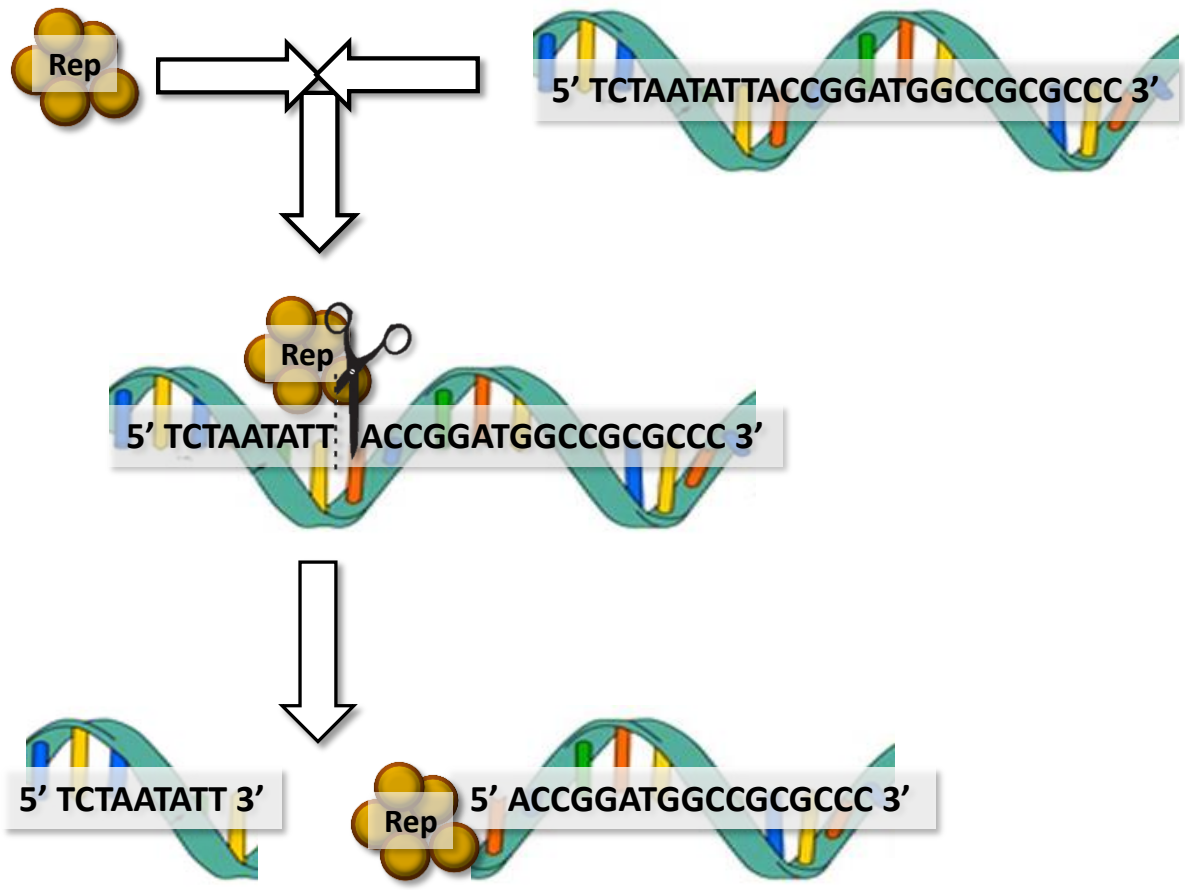

Figure 2.21 Schematic representation of DNA binding and cleavage activity of Rep with probe SACMVN26V. Upon binding to the conserved nonanucleotide sequence, Rep cleaves the probe at position 9, and binds to the 5′ end of the cleaved product, releasing the 3′- end.

The calculated molecular weight of the resultant adduct suggests cleavage at the 9th position on the probe by Rep and it can be suggested with confidence, as has been shown in previous research of geminivirus Rep proteins that Rep remained bound to the 5′ end of the cleaved product (Campos-Olivas et al., 2002; Hipp et al., 2014; Kittelmann et al., 2009; Laufs et al., 1995a). The cleavage reaction would have resulted in a DNA product of 19 nt. Calculations of the weight reveal a predicted DNA cleavage product with a molecular weight of approximately 5 kDa. The protein-DNA adduct in turn would be approximately 50 kDa in size. Due to the variation in SDS-PAGE gel and how the protein standards were resolved, the calculated molecular weights of free protein and protein-DNA adduct may not reflect the true molecular weight, as the calculated weight of free protein (Figure 2.20 C) differs to that calculated previously (Figure 2.13). However, the ~5 kDa difference in weight between free protein and protein-DNA adduct as observed by SDS-PAGE suggests that the cleaved DNA product is bound to Rep. The increase in molecular weight, due to the bound cleavage product, induces a shift as seen on the SDS-PAGE gel. The results of the western blot

confirm the identity of unbound Rep, and DNA-bound Rep (Figure 2.20 A and B). However, further evidence would need to be obtained to verify the cleavage of the nonanucleotide ssDNA probe by recombinant SACMV Rep. The unbound DNA could serve as evidence for the cleavage reaction, therefore future work would include a Southern Blot to confirm the presence of the 3′- end of the cleaved product (unbound product) (~3 kDa) (Southern, 1975).

The lack of binding of Rep to the double-stranded DNA probe could be explained by a transient interaction between Rep and the dsDNA probe. Rep may have bound to and cleaved the dsDNA probe, but SDS-PAGE conditions were not conducive to a Rep-dsDNA probe adduct. In order to confirm that Rep did not cleave dsDNA, the integrity of the dsDNA probe would have to be investigated.

Contrastingly, the lack of a protein-dsDNA adduct may be rationalised by the absence of iterons and the inability of Rep to carry out cleavage activity on a double-stranded template. Geminivirus-specific iterons in the IR confer virus specificity, and are crucial for initial binding of Rep to dsDNA (Singh et al., 2008). The single stranded nature of the probe enabled the cleavage thereof; however, a double-stranded nonanucleotide is known to not be cleaved by Rep (Singh et al., 2008). If analysing the cleavage activities of Rep on the double-stranded probe, it can be proposed that the iterons found to the left of the stem-loop are required for formation of the stem-loop. The consequent single stranded nature of the stem-loop would enable Rep induced cleavage to take place (Singh et al., 2008).

The His-Tagged recombinant SACMV Rep protein expressed in the soluble fraction in *E. coli* with the help of chaperones may be indicative of a structurally competent recombinant protein. The results obtained from structural and enzymatic experiments suggest that the presence of the His-Tag did not affect the native folding of recombinant SACMV Rep or interfere with enzymatic activity and thus His-Tag cleavage was not necessary. DNA binding and cleavage assays, as performed in this study, have been carried out to confirm the enzymatic activity using other recombinant-expressed begomovirus Rep proteins (Campos-Olivas et al., 2002; Hipp et al., 2016; Kittelmann et al., 2009; Laufs et al., 1995a). The preliminary results of the experiments of this study, that analysed protein-DNA binding and cleavage activity, indicate recombinant SACMV Rep, which was successfully purified using IMAC, is likely to be folded in its native conformation as it displays enzymatic functionality. This strongly suggests that the results of the biophysical analysis obtained herein are likely to be reliable in their depiction of viral protein molecular structure.

2.6 Conclusion

SACMV Rep was successfully expressed as a recombinant protein in an *E. coli* expression system, and purified making use of His-Tags and Ni^{2+} charged IMAC methods. The presence of recombinant expressed protein was probed using western blotting, and the purified protein was quantified. The structural characteristics of recombinant Rep were probed, and revealed hydrophobic pockets that could serve as binding sites for ATP, DNA or other plant host cell protein binding partners. Secondary structure was investigated *in vitro*, but yielded unsatisfactory data and no definitive conclusions could be made in relation to what the secondary structural features of SACMV Rep are. However, bioinformatics analysis revealed the similarities of Rep secondary structures across begomovirus species and helicase proteins of SF3. Functional assays proposed that recombinant Rep possessed enzymatic ability, namely the DNA binding and cleavage activities characteristic of all geminiviruses (Hanley-Bowdoin et al., 2013; Laufs et al., 1995a).

The preliminary structural and functional characterisation of SACMV Rep was achieved but further studies need to be undertaken to fully characterise the secondary and tertiary structure of the protein. Furthermore, the quaternary and 3D structure of partial or full-length geminivirus Rep proteins can be elucidated by nuclear magnetic resonance (NMR), X-ray crystallography or cryo-electron microscopy. Functional assays will further aid in determining if SACMV Rep activities and interaction with host proteins differ from those of other members of the genus, and further analysis may lead to fully characterising the structure of Rep, and the role of the protein structure in host interactions.

Chapter III – Elucidation of protein binding partners of SACMV Rep in cassava

3.1 Abstract

Functional genomics has used the vast amounts of sequencing data to annotate genes, and more recently has delved into using proteomics and protein-protein interaction (PPIs) data to assign function to unknown genes. Studying PPIs is crucial to further understanding the complex systems that are involved in plant-pathogen interactions thereby expanding the relevant interactome network, which for many plant-pathogen interactions is still not fully understood. The yeast two-hybrid (YTH) assay is a straight forward and easy to use technique that aims to reveal direct PPIs between a chosen bait and prey protein. This study aimed to identify putative cassava proteins that would act as SACMV Rep binding partners. The interactions were confirmed using an X-gal based YTH assay. Three prey proteins were chosen based on previous research and suggestions of an interaction with Rep in literature. Histone H3, an interaction candidate was shown to interact with TGMV Rep in previous studies. The other candidate interactors, cyclin D3;2 and cyclin D4;2 were selected due to the role that Rep plays during the host cell cycle, as well as a cyclin binding motif present within the sequence of SACMV Rep. The putative binding partners served as prey proteins, with SACMV Rep selected as the bait protein. YTH assay suggested that there was not a direct interaction between SACMV Rep and the chosen prey proteins, with future research seeking to confirm the formation of a multi-protein binding complex with the candidate interactors and SACMV Rep.

3.2 Introduction

3.2.1 Proteomics and its role in functional genomics

Functional genomics is a field of research that aims to ascertain gene function (Clara, 1999; Holtorf et al., 2002). The proteome encompasses all proteins present within the lifespan of a cell or tissue, the structures therein, their interactions, higher order complexes as well as interactions with other macromolecules (Honoré et al., 2004). The genomics era has provided whole genome sequencing and high throughput RNA sequencing data which contributed to the analysis of thousands of genes (Auerbach et al., 2002). In the post-genomics era, research has used the vast amounts of genomic and proteomic data, and protein-protein interactions (PPIs), to determine the function of the genes and assist in whole genome annotations (Suter et al., 2008).

Proteins carry out their function through complex formation with other macromolecules, including other proteins, forming PPIs (Fields, 2005). PPIs play an important role in all biological systems and participate in cellular processes, including cell signalling and metabolic pathways, as well as in pathogen-host interactions, therefore the study of these interactions is crucial (Parrish et al., 2006). By analysing PPIs the roles of the individual proteins can be determined, as well as their functions within the complexes they partake in leading to a more comprehensive picture of biological systems (Brückner et al., 2009; Suter et al., 2008). Additionally, if the function of one protein interactor is known, the functional relationship of associated proteins can be elucidated (Auerbach et al., 2002). Studying PPIs is essential if the proteome is to be understood in its complexity, results of which would contribute greatly to the expansion of systems biology (Suter et al., 2008). Particularly in plant-microbe studies, understanding the protein interaction networks allows for the better understanding of virulence proteins and how they trigger infection (Garbutt et al., 2014).

The complexity of protein structures is attributed to the dynamic nature of these macromolecules which can make them difficult to study in isolation. Furthermore, large scale analyses of the proteome have proven difficult due to protein sample degradation and limited amount of starting material (Tyers and Mann, 2003). Since the beginning of protein interaction research, two techniques have been preferred: the YTH approach and mass spectrometry (Fields, 2005). The resultant data has provided the starting point for protein-

protein interactome studies, and despite the incomplete picture it provides, the insights that have been gained from these techniques are significant (Parrish et al., 2006).

3.2.2 The yeast two-hybrid approach in interactome mapping

The YTH method was first established in the early 1990s to carry out high throughput screenings, during a time when other such research techniques were not available. The YTH approach is simple and quick to set up, as well as being sensitive (Fields, 2005). Consequently, the technique was used to generate the first interactome map for the genome of bacteriophage T7, and large scale YTH screens have been carried out for numerous viruses, *Helicobacter pylori*, budding yeast as well as in human studies (Parrish et al., 2006). These YTH studies in particular have contributed data on individual proteins, as well as large amounts of binary PPI data aiding in the expansion of interactome studies (Parrish et al., 2006).

The first step in a small scale YTH assay is creating two recombinant proteins; the proteins of interest and putative or predicted binding partners are fused to one part of a transcription factor. One (putative) binding partner is fused to the DBD and is labelled the "bait", whereas the other heterologous protein is a DNA activation domain (AD) fused to the other (putative) binding partner. The latter is labelled the "prey" (Fields, 2005). Reconstitution of the split transcription factor is indicative of the formation of a binding complex being formed between the two proteins of interest and results in the expression of a reporter gene (Fields, 2005). The original yeast two-hybrid system was modelled on the *Saccharomyces cerevisiae* Gal4 transcription factor and its DNA-binding and activation domains. The interaction of the fusion proteins resulted in the transcription factor binding to the Gal1 promoter, which in turn induced the expression of the *Escherichia coli lacZ* reporter gene (Fields and Song, 1989). The *lacZ* reporter gene encodes the chromogenic reporter β-galactosidase, which cleaves the artificial chromogenic substance 5-bromo-4-chloro-3-indolyl-β-d-galactopyranoside (X-gal) to produce an easily detectable blue product.

Over the years, the basis of the YTH approach has been improved, and even converted into a kit-based format for easy use (Fields, 2005). To improve sensitivity and selectivity, a range of reporter genes have been made available, including using prototrophic reporter genes. Many YTH approaches use more than one reporter gene to widen the spectrum and decrease the rate of false positives (Stynen et al., 2012). Large scale YTH assays have contributed a

considerable amount of PPI experimental data; however, the issue of false positives in the resultant data persists (Stynen et al., 2012).

The YTH assay has been used to answer the question of whether pathogen and host proteins interact, even on a large scale. One example of a large scale study elucidating the host-pathogen interactome was performed on herpesvirus and human proteins (Uetz et al., 2006). The two-hybrid system, which specifically looks at the interaction of two particular proteins, has also been used in the study of a geminiviral-plant host interaction. This experimental approach has been used in successfully identifying multiple host interactions, contributing to the interactome network of geminiviruses and their plant hosts (Kong and Hanley-Bowdoin, 2002; Kushwaha et al., 2017; Sánchez-Durán et al., 2011).

The Gateway® cloning method utilises the site specific recombination reaction carried out by bacteriophage lambda. The gene to be cloned requires flanking attachment sites (*att*) that transfer the gene between plasmids that also containing the recombination *att* sites. The system uses two separate reactions to transfer the gene of interest to its destination vector. The reaction is carried out between two recombination sites to yield the vector containing the cloned gene: *att*B x *att*P→ *att*L + *att*R (Hartley et al., 2000). The advantage of the system is that the gene of interest is maintained in one orientation. The reactions are induced through the commercially labelled BP clonase and LR clonase (Liang et al., 2013), enzymes mixtures that contain integrase, as well as integration host factor proteins (Hartley et al., 2000).

3.2.3 The role of geminivirus Rep in manipulating its plant host to facilitate viral replication

Geminiviruses have been shown to infect mature, differentiated plant cells, where replication machinery to facilitate viral replication, would not be readily available. As Rep cannot act as a DNA polymerase, the protein induces expression of host proteins that make up the replisome and that are required for viral replication (Hanley-Bowdoin et al., 2004). The replisome of consist DNA replication and repair proteins (Hanley-Bowdoin et al., 2013).

During the plant cell cycle, the RBR controls the progression of the plant cell through the cell cycle by preventing the expression of DNA replication associated genes. RBR also regulates stem cell maintenance and cell differentiation (Gutzat et al., 2012). As identified in animals, the RBR forms a complex with E2F transcription factors, thereby directing and inhibiting the

expression of replication dependant proteins, specifically those required for entering the S-phase of the cell cycle (Kong et al., 2000). In actively dividing cells, RBR is phosphorylated which disrupts the RBR-E2F complex. The phosphorylation of RBR inhibits the formation of the RBR-E2F complex, allowing for the expression of genes targeted by E2F during the late G_1 phase of the cell cycle, enabling entry of the plant cell into the S-phase of the cell cycle (Gutzat et al., 2012). In geminivirus infected cells, inactivation of RBR is facilitated through binding of Rep to RBR, thereby circumventing the need for its phosphorylation (Figure 3.1). Binding of Rep to RBR results in the expression of replisome related proteins regulated by the E2F transcription factors (Egelkrout et al., 2001; Nagar et al., 1995). In this way geminivirus infection allows for terminally differentiated plant cells to re-enter the cell cycle through the binding of Rep to RBR (Egelkrout et al., 2001; Kong et al., 2000; Nagar et al., 1995). This manipulation of the host cell cycle is conserved in small ssDNA viruses infecting both animals and plants (Hanley-Bowdoin et al., 2013).

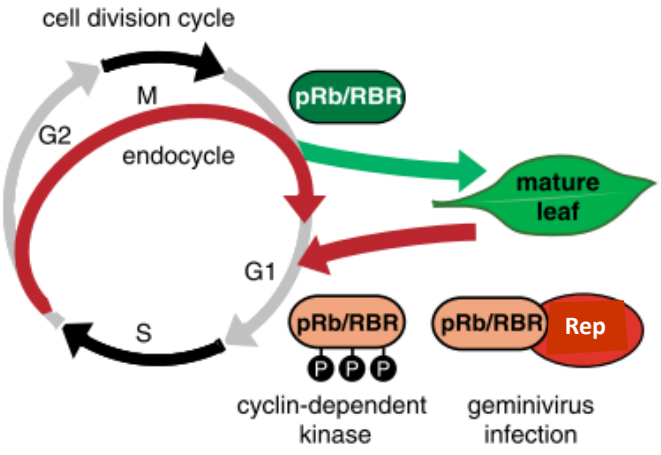

Figure 3.1 Schematic representation of the role of geminivirus Rep in the plant cell cycle. Retinoblastoma-related protein (RBR) (pRB in animal cells) is regulated through G_1–associated cyclin-dependent kinases, which phosphorylate RBR, preventing its inhibition of cell cycle progression and plant division. During geminivirus infection, Rep binds to RBR, circumventing the need for phosphorylation, enabling the expression of S-phase associated genes. Presence of Rep facilitates the re-entering of mature plant cell into the cell cycle or the endocycle (Hanley-Bowdoin et al. (2004) (slightly altered)).

The binding motif of Rep to RBR has been characterised in mastreviruses, as well as in oncoproteins of animal ssDNA viruses and nanoviruses. The RBR-binding motif was characterised as the aa sequence LXCXE (Liu et al., 1999). The RBR-binding motif has not been identified in SACMV or other begomoviruses; however, the putative binding motif has been narrowed down to a highly predicted alpha helix found within the oligomerisation domain (Arguello-Astorga et al., 2004). An RBR binding site within the region 101-180 was identified in TGMV, a virus of the genus *Begomovirus.* This region on the protein plays a role in the catalytic activity, DNA binding functions, and forms part of the oligomerisation domain. Two highly predicted alpha helices within this region are revealed to be involved in Rep-RBR interactions (Arguello-Astorga et al., 2004; Kong et al., 2000).

A central component of the replisome is PCNA, a protein that binds to DNA and modulates the binding of other replication proteins to DNA and aids in DNA replication alongside polymerase δ (Castillo et al., 2003; Kong et al., 2000). Also described as a sliding clamp, PCNA facilitates DNA-protein complexes involved in DNA methylation, DNA repair and cell cycle control, amongst others (Castillo et al., 2003). This replisome component is not present in healthy, terminally differentiated plant cells (Hanley-Bowdoin et al., 2013; Nagar et al., 1995). However, geminivirus infected plant cells display PCNA expression, a development that been attributed to the presence and activity of Rep (Ach et al., 1997; Egelkrout et al., 2001; Kong et al., 2000; Nagar et al., 1995). Expression of PCNA has been

shown to be negatively regulated by E2F transcription factors (Egelkrout et al., 2001). The presence of PCNA can be attributed to the influence of Rep on the cell cycle, through binding of Rep to RBR, thereby enabling the expression of genes regulated by the E2F transcription factor.

Further studies have been carried out to delve into the mechanism of action of Rep, and its ability to manipulate cells to re-enter the cell cycle. This was carried out in fission yeast, where an RBR homologue is not present, which suggests that Rep induces re-entry into the S-phase through other mechanisms (Kittelmann et al., 2009). Hipp et al. (2014) suggested that Rep influences the cell cycle through means other than binding to RBR. This theory is supported by the observation of up regulated S and G_2 phase genes from CaLCuV infected *Arabidopsis thaliana* (Ascencio-Ibáñez et al., 2008). The presence of a putative cyclin binding motif (RXL), conserved throughout geminivirus genera, further supports this theory, as it has also been identified in the replication initiator protein (E1) and E4 protein of mammalian ssDNA viruses, human papillomavirus. The RXL motif in these proteins contributes to the cell cycle arrest at G2/M, in turn maintaining a pseudo S-phase (Knight et al., 2011; Ma et al., 1999).

Plant infecting geminiviruses in turn induce the endocycle (Figure 3.1), a plant cycle characterised by increased chromosomal replication and cell expansion; however, cell division is typically inhibited (Ascencio-Ibáñez et al., 2008; Nagar et al., 2002) as shown in *A. thaliana* infected with CaLCuV (Ascencio-Ibáñez et al., 2008) and SACMV (Pierce and Rey, 2013). Some suggest that the Rep-RBR complex inhibits RBR from preventing the cell entering the endocycle, which is characterised by large nuclei and polyploidy. Hence, endoreduplication and the endocycle have been identified in geminivirus infected cells (Ascencio-Ibáñez et al., 2008). However, endoreduplication has also been exhibited in yeast cells that lack an RBR-equivalent. Furthermore, in the study carried out by Kittelman et al (2009), the yeast cells (*Schizosaccharomyces pombe*) were not in possession of a viral origin of replication, and yet Rep established an environment that stimulated replication of the yeast genome nonetheless (Kittelmann et al., 2009). This suggests that other host factors could contribute to the geminivirus infected cells carrying out endoreduplication. Gene regulation studies have also shown that mitosis-coupled genes as well as G_1 cyclins and cyclin-dependent kinases are differentially expressed during CaLCuV infection of *A. thaliana* (Ascencio-Ibáñez et al., 2008). Similar expression patterns were found in SACMV infected *A. thaliana*, suggesting the regulation of cyclin-associated processes by SACMV (Pierce and

Rey, 2013). A Rep-cyclin or Rep—kinase complex has been hypothesised, as a cyclin binding motif, RXL, was identified as a conserved sequence in geminivirus Rep (Hipp et al., 2014). Binding of Rep to cyclins may provide an alternative method of host cell cycle regulation (Hipp et al., 2014). These results point to the fact that Rep enables re-replication in plants, and prevents the progression of the cell cycle into the M-phase via an unknown pathway (Ascencio-Ibáñez et al., 2008; Kittelmann et al., 2009). It can be argued that since geminiviruses infect mature leaves, that geminivirus infection can more easily induce the endocycle, as the cell environment is primed for this (Kong and Hanley-Bowdoin, 2002). The environment that the endocycle creates is an another indication of how these plant pathogens control and alter the plant cell cycle to ensure an environment compatible for viral replication.

An additional Rep binding partner was identified in a study carried out by Kong et al., (2002). The geminivirus Rep-interacting motor protein (GRIMP) was shown to bind to TGMV Rep and CaLCuV Rep. GRIMP has an N-terminal kinesin domain and the central domain binds to a cyclin-dependent kinase (CDC2A). Phylogenetic analysis has placed it within the C-terminal kinesin subfamily (Kong and Hanley-Bowdoin, 2002). This kinesin subfamily has previously been shown to co-localise with spindles and aid in mitosis through microtubule regulation, suggesting it plays a significant role during mitosis (Goldstein and Philp, 1999). It has been theorised, that by binding to GRIMP, Rep inhibits phosphorylation of GRIMP by cyclin-dependent kinases (CDK), thereby negatively regulating the GRIMP and preventing it from localising at spindles and aiding the progression of mitosis (Kong and Hanley-Bowdoin, 2002).

3.2.4 Fluctuating core cell cycle expression upon geminivirus infection

Expression levels of a number of cell cycle related genes were altered as revealed by the transcriptome analysis of CaLCuV and SACMV-infected *A. thaliana* (Ascencio-Ibáñez et al., 2008; Pierce and Rey, 2013), as well as in SACMV-infected cassava (*Manihot esculenta*) (Allie et al., 2014). Amongst them were a number of cyclins from the plant cyclin D family (CYCD). These cyclins are involved in aiding quiescent cells to re-enter the cell cycle and also enable cell cycle progression by inducing the G_1 to S-phase transition (Dewitte and Murray, 2003; Kono et al., 2007). CDKs bound to D-type cyclins phosphorylate RBR, inhibiting the RBR-E2F complex. This enables the expression of E2F dependant genes,

which are genes involved in replication and associated with the S-phase (Ach et al., 1997). Expression of plant D-type cyclins is dependent on external signals. CYCD3 are present at the entry of the cell into the G_1 phase, but are most expressed before the cell enters the S-phase. This class of the D-type cyclins have been described as the main driving force for the transition between G_1 and S-phase (Menges, 2006). CYCD3;2 are known for their involvement in the RBR/E2F pathway, as they regulate the activity of RBR through binding in turn controlling G_1 to S-phase transition (Ach et al., 1997). In addition CYCD3 prevent the cell from entering the endocycle. CYCD4 is labelled a late G_1 phase cyclin (Menges et al., 2005) and has been identified as a driving force of cell division in formation of the stomata in hypocotyls (Kono et al., 2007). This D-type cyclin differs from other cyclins in the lack of an RBR binding domain. This suggests that CYCD4 phosphorylates other cell cycle inducers, and, as reported in animal cells, may be forming complexes with transcription factors other than E2F (Kono et al., 2007).

The transcriptome data of both Ascencio-Ibáñez et al. (2008) and Pierce and Rey (2013) revealed the up-regulation and down-regulation of CYCD4;2 and CYCD3;2, respectively (Figure 3.2). Further experimentation showed that overexpressed CYCD3;1 hindered symptom development in CaLCuV infected *A. thaliana*. Contrastingly, *A. thaliana* expressing mutated CYCD3 genes and infected with CaLCuV resulted in a high viral load and severe disease symptoms (Ascencio-Ibáñez et al., 2008). The down-regulation of CYCD3 could be justified due to the involvement of this cyclin subfamily in prevention of the endocycle (Dewitte et al., 2007). Previous research has shown that geminiviruses prefer infecting the 4C cell population that is programmed to enter the endocycle. With decreasing CYCD3 RNA transcripts, more cells enter the endocycle. Consequently the population of 4C increases, which in turn enables geminivirus infection (Ascencio-Ibáñez et al., 2008).

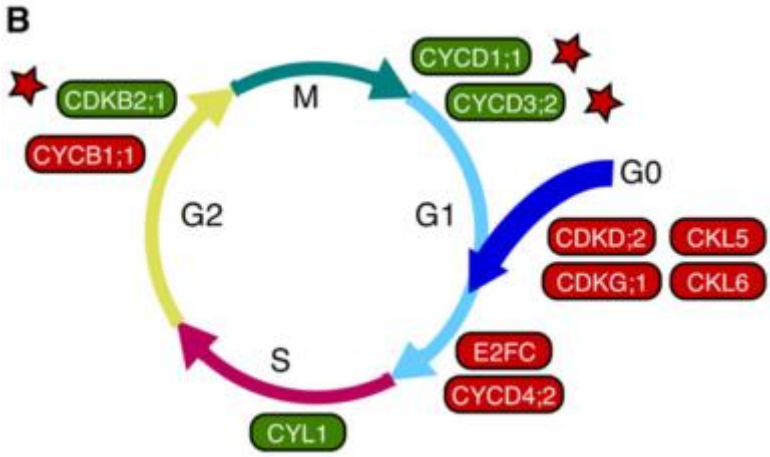

Figure 3.2 A model showing differentially expressed plant core cell cycle genes. The diagram shows at which time point during the cell cycle these core cell cycle genes are being expressed. Red is indicative of upregulation, green of downregulation. The cell cycle genes marked with the star are downregulated genes and show their functions at specific time point of the cell cycle (Ascencio-Ibáñez et al., 2008).

3.2.5 The function of histone H3 in transcriptional regulation and modification generated by geminivirus Rep

Histone 3 forms part of the nucleosome core particle, alongside H2A, H2B and H4. The DNA and histone core form chromatin. The structure of the chromatin and domains within the chromatin controls replication of the DNA and transcription. Post-translational modification of core histones can act as switches, activating transcription of genes or DNA replication (Mariño-Ramírez et al., 2005).

The geminivirus genome is known to form minichromosomes in its double-stranded form (Pilartz and Jeske, 1992). To analyse the role of histones and histone modifying enzymes in geminivirus infection these proteins were analysed through a transcriptomics study of SACMV infected cassava. A number of histone were up- and down-regulated in the susceptible cassava landrace, while the tolerant land race displayed the down-regulation of a histone acetyltransferase, suggesting that the tolerant plant was counteracting cell cycle dependant replication of the viral genome (Allie et al., 2014).

Due to the formation of minichromosomes, the viral origin of replication and promoters may not be accessible for Rep to initiate replication (Hanley-Bowdoin et al., 1999). Research has shown that TGMV Rep circumvents this by binding to *A. thaliana* Histone H3. It is suggested that this interaction enables the removal of nucleosomes so that viral replication and transcription of viral genes can take place (Kong and Hanley-Bowdoin, 2002). Interaction of

SV40 large T antigen with Histones H1 and H3 has also recorded, and was found to play a role in chromatin remodelling (Ramsperger and Stahl, 1995), as has been confirmed experimentally for the TGMV Rep-H3 interaction (Kong and Hanley-Bowdoin, 2002). The Rep-H3 interaction was also suggested to play a role in preventing methylation of lysine 9 of histone H3. This post-translational modification was found to negatively affect viral pathogenicity (Raja et al., 2008).

3.2.6 Aim

The aim of this study is to identify potential binding partners of SACMV Rep and assess binding through the use of a yeast two-hybrid assay.

3.2.7 Specific objectives

- To clone AC1 (ORF encoding for SACMV Rep) into the bait vector (pDEST^{TM}32) and the prey vector (pDEST^{TM}22)
- To identify cassava host proteins that may act as potential SACMV Rep binding partners using transcriptomics data
- To clone prey genes (*Histone H3*, *CYCD3;2*, *CYCD4;2*) into the bait vector (pDEST^{TM}32)
- To transform yeast strain, MAV203 with YTH controls and bait and prey clones
- To analyse binding interactions through the use of an X-gal assay using the yeast two-hybrid filter-lift method

3.3 Methods and Materials

3.3.1 Cloning bait gene, AC1 into pDEST^{TM}32 and pDEST^{TM}22

The chosen bait gene, AC1 (encoding for the protein SACMV Rep), was cloned into Gateway® pDONR221^{TM} vector through PCR. SACMV AC1 was amplified from pJET1.2/AC1 using primers containing the attB1 adaptors (Table 3.1). It is important to note that the pJET1.2/AC1 utilised in the present study contains a single nucleotide mutation in the AC1 ORF at nucleotide position 620 (Figure 3.8). This mutation arose spontaneously and results in an amino acid mutation at position 207 from a phenylalanine in the originally-deposited sequence (NP_620665.1) to a serine in pJET1.2/AC1. This amino acid mutation is located outside the functional motifs of Rep and is present in the Zimbabwean variant of SACMV (CAE01424.1) and other SACMV AC1 variants.

Table 3.1 Primer names and sequences used to amplify AC1 for Gateway® cloning purposes. The respective sequences of the genes are shown in bold, the *att*B adaptors are in normal font.

Primer name	Primer sequence
Rep*att*B1F	5′-GGGGACAAGTTTGTACAAAAAAGCAGGCT**TCATGCCGA GGGCTGGTCGTTT**
Rep*att*B1R	5′-GGGGACCACTTTGTACAAGAAAGCTGGGTC**CTAGTCTTC GCGGTGCGGTGTT**

The PCR reaction was carried out using 2× Phusion High Fidelity PCR Master Mix with primers RepattB1F and RepattB1R. The components of the PCR reaction are shown in Table 3.2.

Table 3.2 Polymerase chain reaction components for amplification of AC1 from pJET1.2.

Reaction component	Volume (µL)	Concentration
H_2O	7.5	-
2× Phusion High Fidelity PCR Master Mix	10	1x
RepattB1F	1	0.5 mM
RepattB1R	1	0.5 mM
pJET1.2/AC1	0.5	125 ng
Total reaction volume	20	-

Initial denaturation was carried out at 98° C for 30 s, with denaturation completed at 98° C for a further 10 s. Primer annealing was performed at 65° C for 20 s. Extension and final extension was carried out at 72° C for 20 s and 5 min, respectively. The denaturation to extension steps were repeated 35 times.

The PCR products were subjected to a PCR clean-up reaction as described in Gateway® Technology with Clonase® II manual (Thermo Fisher Scientific, Waltham, USA). Briefly, 150 µL of TE buffer (pH 8.0) was added to 50 µL of PCR amplified *att*B-PCR product. An additional 100 µL of 30% PEG 8000/30 mM $MgCl_2$ was added, vortexed and centrifuged at 10000 ×g for 15 min. The resulting supernatant was extracted, and the pellet was resuspended in 50 µL of TE buffer, pH 8.0. The recovered *att*B-PCR product was analysed on a 1% agarose gel and quantified using the NanoDrop 1000 (Thermo Fisher Scientific, Waltham, USA).

The methodology for the BP recombination reaction was provided by the Gateway® Technology with Clonase® II manual (Thermo Fisher Scientific, Waltham, USA). The BP recombination reaction was carried out using 20 femtomoles of the *att*B-PCR product and 150 ng of the donor vector, pDONR^{TM}221. The nanograms of PCR product that equals 20 femtomoles was calculated using Equation 4.

$$ng = (fmol) \times (N) \times (\frac{660\ fg}{fmol}) \times (\frac{1\ ng}{10^6 fg}) \qquad \text{Eq. 4}$$

Where N is the size of DNA in base pairs (bp)

The BP recombination reaction was made up as described in Gateway® Technology with Clonase® II manual (Thermo Fisher Scientific, Waltham, USA), using the components as shown in Table 3.3.

Table 3.3 Components of the BP recombination reaction. Reaction components required for the BP recombination reaction of the polymerase chain reaction (PCR) fragment AC1 into pDONR^{TM}221 using Gateway® cloning.

Reaction component	Volume (µL)
*att*B-PCR product	2
pDONR^{TM}221	1
TE buffer, pH 8.0	5
Gateway® BP Clonase® II enzyme mix	2
Total reaction volume	10

The BP Clonase II enzyme mix was incubated on ice two minutes prior to use and vortexed twice for 2 s then added to the reaction mixture. The reaction mixture was incubated at 25° C for an hour. The reaction was suspended by adding Proteinase K and incubating at 37° C for 10 min. The entry clone that was yielded from the reaction was transformed into *Escherichia coli* DH5α using the heat-shock method. Briefly, 1 µL of the recombination reaction was added to competent *E. coli* DH5α, and incubated on ice for 30 min. The cells were subjected to 42° C for 1 min, and placed on ice for 2 min. SOC (450 µL) was added to the cells. The cells were incubated at 37° C for 1 hour to allow for growth and 100 µL was spread onto SOB media containing 50 µg.ml^{-1} kanamycin and incubated overnight at 37° C. Resultant colonies were used to inoculate broth which was subsequently used to make 2:1 glycerol stocks (cell culture: 50% glycerol).

Glycerol stocks were used to inoculate an overnight culture, from which plasmid was extracted using GeneJet Gel Extraction Kit (Thermo Fisher Scientific, Waltham, USA). The LR recombination reaction was carried out using 150 ng of pDONR^{TM}221/AC1 and 150 ng of the destination vector, pDEST^{TM}32/pDEST^{TM}22. The reactions were set up as recommended in ProQuest™ Two-Hybrid System with Gateway® Technology manual (Table 3.4).

Table 3.4 Components required for the LR recombination reaction. Reaction components required for the LR recombination reaction of AC1 from pDONR™221 into both pDEST™32 and pDEST™22.

Reaction component	Volume (µL)
pDONR™221/AC1	1
pDEST™32/pDEST™22	1
TE buffer (pH 8.0)	6
LR Clonase™ II enzyme mix	2
Total reaction volume	10

The LR Clonase™ enzyme mix was placed on ice for 2 minutes prior to usage, vortexed twice for 2 s, and then aspirated into the reaction mixture. The reaction was incubated at 25° C for 1 h, then terminated by adding 1 µL Proteinase K and incubated for 10 min at 37° C. The LR reactions were transformed into DH5α, as described previously. Glycerol stocks of the newly constructed bait and prey plasmids, pEXP™32-AC1 and pEXP™22/AC1, were made, as described previously for the pDONR™221/AC1 stock.

The recombinant destination vectors, pEXP™32/AC1 and pEXP™22/AC1, were sequenced (Inqaba Biotech™, Pretoria, South Africa) to verify their nucleotide identities.

3.3.2 Identification of prey genes from transcriptome data and cloning into pDEST™22 and pDEST™32

Cassava cv. TME3 (tolerant) and T200 (susceptible) transcriptome data (Allie et al., 2014) and *A. thaliana* transcriptome data (Pierce et al., 2013) was used to select prey targets. The cassava genes that encode for Histone H3, cyclin D3;2 (CYCD3;2) and cyclin D4;2 (CYCD4;2) proteins were selected.

The Histone H3 protein was identified as a suitable candidate based on research conducted by Kong and Hanley-Bowdoin (2002). The study carried out by Allie et al. (2014), analysed the differentially expressed histone-related genes in SACMV-infected cassava. Allie et al. (2014) (Supplementary file 13) reported the differentially expressed histone-related genes at each time point and in each of the two cultivars infected with SACMV. The *Histone H3* gene with the highest log2 fold change with the Phytozome identifier, Manes.13G097500.1.p (Allie et al., 2014), was chosen.

Using the transcriptome data from Pierce et al. (2013) and Allie et al. (2013), the sequences for *CYCD3;2* and *CYCD4;2* were obtained from Phyotozome (Goodstein et al., 2012). *CYCD3;2* was identified using supplementary data of Pierce et al. (2013) (Supplementary file 6). The *A. thaliana* accession number was used to identify the cassava homologue in the cassava genome annotation data (Manes.13G119100) (Supplementary file 2) of Allie et al. (2014). The *A. thaliana* gene *CYCD4;2*, as identified by Pierce et al. (2013) was not available in the cassava annotation data of Allie et al. (2014) (Supplementary file 2). To obtain the cassava homologue of *A. thaliana CYCD4;2*, BLAST (Altschul et al., 1990) was used to identify the cassava homologue (Manes.12G128000) using the annotated cassava genome available through phytozome (Goodstein et al., 2012).

Synthesis and cloning of the prey genes into the destination vector pDEST^{TM}22 was conducted by GenScript (Piscataway, NJ, USA) and supplied as a purified recombinant vector.

To clone the three prey genes separately into pDEST^{TM}32, the genes were amplified using PCR primers designed to add the *att*B adaptors (Table 3.5).

Table 3.5 Primers used to amplify the prey genes from their respective pDEST^{TM}22 vectors. Polymerase chain reaction (PCR) amplified gene is used for cloning into pDONR^{TM}221. The respective sequences of the genes are shown in bold, the *att*B adaptors are in normal font.

Primer name	Primer sequence
H3*att*B1	5′- GGGACAAGTTTGTACAAAAAAGCAGGCTTC**ATGGCTCGCAC**
H3*att*B2	5′- GGGGACCACTTTGTACAAGAAAGCTGGGTC**CTAAGCCCTCTC**
CYCD3*att*B1	5′- GGGGACAAGTTTGTACAAAAAAGCAGGCTTC**ATGGCTTTG**
CYCD3*att*B2	5′- GGGGACCACTTTGTACAAGAAAGCTGGGTC**TCAAATAAAG**
CYCD4*att*B1	5′- GGGGACAAGTTTGTACAAAAAAGCAGGCTTC**ATGGCACAG**
CYCD4*att*B2	5′- GGGGACCACTTTGTACAAGAAAGCTGGGTC**CTATCTTTGGAC**

Histone H3 was PCR amplified from templates pDEST^{TM}22/H3 using Phusion U Hot-Start DNA polymerase (Thermo Fisher Scientific, Waltham, USA). The PCR reaction was set up as described in Table 3.6.

Table 3.6 PCR reaction components for amplification of *histone H3*. Polymerase chain reaction (PCR) components for the amplification of the gene for histone H3 from pEXP^{TM}22/H3 vectors for cloning into pDONR^{TM}221.

Reaction component	Volume (µL)	Concentration
H_2O	32.5	-
5× Phusion HF Buffer	10	1x
dNTPs	1	200 µM
Forward primer	2.5	0.5 mM
Reverse primer	2.5	0.5 mM
Phusion U Hot-Start DNA polymerase	1	0.4 U
DNA template	0.5	150 ng
Total reaction volume	50	

The PCR reaction was carried out using the three step protocol, as shown in Figure 3.3.

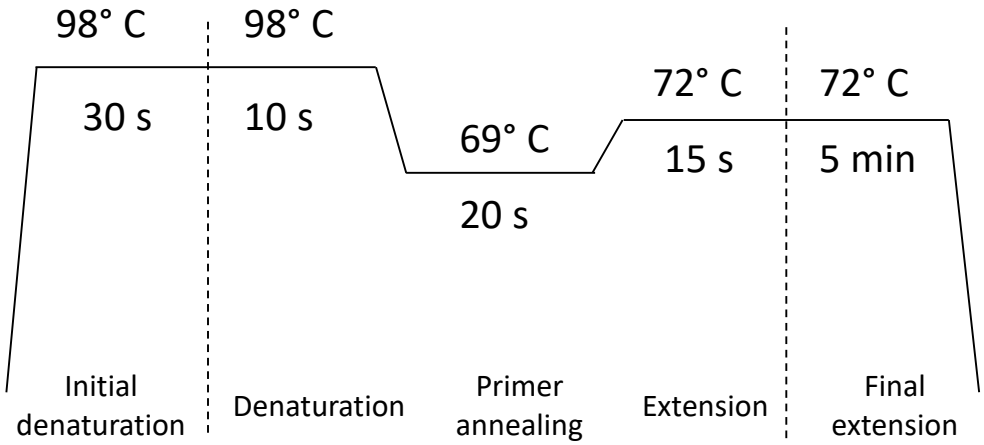

Figure 3.3 PCR profile for amplification of *histone H3*. Polymerase chain reaction (PCR) amplification profile, showing the different stages of gene amplification. The termperature for each step is shown above the line, time taken for each step is shown below the line.

The PCR product was subjected to a PCR clean-up reaction as described in Gateway® Technology with Clonase® II manual (Thermo Fisher Scientific, Waltham, USA) and previously with *att*B-PCR product *AC1* in Section 3.3.1.

The other two prey genes, *CYCD3;2* and *CYCD4;2* were PCR amplified from DNA templates pDONR^{TM}221/CYCD3;2 and pDONR^{TM}221/CYCD4;2, respectively, using primers as shown

in Table 3.5. The reactions were set up as shown in Table 3.7, using Phusion U Hot Start DNA polymerase.

Table 3.7 PCR reaction components for the amplification of *CYCD3;2* and *CYCD4;2*. Polymerase chain reaction (PCR) components for the amplification of the gene for CYCD3;2 and CYCD4;2 from pEXP^{TM}22/CYCD3;2 and pEXP^{TM}22/CYCD4;2 for cloning into pDONR^{TM}221 using the Gateway® system.

Reaction component	Volume (µL)	Concentration
H_2O	11.8	-
5× Phusion HF Buffer	4	1x
dNTPs	0.4	200 µM
Forward primer	1	0.5 mM
Reverse primer	1	0.5 mM
Phusion U Hot Start DNA polymerase	0.2	0.4 U
100% DMSO	0.6	3%
DNA template	1	110 ng
Total reaction volume	20	

Key: 100% DMSO- Dimethyl sulfoxide

Due to the length of the primers for amplification of *CYCD3;2* and *CYCD4;2*, a two-step protocol PCR was carried out using the PCR profile described in Figure 3.4.

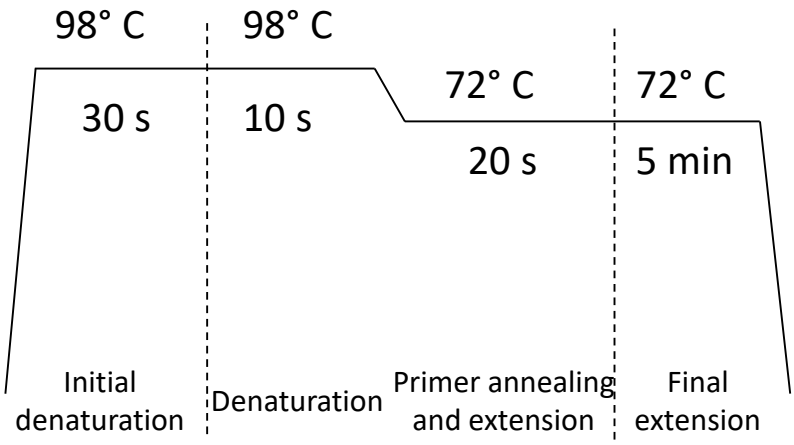

Figure 3.4 PCR amplification profile for amplification of *CYCD3;2* and *CYCD4;2*. Polymerase chain reaction (PCR) amplification profile indicating the degrees and time taken for each steo of PCR for amplification of *CYCD3;2* and *CYCD4;2* for Gateway® cloning.

The PCR products were analysed on a 1% agarose gel. The resultant product, *att*B-H3 product was cloned into pDONR^{TM}221 using the BP reaction method as described in the

Gateway® Technology with Clonase® II manual (Thermo Fisher Scientific, Waltham, USA). This was performed as described in Section 3.3.1, Table 3.3. As before, 50 femtomol of recovered *att*B-PCR product and 150 ng of pDONR^{TM}221 were used for each BP reaction. The resultant pDONR^{TM} clones (pDONR^{TM}221/H3, pDONR^{TM}221/CYCD3;2, pDONR^{TM}221/CYCD4;2) were transformed into DH5α and bacterial glycerol stocks were made as described in Section 3.3.1.

The LR recombination reaction followed as described previously in Section 3.3.1, Table 3.4, with the reaction taking place between pDONR^{TM}221/H3, pDONR^{TM}221/CYCD3;2, pDONR^{TM}221/CYCD4;2 and pDEST^{TM}32.

A restriction digest was carried out to confirm the presence of recombinant clones. Recombinant bait vectors, pEXP^{TM}32/H3 and pEXP^{TM}32/CYCD3;2 (500 ng each) were digested using Fast digest^{TM} *Xho*I (Thermo Fisher Scientific, Waltham, USA) and pEXP^{TM}32/CYCD4;2 (500 ng) was digested using Fast digest^{TM} *Hind*III (Thermo Fisher Scientific, Waltham, USA) at 37° C for 1 h. The restriction digest was run on a 1% agarose gel and the clones were sequenced to confirm accuracy of the sequences (Inqaba Biotech, Pretoria).

3.3.3 Transformation of yeast strain MAV203 with YTH controls and bait and prey clones

The yeast strain supplied with the ProQuest ™ Two-Hybrid System with Gateway® Technology kit, *S. cerevisiae* MAV203, was made competent using the *S.c.* EasyComp™ Transformation Kit (Thermo Fisher Scientific). The competent yeast cells were transformed as per the manufacturer's instructions. A 50 µL aliquot of competent *S. cerevisiae* MAV203 was inoculated with up to 5 µg of plasmid, and supplemented with 500 µL of solution III (Thermo Fisher Scientific) was added and the mixture was vortexed. The yeast cells were incubated in a 30 ° C water bath for 1 h. The transformation mixture was mixed vigorously every 15 min by vortexing. To select for transformed yeast cells, 100 µL of the transformation mixture was plated onto selection media, synthetic dropout media without leucine and tryptophan (SC-Leu-Trp) (Clontech, Takara Bio Inc., Japan).

Control bait and prey plasmids provided with the ProQuest^{TM} kit were transformed into competent *S. cerevisiae* MAV203. The control bait protein is Rap1A, also labelled Krev1, of

the Ras family of GTP-binding proteins. The strong control consists of the prey protein expressed alongside the wild type prey protein, RalGDS-wt (the Ral guanine nucleotide dissociator stimulator protein). The weak and negative control are formed by the interaction of the bait protein, Rap1A with two mutant prey protein. RalGDS-mutant 1 (RalGDS-m1) interfering slightly with the interaction between the bait and prey, giving rise to the weak control. RalGDS-mutant 2 cannot interact with the bait protein, Rap1A, thereby this bait and prey combination gives rise to the negative control (Herrmann et al., 1996; Serebriiskii et al., 1999).

Yeast cells transformed with both a bait and prey plasmids were plated onto selective media. Dropout media was used to select for the bait and prey vectors, respectively, as these encode for leucine and tryptophan, enabling growth on selective media deficient in these two essential amino acids. Different combinations of the generated prey and bait plasmids containing AC1, *Histone H3*, *CYCD3;2* and *CYCD4;2* were transformed into *S. cerevisiae* MAV203 alongside pDEST^{TM}32 and pDEST^{TM}22 to form controls for YTH (Table 3.8).

Table 3.8 Control bait and prey vectors, combinations and control types. Combinations of bait and prey clones transformed into *Saccharomyces cerevisiae* MAV203 that serve as controls alongside the experimental bait and prey clones for the yeast two-hybrid X-gal assay.

Control #	Bait vector	Prey vector	Type of control
2	pEXP^{TM}32/Krev1	pEXP^{TM}22/RalGDS-wt	Strong positive interaction control
3	pEXP^{TM}32/Krev1	pEXP^{TM}22/RalGDS-m1	Weak positive interaction control
4	pEXP^{TM}32/Krev1	pEXP^{TM}22/RalGDS-m2	Negative interaction control
5	pDEST^{TM}32	pDEST^{TM}22	Negative activation control
6	pEXP^{TM}32/AC1	pDEST^{TM}22	Negative activation control
7.1	pDEST^{TM}32	pEXP^{TM}22/AC1	Negative activation control
7.2	pEXP^{TM}32/H3	pDEST^{TM}22	Negative activation control
7.3	pEXP^{TM}32/CYCD3;2	pDEST^{TM}22	Negative activation control
7.4	pEXP^{TM}32/CYCD4;2	pDEST^{TM}22	Negative activation control

Finally, various combinations of generated bait and prey clones were transformed into *S. cerevisiae* MAV203 to perform interaction studies using YTH (Table 3.9).

Table 3.9 Experimental bait and prey clones and combinations thereof. Combinations of bait and prey clones containing AC1 (Rep) and *Histone H3*, *CYCD3;2* and *CYCD4;2*, and one experimental sample with Rep expressed from the bait and prey vector.

Bait vector	Prey vector
pEXP^{TM}32/AC1	pEXP^{TM}22/H3
pEXP^{TM}32/AC1	pEXP^{TM}22/CYCD3;2
pEXP^{TM}32/AC1	pEXP^{TM}22/CYCD4;2
pEXP^{TM}32/H3	pEXP^{TM}22/AC1
pEXP^{TM}32/CYCD3;2	pEXP^{TM}22/AC1
pEXP^{TM}32/CYCD4;2	pEXP^{TM}22/AC1
pEXP^{TM}32/AC1	pEXP^{TM}22/AC1

3.3.4 Analysing protein binding partners using the X-gal assay

To carry out the YTH assay, relevant controls and experimental interaction recombinant yeast strains were plated onto selective media (SC-Leu-Trp) (Clontech, Takara Bio Inc., Japan) from 24 h plated cultures, as shown in Figure 3.5 to form the master plate.

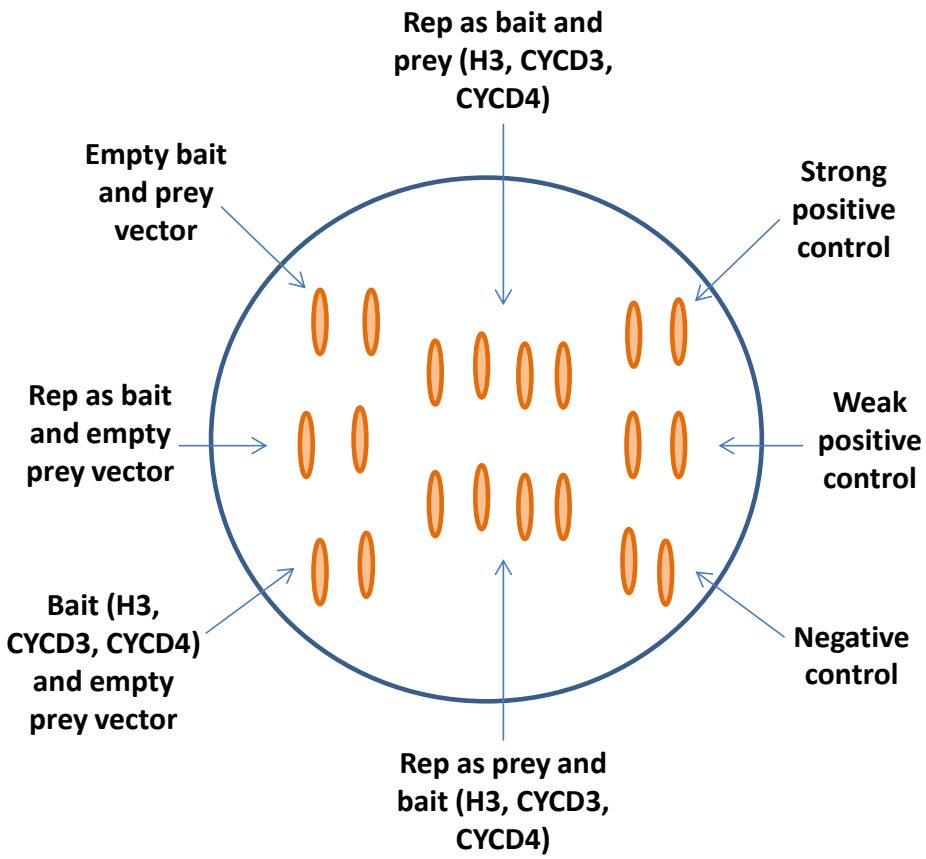

Figure 3.5 Schematic model showing the set up of the master plate for the Y2H. All respective controls are included on the sides of the plate, with the experimental interactions streaked down the middle of the plate. Two colonies are streaked for each control, with the experimental interactions requiring four colonies each.

The plates were incubated for 18 h at 30° C. The colonies were replica plated onto YPAD plates (see Appendix), containing a Whatman™ Nytran®N nylon membrane (GE Healthcare, USA). These plates were incubated for up to 24 h at 30° C. The YTH was carried out by performing a colony lift filter assay on the nylon membrane. Per membrane, 10 mg of X-gal (Thermo Fisher Scientific, Waltham, USA) was dissolved in 100 µL dimethylformamide (DMF), which was in turn combined with 60 µL β-mercaptoethanol and 10 ml buffer Z (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, pH 7.0) (ProQuest ™ Two-Hybrid System with Gateway® Technology) (Thermo Fisher Scientific, Waltham, USA). This solution was used to soak two Whatman 541 filter papers (Whatman® Schleicher & Schuell®, GE Healthcare) in a petri dish. The membranes were removed from the plate and frozen in liquid nitrogen for 20-30 s and placed on top of the soaked Whatman filters. The plates were wrapped in aluminium foil to protect from light and were incubated at 37° C

for up to 24 h. The presence of blue colonies was observed. The YTH assays were carried out in triplicate for each interaction.

3.4 Results

3.4.1 Cloning Rep into pDEST^{TM}32 and pDEST^{TM}22

The sequence encoding the SACMV replication-associated protein (Rep) was amplified from pJET1.2 using primers containing the *att*B adaptors as described in Gateway® Technology with Clonase® II manual (Thermo Fisher Scientific, Waltham, USA). The amplified product was 1115 bp in length (Figure 3.6).

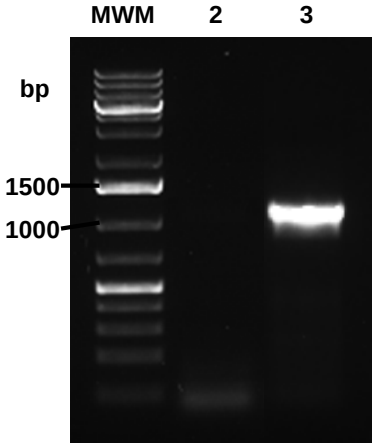

Figure 3.6 1% agarose gel showing PCR amplified AC1 with *att*B adaptors. Polymerase chain reaction (PCR) amplification of AC1 (Rep) from pJET1.2 using primers containing *att*B adaptors is shown in lane three. DNA molecular weight marker (MWM) is shown in the first lane in base pairs (bp), with the no-template control shown in lane two. MWM is 1 kB plus DNA marker (Thermo Fisher Scientific, Waltham, USA).

A clean-up was performed on the *att*B-PCR product and this was used to carry out the BP recombination reaction, with the cloning of AC1 (Rep) into pDONR^{TM}221. The LR recombination reaction was carried out using pDONR^{TM}221/AC1 with pDEST^{TM}32 and pDEST^{TM}22, with the resultant recombinant vectors, pEXP^{TM}32/AC1 and pEXP^{TM}22/AC1 (Figure 3.7) being sequenced. The sequence was aligned with the nucleotide sequence for SACMV AC1 (replicase, NP_620665.1). As mentioned previously, the mutation present at nucleotide 620 results in a single amino acid change of a phenylalanine to a serine (Figure 3.8). This mutation is prevalent in other SACMV Rep sequences and is present in the Zimbabwe variant of SACMV Rep (Accession number CAE01424.1).

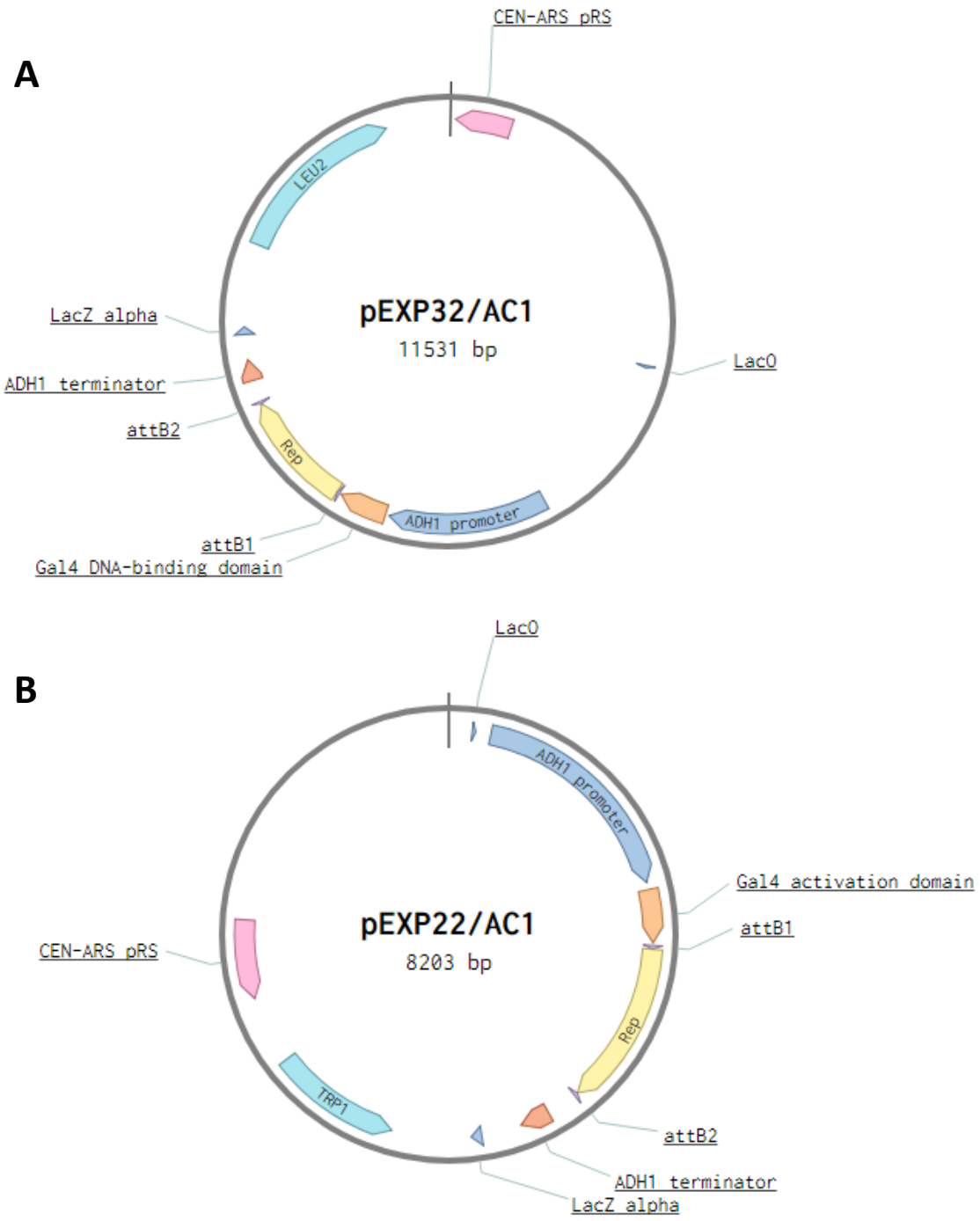

Figure 3.7 Schematic representation of bait and prey vectors containing AC1 (Rep-encoding ORF). The gene encoding for an enzyme involved in the leucine biosyntheis pathway is shown in cyan in vector **A**. Vector **B** includes a gene that encodes for an enyme involved in biosynthesis pathways of tryptophan. The genes on these vectors allow for selection in transformed yeast. The Cen-ARS gene replicates and maintains the plasmids at low copy number. The *att*B1 and 2 adaptors are indicated in vectors **A** and **B**. The ADH1 promoter and terminator are shown in blue and red, respectively, with the Gal4 DNA-binding domain present in vector **A** and Gal4 activation domain in vector **B**, respectively. The lac operon and LacZ alpha are indicted in blue in both vectors. The AC1 gene encoding Rep is shown in yellow for both bait (**A**) and prey (**B**) vectors.

```
Query  1     ATGCCGAGGGCTGGTCGTTTTAGCATAAAAGCCAAAAATTATTTCCTCACGTATCCGAAA  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2647  ATGCCGAGGGCTGGTCGTTTTAGCATAAAAGCCAAAAATTATTTCCTCACGTATCCGAAA  2588

Query  61    TGCACTCTCTCGAAAGAAGCGGCATTAGATCAACTCCGACAACTCCAAACCCCAACAAAT  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2587  TGCACTCTCTCGAAAGAAGCGGCATTAGATCAACTCCGACAACTCCAAACCCCAACAAAT  2528

Query  121   AAATTGTTCATCAAGATCTGCAGAGAACTCCATGAAAATGGGGAACCTCATTTGCATGCC  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2527  AAATTGTTCATCAAGATCTGCAGAGAACTCCATGAAAATGGGGAACCTCATTTGCATGCC  2468

Query  181   CTCATTCAGTTCGAGGGCAAGTACAATTGTACCAACCAACGATTCTTCGACCTCATATCC  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2467  CTCATTCAGTTCGAGGGCAAGTACAATTGTACCAACCAACGATTCTTCGACCTCATATCC  2408

Query  241   CCTTCCAGGTCAACACATTTCCATCCAAACATTCAGGGAGCTAAATCCAGTTCTGACGTC  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2407  CCTTCCAGGTCAACACATTTCCATCCAAACATTCAGGGAGCTAAATCCAGTTCTGACGTC  2348

Query  301   AAGTCCTATTTGGACAAGGACGGAGACACCATCCAATGGGGCGAGTTTCAGATCGACGGA  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2347  AAGTCCTATTTGGACAAGGACGGAGACACCATCCAATGGGGCGAGTTTCAGATCGACGGA  2288

Query  361   CGATCTGCTCGCGGCGGACAACAATCCGCCAATGACGCTTACGCCAAGGCTCTTAACGCA  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2287  CGATCTGCTCGCGGCGGACAACAATCCGCCAATGACGCTTACGCCAAGGCTCTTAACGCA  2228

Query  421   GCAAGTAAAACAGAGGCTCTTAATGTAATCCGGGAACTAGCCCCAAAGGATTTTGTTTA  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2227  GCAAGTAAAACAGAGGCTCTTAATGTAATCCGGGAACTAGCCCCAAAGGATTTTGTTTA  2168

Query  481   CAGTTTCATAATTTAAATAGCAATTTAGATAGGATTTTTCAGGAGCCTCCGATTCCTTAT  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2167  CAGTTTCATAATTTAAATAGCAATTTAGATAGGATTTTTCAGGAGCCTCCGATTCCTTAT  2108

Query  541   ATTTCTCCCTTTCTTTCTTCTTCTTTCACTCATGTTCCTGAGGAACTTGAAGACTGGGTT  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2107  ATTTCTCCCTTTCTTTCTTCTTCTTTCACTCATGTTCCTGAGGAACTTGAAGACTGGGTT  2048

Query  601   TCCGAGAACGTGATGGGTTCCGCTGCGCGGCCATGGAGACCGAGTAGTATCGTCATCGAG  660
             ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Sbjct  2047  TCCGAGAACGTGATGGGTTTCGCTGCGCGGCCATGGAGACCGAGTAGTATCGTCATCGAG  1988

Query  661   GGCGATAGTAGGACAGGGAAGACGATGTGGGCCCGATCTCTGGGACCACACAACTACTTA  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1987  GGCGATAGTAGGACAGGGAAGACGATGTGGGCCCGATCTCTGGGACCACACAACTACTTA  1928

Query  721   TGTGGACATTTGGATCTCAGTCCAAAGGTTTACAGCAACGACGCATGGTACAACGTCATT  780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1927  TGTGGACATTTGGATCTCAGTCCAAAGGTTTACAGCAACGACGCATGGTACAACGTCATT  1868

Query  781   GATGACGTCGACCCCCATTACCTCAAGCACTTCAAAGAATTCATGGGGGCCCAAAGGGAC  840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1867  GATGACGTCGACCCCCATTACCTCAAGCACTTCAAAGAATTCATGGGGGCCCAAAGGGAC  1808

Query  841   TGGCAAAGCAATACCAAGTACGGGAAGCCGATTCAAATTAAAGGCGGCATTCCCACTATC  900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1807  TGGCAAAGCAATACCAAGTACGGGAAGCCGATTCAAATTAAAGGCGGCATTCCCACTATC  1748

Query  901   TTCCTATGCAATCCAGGACCGACATCATCATATAAAGAGTTTCTGGACGAGGAAAAGAAC  960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1747  TTCCTATGCAATCCAGGACCGACATCATCATATAAAGAGTTTCTGGACGAGGAAAAGAAC  1688

Query  961   CAGTCCCTTAAAGCCTGGGCTTTAAAGAATGCAACCTTCATCACCCTCCACGAGCCATTG  1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1687  CAGTCCCTTAAAGCCTGGGCTTTAAAGAATGCAACCTTCATCACCCTCCACGAGCCATTG  1628

Query  1021  TTCTCAAGTGCCCATCAAAGTCCAACACCGCACCGCGAAGACTAG  1065
             |||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1627  TTCTCAAGTGCCCATCAAAGTCCAACACCGCACCGCGAAGACTAG  1583
```

Figure 3.8 Nucleotide sequence alignment of cloned SACMV Rep with SACMV replicase.Alignment (Altschul et al., 1990) of South African cassava mosaic virus (SACMV) AC1 (Accession number NP_620665.1) (Berrie et al., 2001) cloned into pDEST^{TM}32 and pDEST^{TM}22. The mutation is located at query nucleotide position 620.

3.4.2 Cloning prey genes into pDEST^{TM}32

The three prey genes *histone H3* (Manes.13G097500.1), c*yclin D3;2* (Manes.13G119100.1) and *cyclin D4;2* (Manes.12G128000.2) as selected previously were successfully cloned into pDEST^{TM}22 by GenScript (Piscataway, NJ, USA), and sequences of each clone provided

were evaluated to confirm sequence identity (See Appendix for Certificate of Analysis for pEXP^{TM}22/H3, pEXP^{TM}22/CYCD3;2 and pEXP^{TM}22/CYCD4;2).

The prey genes were amplified using primers with the *att*B adaptor to facilitate Gateway® cloning into the bait vector, pDEST^{TM}32 (Figure 3.9). Amplification of *histone H3* gave rise to the expected DNA product of 411 base pairs, with the amplficiation of *CYCD3;2* and *CYCD4;2* resulting in the required DNA products of 1185 bp and 900 bp, respectively.

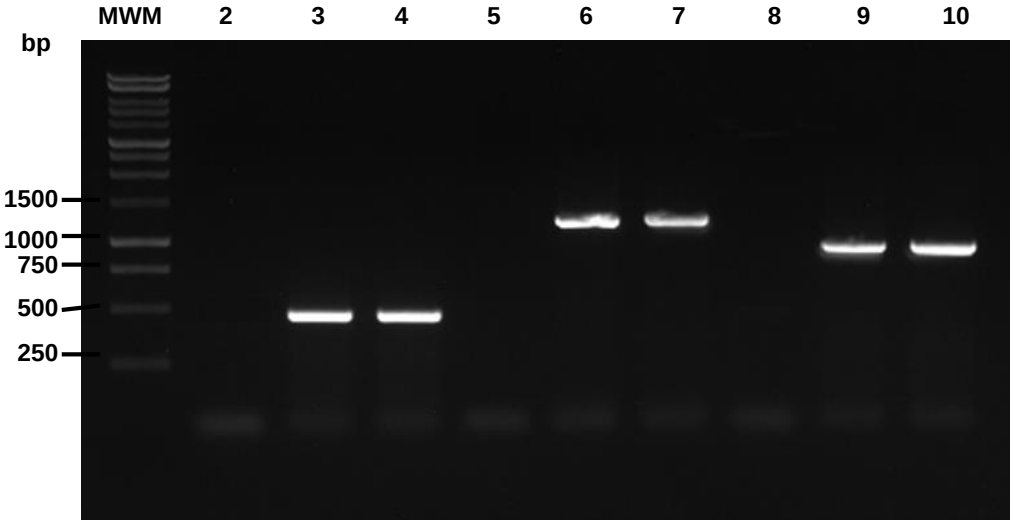

Figure 3.9 1% agarose gel displaying DNA products following PCR amplification of *histone H3, CYCD3;2* and *CYCD4;2*.Polymerase chain reaction (PCR) amplification of the genes encoding *histone H3*, *CYCD3;2* and *CYCD4;2* for cloning into pDEST^{TM}32 using the Gateway® recombination reaction. Lanes 2-4 show the non-template control (NTC) of the PCR reaction of *histone H3.* Lanes 5-7 show the NTC and PCR of *CYCD3;2*. Lanes 8-10 show the NTC and PCR amplification of *CYCD4;2*. The molecular weight marker (MWM) is the Promega 1kB DNA ladder (Promega, Madison, USA) and sizes of DNA standards are shown.

Successfully amplified *att*B-PCR products were cloned into pDONR^{TM}221 using the BP recombination reaction method. The LR recombination reaction was carried out to transfer the genes of interest into the prey vector, pDEST^{TM}32. To analyse transformants, total plasmid was extracted using GeneJET plasmid miniprep kit (Thermo Fisher Scientific, Waltham, USA) and quantification of plasmids was done using NanoDrop 1000 (Thermo Fisher Scientific, Waltham, USA). The vectors were subjected to a restriction digest. Fast digest^{TM} *Xho*I (Thermo Fisher Scientific, Waltham, USA) was used to digest pEXP^{TM}32/H3 and pEXP^{TM}32/CYCD3;2, with Fast digest^{TM} *Hind*III (Thermo Fisher Scientific, Waltham, USA) used to cut pEXP^{TM}32/CYCD4;2. The banding pattern of the restriction digest of pEXP^{TM}32/CYCD3;2 was indicative of two DNA products, 9974 bp and 1706 bp. The

restriction digests of pEXP^{TM}32/H3 and pEXP^{TM}32/CYCD4;2 resulted in DNA products of 9181 bp and 1 706 bp as well as 10 466 bp and 926 bp, respectively (Figure 3.10).

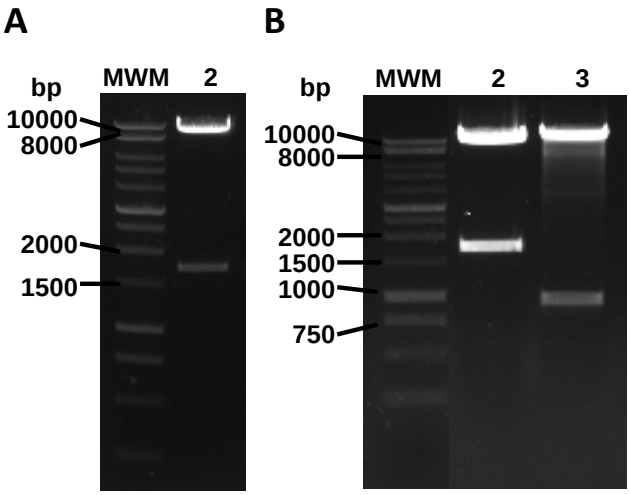

Figure 3.10 Restriction digest of pDEST^{TM}32 containing prey genes. **A** 1% agarose gel shows the DNA products of a restriction digest of pEXP^{TM}32/CYCD3;2 with Fast digest^{TM} *Xho*I (Thermo Fisher Scientific, Waltham, USA). **B** Restriction digest of pEXP^{TM}32/H3 and pEXP^{TM}32/CYCD4;2 using Fast digest^{TM} *Xho*I (Thermo Fisher Scientific, Waltham, USA) and Fast digest^{TM} *Hind*III (Thermo Fisher Scientific, Waltham, USA) in lane two and three, respectively. Molecular weight marker is shown in the first lane of **A** and **B**, with Promega 1kB DNA ladder (Promega, Madison, USA) the DNA standard in both gels.

The prey plasmids containing either *Histone H3*, *CYCD3;2* or *CYCD4;2* were sequenced (Inqaba Biotech, Pretoria) to confirm the presence thereof (see Appendix).

The combinations of bait and prey plasmids, as listed in Tables 3.8 and 3.9 (Section 3.3.3) were transformed into *S. cerevisiae* MAV203, and selected for on dropout media without leucine and tryptophan. Only those yeast cells containing both the respective bait and prey plasmids, encoding for leucine and tryptophan, respectively, grew on selective media.

3.4.3 Yeast two-hybrid assay

The master plate for each interaction was streaked with controls 2-6 and one each of controls 7.1-7.4 (Table 3.8), depending which interaction was being examined. The master plate for the interaction contained the recombinant yeast with AC1 present in the bait vector as well as the reverse, AC1 being present in the prey vector, with each prey gene of interest being present in the corresponding other vector (Table 3.9). Master plates were generated for all three interactions, Rep-H3, Rep-CYCD3;2, Rep-CYCD4;2 (Figure 3.11-3.13). An interaction of SACMV AC1-encoded Rep with itself was carried out (Figure 3.14).

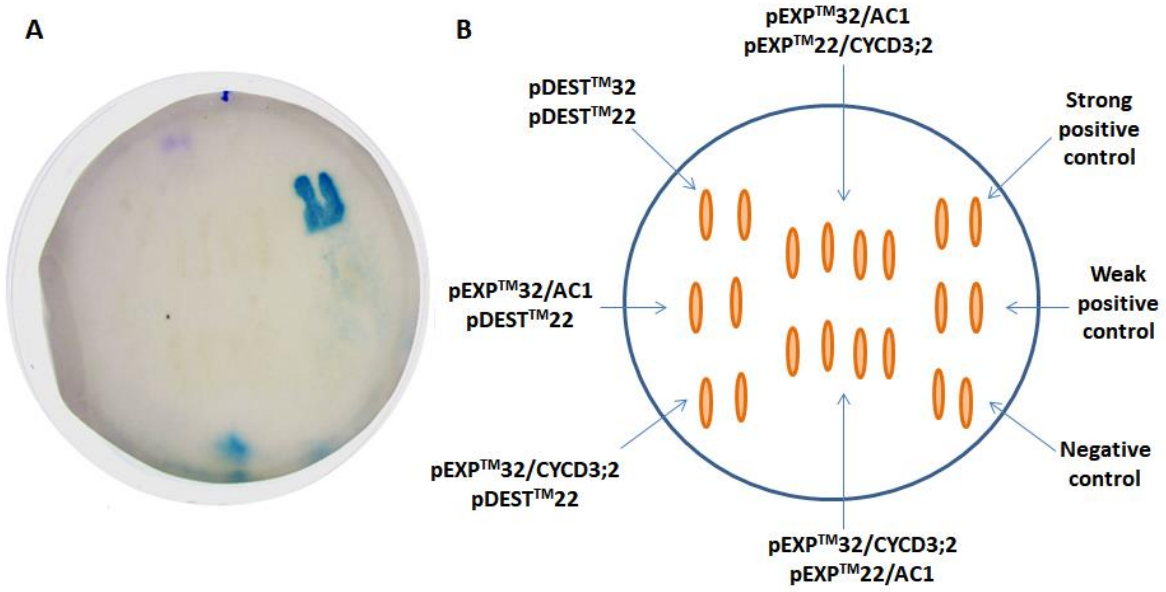

Figure 3.11 YTH assay of Rep-CYCD3;2. **A** Yeast two-hybrid (YTH) of SACMV Rep-CYCD3;2 interaction with relevant controls obtained after incubation in the presence of X-gal for 24 h. **B** Schematic representation of the master plate showing experimental and control recombinant *Saccharomyces cerevisiae* MAV203, with the streaking pattern for each control/interaction indicated in orange.

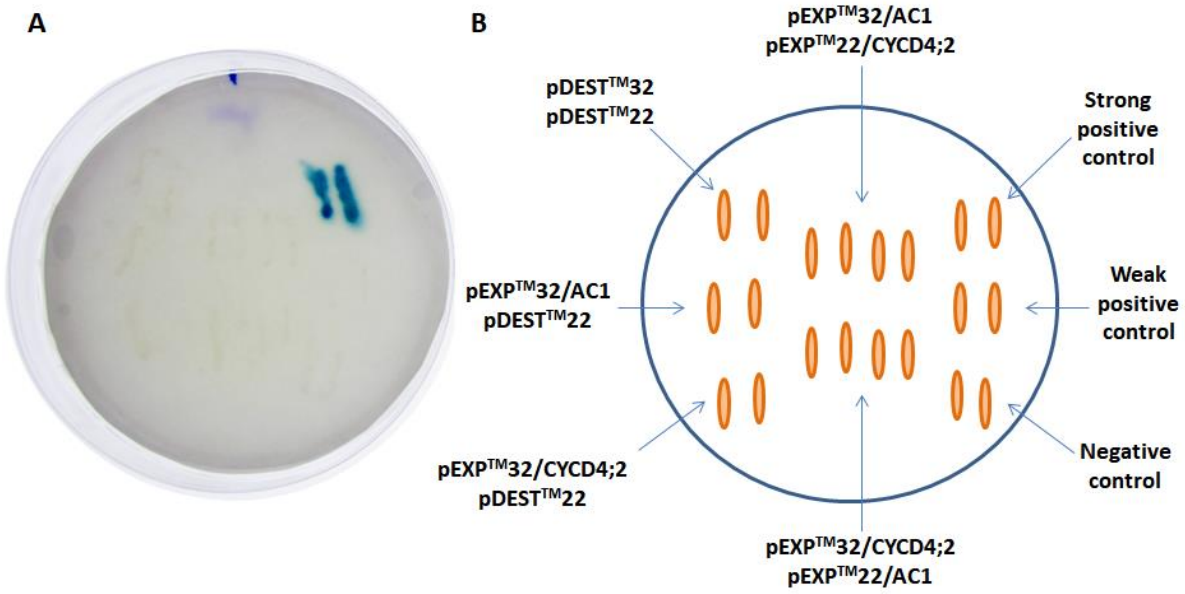

Figure 3.12 YTH assay of Rep-CYCD4;2. **A** Yeast two-hybrid (YTH) of SACMV Rep-CYCD4;2 interaction with relevant controls obtained after incubation in the presence of X-gal for 24 h. **B** Schematic representation of the master plate showing experimental and control recombinant *Saccharomyces cerevisiae* MAV203, with the streaking pattern for each control/interaction indicated in orange.

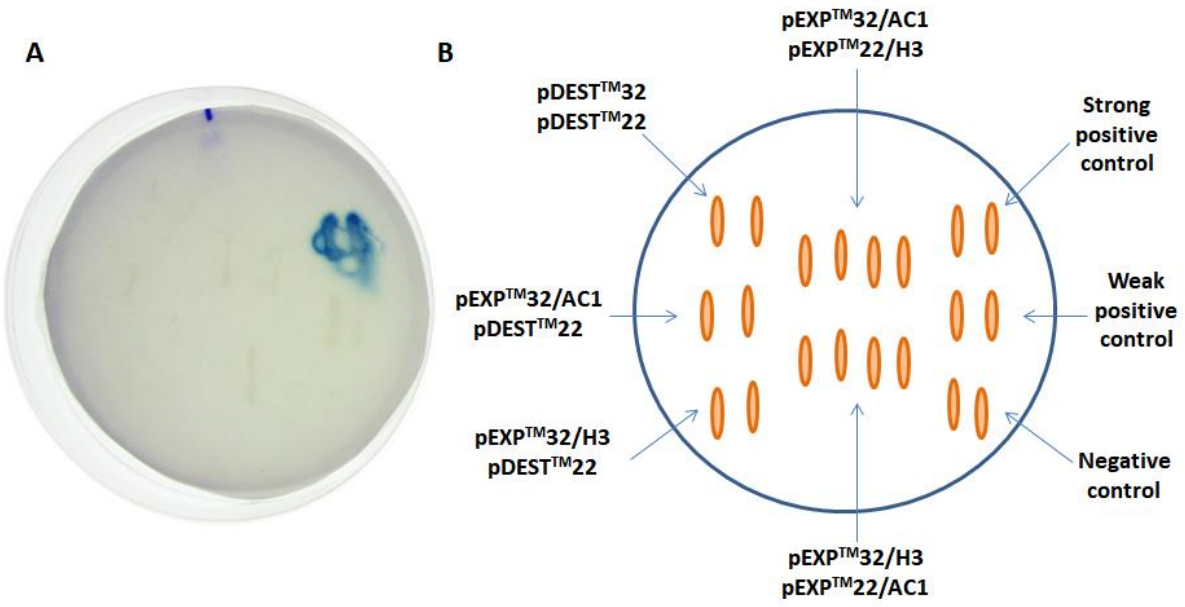

Figure 3.13 YTH assay of Rep-H3. **A** Yeast two-hybrid (YTH) of SACMV Rep-H3 interaction with relevant controls obtained after incubation in the presence of X-gal for 24 h. **B** Schematic representation of the master plate showing experimental and control recombinant *Saccharomyces cerevisiae* MAV203, with the streaking pattern for each control/interaction indicated in orange.

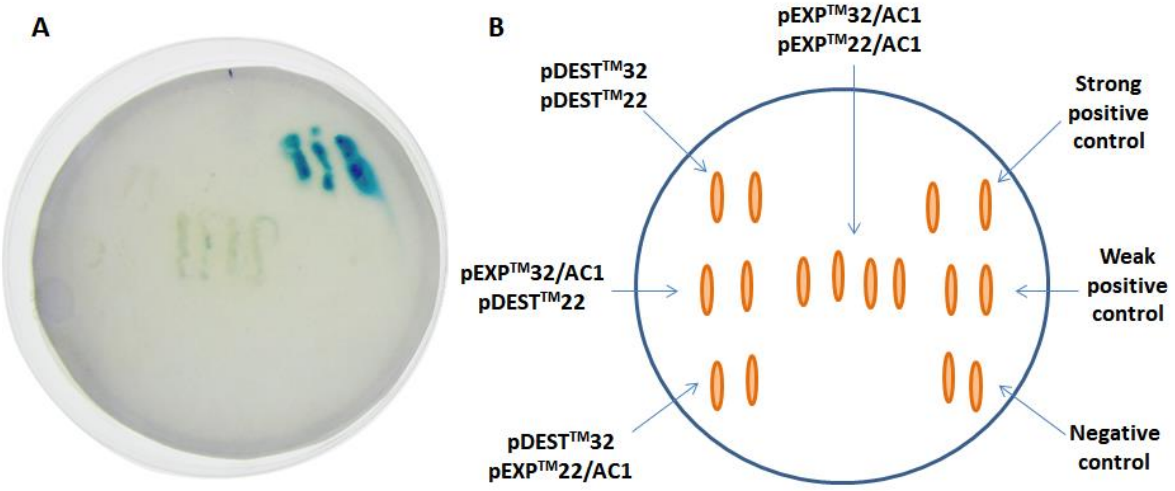

Figure 3.14 YTH assay of Rep-Rep. **A** Yeast two-hybrid (YTH) of SACMV Rep-SACMV Rep interaction with relevant controls obtained after incubation in the presence of X-gal for 24 h. **B** Schematic representation of the master plate showing experimental and control recombinant *Saccharomyces cerevisiae* MAV203, with the streaking pattern for each control/interaction indicated in orange.

The interactions between SACMV Rep and the chosen prey gene products yielded a negative interaction as observed through YTH assay. The YTH positive control (control 2) reacted positively as indicated by the development of blue colour after transfer and reaction with X-gal (Figures 3.11–3.14). The weak positive control (control 3) did not develop a faint blue

colour in all the YTH assays except for the SACMV Rep-CYCD3;2 YTH assay (Figure 3.11 A). The YTH assay carried out using SACMC AC1-encoded Rep expressed from both bait and prey plasmids, thus interacting with itself, suggesting a weakly positive interaction as seen by the presence of faint blue colour (Figure 3.14). This suggests that the protein interacts with itself, although verification of the suspected interaction would need to be confirmed using other techniques.

3.5 Discussion

In this study, the interaction of the Rep protein with three different cassava host proteins was evaluated using the YTH assay.

The two cyclin proteins, CYCD3;2 and CYCD4;2 (Allie et al., 2014; Goodstein et al., 2012), were selected as potential interactors in this study. The cassava cyclin D2, a homologue of *A. thaliana* cyclin D4;2 was identified in the genome annotation database (Phytozome). It should be noted that cyclin D2 and cyclin D4 annotations are interchangeable as they are classified under the same cyclin group (Kono et al., 2007). The cassava homologue is thus referred to as cyclin D4 in this study. Selection of these cyclins as candidate interactors was based on the fact that they are involved in cell cycle regulation in leaf tissue. These classes of the D-type cyclins have been described as the main driving force for the transition between G_1 and S-phase (Menges, 2006), and geminiviruses induce this transition in order to replicate (Hanley-Bowdoin et al., 2004; Nagar et al., 2002). While differential expression is not necessarily a hallmark of direct protein-protein interaction, transcriptomics data generated through CaLCuV infected *A. thaliana* (Ascencio-Ibáñez et al., 2008), and SACMV infected *A. thaliana* and cassava (Allie et al., 2014; Pierce and Rey, 2013) also demonstrated differential expression of these two genes. Histone 3, besides interacting with Rep (Kong and Hanley-Bowdoin, 2002), was also selected as up-regulation of a histone H3 homologue in SACMV-infected tolerant cassava TME3 (Allie et al., 2014) was noted, suggesting a possible role in retardation of SACMV replication or transcription.

The lack of an interaction of SACMV Rep and the cyclin proteins must be scrutinised. Since Rep is involved in cell cycle regulation in leaf cells, and binds to RBR (Ach et al., 1997), a putative interaction between cyclins, Rep and RBR was hypothesised. Since direct binding of CYCD3;2 and CYCD4;2 to Rep was not observed, the results from this study suggest that SACMV Rep binding is only to RBR, or that other intermediate host protein(s) are involved in the binding complex. The limitations of YTH assays are that they only detect the interaction of two proteins (Brückner et al., 2009). Additional affinity purification and mass spectrometry studies (Wang et al., 2017) could to be undertaken to confirm the results from this study.

Despite SACMV Rep protein containing a cyclin interaction motif, Rep has not previously been shown to bind to these cell cycle regulators. As suggested previously, geminivirus Rep may not interact directly with host plant protein cyclins (Hipp et al., 2014; Kong and Hanley-

Bowdoin, 2002). Rep and cyclin proteins may form a complex, by binding to additional proteins such as CDKs. The YTH assay is limited as it relies on a binary interaction to reveal a positive result, which may omit other protein binding partners (Brückner et al., 2009). As suggested by Hipp et al. (2014), begomovirus Rep proteins lacking an RBR binding motif (LXCXE) may be inhibiting RBR-E2F interaction through other mechanisms. The Rep-Cyclin/CDK complex may enable geminivirus Rep protein to interact and phosphorylate RBR thereby inhibiting the RBR-E2F complex. This would in turn facilitate the transcription of replication-associated genes regulated by the E2F transcription factor, providing the geminivirus replication factors necessary for viral replication (Ach et al., 1997).

The involvement of other proteins, directly or indirectly, in Rep-cyclin binding can be further explored in context of the cell cycle. It is possible that Rep down-regulates the expression of CYCD3;2, as this cyclin is involved in the RBR-E2F regulatory network, and promotes the G_1 to S phase transition (Ascencio-Ibáñez et al., 2008; Hanley-Bowdoin et al., 2004; Kono et al., 2007). Furthermore, the cyclin D3 family and associated CDKs are involved in inhibiting the endocycle (Dewitte et al., 2007), a cellular state that geminiviral infection is known to induce (Ascencio-Ibáñez et al., 2008; Nagar et al., 2002). Ascencio-Ibáñez et al (2008) has suggested that down-regulation of CYCD3 would enable E2F and RBR to facilitate the entering of the cell into the endocycle. It is speculated that up-regulation of CYCD4 may also be mediated by Rep. CYCD4 is distinct in sequence, as it lacks the RBR binding site and may be interacting with other transcription factors, which has been revealed for the cyclin D4 family in animal cells (Kono et al., 2007). Hipp et al. (2014) suggested that Rep influence the cell cycle through other means, other than RBR binding, especially for begomoviruses that lack the RBR binding motif. It can be speculated that the binding of Rep to cyclin D4 would curtail the need for Rep to bind to RBR. The up-regulated cyclin D4;2 may bind to transcription factors, influencing expression of cell cycle related genes, and facilitating geminivirus infection and cell cycle modification. Further binding studies would have to be carried out to confirm the hypothesis mentioned above.

The E1 protein of the family *Papillomaviridae* is described as the replication initiation protein, and has been shown to share structural and functional similarities to geminivirus Rep proteins (Campos-Olivas et al., 2002; George et al., 2014). The E1 Rep protein of human papillomavirus (HPV) has been shown to interact with animal cyclin E and be phosphorylated by the cyclin E/CDK complex. The interaction was located to the RXL motif and was deemed necessary for successful viral replication *in vivo* (Ma et al., 1999). Due to

the conserved nature of DNA replication initiation and helicase domains across ssDNA viruses (Campos-Olivas et al., 2002; George et al., 2014; Nawaz-ul-Rehman and Fauquet, 2009), it can be suggested that geminivirus Rep proteins also bind to cyclin E/CDK complexes. The reasons for the complex formation may differ between the two virus families, but binding of geminivirus Rep to cyclin E/CDK complexes may offer an alternative method to RBR inactivation. The study mentioned that cyclin D proteins were shown not to interact with E1 (Ma et al., 1999). However, as this study was carried out using a human proteins, cyclin E in human cells is part of the regulation of the E2F/RBR pathway (Oakenfull et al., 2002), not cyclin D, as shown in plant cells (Dewitte and Murray, 2003). The RXL motif of E1 is located in the N-terminus, between residues 121-131, whereas the putative cyclin binding motif of geminiviruses is predicted to lie in the C-terminus, adjacent to the Walker A motif. Taking these findings into account, further investigation with SACMV Rep and cyclins is warranted.

In a study conducted by Kong and Hanley-Bowdoin (2002) full length TGMV and CaLCuV Rep protein (encoded for by ORF AL1) was found to interact with three host proteins following a Gal4-based two-hybrid system (Matchmaker^{TM} GAL4 Two-Hybrid system) using an *A. thaliana* cDNA library. These were identified as the Geminivirus Rep-interacting motor protein (GRIMP), Geminivirus Rep interacting kinase (GRIK) and a histone H3 protein.

The putative Rep-H3 interaction was intended to serve as a positive interaction control in this study to replicate the results of earlier studies. While up-regulation of a histone H3 in SACMV-infected tolerant cassava TME3 (Allie et al., 2014) suggested a possible role in retardation of SACMV replication or transcription, negative results for the histone H3 interaction with SACMV Rep were obtained. The negative result of the Rep-H3 interaction was not expected since the interaction of TGMV Rep and CaLCuV Rep was identified with *A. thaliana* histone H3 on triple dropout media (SD-Leu-Trp-His) (Kong and Hanley-Bowdoin, 2002). Interestingly, TGMV and CaLCuV Rep were not shown to interact with *A. thaliana* RBR in an earlier YTH analysis (Kong et al., 2000), and RBR is known to interact with Rep as shown in other studies (Ach et al., 1997; Arguello-Astorga et al., 2004; Kong et al., 2000) suggesting a possible false negative result. The histone H3 interaction with TGMV and CaLCuV Rep was not validated using protein immunolocalisation in the study conducted by Kong and Hanley-Bowdoin (2002), and other methods such as affinity purification-MS and antibody immunoprecipitation could be attempted in the future. The negative result obtained by the SACMV Rep-H3 interaction also does not necessarily indicate

that the bait or prey fusion proteins were indirectly folded or inactive. In this study a full length cassava histone H3 gene was chosen, instead of the truncated gene identified in the YTH carried out by Kong and Hanley Bowdoin (2002), which may have influenced the results. Most importantly the host genes used in this study were chosen from the annotated cassava genome available through Phytozome. The protein interactions taking place between virus proteins and host plant proteins may differ greatly between cassava and *A. thaliana*, despite it being a model organism. The interaction of SACMV Rep with cassava histone H3 may not occur, or may require additional host proteins. The validity of the YTH assay results were called into question by Kong and Hanley-Bowdoin (2002) due to TGMV and CaLCuV Rep not interacting with *A. thaliana* RBR, as has been shown previously with YTH analysis (Kong et al., 2000).

The negative result obtained by the YTH assay investigating the Rep-H3 interaction may be explained by the absence of an additional binding partner. The role of Rep-H3 binding, as suggested by Kong and Hanley-Bowdoin (2002), was to displace nucleosomes, to enable binding of Rep to the viral origin of replication. A proposed protein complex that would yield a positive YTH result may require the inclusion of the geminiviral origin of replication. It was suggested that Rep binds to the origin (Fontes et al., 1992), and then recruits histone H3, implying the DNA-protein interaction as a requirement for binding of Rep-H3 (Kong and Hanley-Bowdoin, 2002). The binding of Rep to histone H3 was also hypothesised to prevent methylation of the lysine 9 (Raja et al., 2008). The involvement of methylation machinery in the Rep-H3 reaction could, therefore, be proposed.

The interaction of Rep with itself was suggested by the positive result obtained by the YTH assay, as Rep was expressed as both the bait, DBD-Rep, and as the prey, AD-Rep. A faint blue colour was noted, suggesting an interaction between Rep proteins and indicating that the DBD and AD fusion proteins were able to interact and lead to reporter expression (Figure 3.14). This was as expected, as geminivirus Rep proteins have been shown to form high order oligomers (Choudhury et al., 2006; Clérot and Bernardi, 2006; George et al., 2014). However, the weak signal obtained from the Rep-Rep interaction was unusual as a stronger interaction was expected, as seen in previous YTH assays with MYMIV (Genus *Begomovirus*) Rep (Suyal et al., 2013, 2014). A weak interaction may be explained by the transience of the interaction, or a weak affinity of the binding partners for each other. Furthermore, the fusion tags may be contributing to incomplete folding of the fusion proteins, DBD-Rep and AD-Rep, resulting in the weak interaction (Rao et al., 2014). A weak

interaction does not completely invalidate the positive result of the YTH assay of Rep-Rep (Van Criekinge and Beyaert, 1999), but the result must be scrutinised. The controls carried out alongside the Rep-Rep interaction were used to evaluate the possibility of fusion-Rep auto-activating the reporter genes. The controls included DBD-Rep alongside an empty prey vector, and conversely AD-Rep expressed alongside an empty bait vector. As neither of these interactions yielded a positive YTH result, the Rep-Rep interaction did not arise due to auto-activation. Despite this, the feint signal might be indicative of a false positive. Further protein interaction studies would need to be carried out, to confirm Rep-Rep binding.

The high levels of false positives that YTH assays give rise to must be taken into account when analysing the results of this study. There are numerous reasons that the interactions carried out in this assay can be deemed as false negatives. The fused yeast reporter proteins, the DBD and AD, may be interfering with the interaction of the fusion bait and prey through steric hindrance (Brückner et al., 2009). The orientation of the fusion tag to the protein may also be inhibiting protein interaction (Causier and Davies, 2002). The bait and prey vectors of the ProQuest™ Two-Hybrid System (Thermo Fisher Scientific, Waltham, USA) are specifically designed so that the putative protein interactor genes are cloned so that the fusion tags are expressed at the N-terminus of the recombinant proteins. This may be inhibiting or altering the native folding of the proteins, thereby preventing the binding partners from interacting. This, in turn, interferes or inhibits the fusion tag interaction and expression of reporter genes (Suter et al., 2008).

Furthermore, false negative interactions may arise from the lack of post-translational modifications made to proteins of higher eukaryotes. This can result in the yeast-expressed fusion protein not being functional (or in its native state) thus preventing it from binding to the putative binding partner (Suter et al., 2008). Geminivirus Rep proteins have been expressed in yeast cells and have found to retain their function (Hipp et al., 2014; Raghavan et al., 2004) suggesting that the native folding of SACMV Rep in yeast is possible. As Rep expressed in yeast in the present study is a recombinant fusion protein, it cannot be ruled out that it has been expressed in its functional conformation. However, the Rep-Rep interaction was found to be weakly positive suggesting the fusion protein was folded in a conformation that enabled Rep to oligomerise. The lack of positive results for Rep-host protein interaction assays suggests that these interactions require additional protein binding partners. In the absence of additional host protein binding partners or viral DNA sequences, the interactions cannot proceed. For the putative binding partners, histone H3, CYCD3;2 and CYCD4;2, it

cannot be confirmed that these fusion-proteins folded correctly, and an absence of post-translational modifications, which are required for the expression of most higher order eukaryotic proteins, may have resulted in the plant proteins not being able to interact with the viral protein Rep. Future YTH assays may involve the addition of a modifying enzymes being co-expressed alongside the bait and prey (Brückner et al., 2009).

To further scrutinise if the system is working, and thereby eliminate chances of false positives and false negatives arising, the presence of the expressed proteins, throughout the YTH assay, could be confirmed. This could be carried out by performing a whole protein extraction from the transformed yeast, and probing for expressed proteins using prey and bait directed primary antibodies and western blotting techniques.

Future studies include the additional analysis of the interactions by evaluating expression of the other reporter genes provided by the ProQuest™ Two-Hybrid System (Thermo Fisher Scientific, Waltham, USA). Positive interactions can be confirmed by selecting for the interactors by plating on selective (dropout) media. The interaction of Rep and the putative binding partners would result in the expression of histidine or uracil, depending on which auxotrophic marker is absent from the growth medium. This would allow for the growth of cells containing positively-interacting bait-prey complexes. The three-hybrid assay could also be undertaken as it examines the potential of a third protein acting as a facilitator of protein binding (Stynen et al., 2012). An additional assay probing protein interaction in yeast includes the split luciferase system. As in the YTH approach, luciferase is split, with the N- and C-terminal fragments fused to the proteins of interest. Interaction of the fused protein binding partners results in bioluminescence due to the reaction of functional luciferase with its substrate, luciferin. This method provides a very important advantage, as the protein binding reaction is reversible, and protein interactions can be observed in real-time (Stynen et al., 2012).

Alternative methods of identifying novel interactions as well as confirming interactions existing in literature include pull down assays and co-immunoprecipitation. The advantages of these methods is in their ability to identify the constituents of a protein complex, and not only a binary, direct interactions as with YTH assays (Brückner et al., 2009; Rao et al., 2014). Pull-down assays require the immobilisation of a protein such as Rep to a resin and allowing for the binding of protein complexes present in plant cell lysate as it is passed through the resin (Wysocka, 2006). The interacting proteins could be identified through the use of mass

spectrometry (MS), which analyses the mass to charge ratio of molecules that have been ionised (Fenn et al., 1989; Tanaka et al., 1988). Co-immunoprecipitation uses antibodies to capture the bait as well as its binding partners. The putative binding partners are identified using MS. MS can be used to identify interacting proteins by using 2D gels and Matrix-assisted laser desorption ionisation time-of-flight (MALDI-TOF) analysis (Honoré et al., 2004; Kaboord and Perr, 2008). MALDI-TOF analysis is used to study protein complexes, as this technique does not break chemical bods, thereby keeping the protein complex intact (Beavis and Chait, 1996; Hillenkamp et al., 1991; Nagy, 2008).

A direct interaction of putative binding partners, as studied with YTH assays, can also be carried out using fluorescent detection of the binding partners in living cells and using FRET to detect interactions. Using fluorescently-tagged antibodies, the proteins of interest could be visualised in living cells, and the co-localisation of proteins could be monitored. Another *in vivo* method uses fluorescent tags (acceptor and donor) that intersect in their excitation/emission spectra. Using confocal microscopy, the interaction of the fluorescently tagged proteins could be monitored by emission of the acceptor fluorophore (Brückner et al., 2009). Bimolecular fluorescence complementation (BiFC) is an *in vitro* method enabling the analysis of protein interactions in living cells. The putative protein binding partners are fused to fragments of a fluorescent protein. Binding of the proteins enables the binding of the fluorescence protein fragments, transpiring in fluorescence that can be analysed with a fluorescence microscope (Kerppola, 2008).

Future work may include co-immunoprecipitation using antibodies directed against a short peptide designed off the Rep sequence and/or affinity purification using His-Tagged Rep. Yeast two-hybrid results have been confirmed through co-immunoprecipitation techniques in studies carried out on geminivirus Rep MCM using a YTH approach, and confirmed the interaction with a co-immunoprecipitation assay (Suyal et al., 2013). Using these techniques a protein complex involving Rep could be extracted, and the individual proteins identified using MS. Using the ProQuest™ Two-Hybrid System (Thermo Fisher Scientific, Waltham, USA) a direct interaction of Rep with the extracted proteins would be verified. FRET techniques analysing *in vivo* interactions of Rep with specific putative interactors could also prove useful in elucidating the evidently complex nature of geminivirus Rep-host protein interactions.

3.6 Conclusion

This study is the first to attempt to identify SACMV Rep interactors in cassava. The YTH assay carried out to establish the direct interaction of Rep with cassava host proteins histone H3, CYCD3;2 and CYCD4;2 yielded a negative result whilst the SACMV Rep-Rep YTH assay yielded a positive result. The host protein-Rep YTH assays may have yielded false negative results, which are a common occurrence in this technique. Reasons for the negative result may be attributed to the absence of a binary interaction between Rep and the chosen putative binding partners. The Rep-H3 binding interaction may require the addition of the viral origin of replication, in order for direct binding between Rep and histone H3 to occur. An interaction of Rep with the cyclin D family of proteins may also only be facilitated in the presence of additional host proteins, such as CDK. Experimental analysis using techniques that extract a protein binding complex directly from the host cell will be helpful in the assessment of the protein binding complexes that SACMV Rep forms a part of in cassava, including those of Histone H3 and cyclin D proteins. Additional analysis of Rep-binding proteins is essential for the establishment of a Rep-host protein interactome map, which will further contribute to the pathogen-plant protein interaction network.

Chapter IV – Thesis overview and concluding remarks

Cassava (*Manihot esculenta*) is the primary staple food source for many subsistence farmers in developing countries, and is one of the most significant sources of starch globally (Baguma et al., 2017). Due to its ability to grow in drought conditions and acidic soils, cassava has been labelled a food security crop, an important food crop in a changing climate (Legg et al., 2014, 2015). The crop is not only used for human consumption, but is also grown for animal feed, and plays an important role in the paper industry, as well as in the production of bioethanol (El-Sharkawy, 2004). Yields of cassava and many other economically important food crops are negatively affected by cassava mosaic disease, caused by the family *Geminiviridae* (FAOSTAT, 2014). This family of viruses is known to infect a number of economically important food crops, and have become a popular point of research due to their ability to manipulate the host plant and their single stranded DNA genome (Hanley-Bowdoin et al., 2013). One of the geminivirus species infecting cassava in sub-Saharan Africa is South African cassava mosaic virus (SACMV) (Berrie et al., 2001). The geminivirus-encoded replication associated protein (Rep) is a multifunctional protein of great importance, a critical component in facilitating plant host manipulation and viral replication, thereby enabling spread of the virus (Hanley-Bowdoin et al., 2013; Laufs et al., 1995c). A comprehensive structural picture of geminivirus Rep proteins has not been established and the full extent of Rep-host interactions has not been elucidated. Studies on SACMV Rep specifically have never been attempted, with this study providing the first look into structural and functional features, as well as aiming to establish potential cassava protein binding partners.

This study exhibited the successful expression and purification of SACMV Rep. The recombinant protein displayed DNA binding and cleavage abilities, as shown for other recombinant expressed geminivirus Rep proteins (Campos-Olivas et al., 2002; Hipp et al., 2014). Structural studies carried out on the protein confirmed the presence of ligand binding pockets, for binding of DNA, ATP or other plant host proteins. Three putative cassava host proteins were shown to not directly interact with SACMV Rep in a binary interaction, but posed further questions for their involvement in host-virus interactions.

Future work will further analyse structural and functional features of this protein, thereby contributing to the knowledge base of geminivirus Rep proteins. Additional structural studies will aim to elucidate the secondary structure and quaternary structure using size exclusion chromatography. To better understand the functional traits of geminivirus Rep proteins,

ATPase abilities of SACMV Rep will be evaluated. Crystallisation and cryo-electron microscopy may further elucidate structural features of SACMV Rep. Any successful attempts to determine the 3D structure of a geminivirus Rep will greatly expand on the knowledge of this protein, and the structure-function relationship. In addition, Rep protein structures will aid in defining the mechanisms of virus-host interactions.

Expanding on the plant-Rep protein interactome map will advance the understanding of how this protein contributes to viral pathogenicity. Experiments carried out *in vitro* (e.g. Co-immunoprecipitation) and *in vivo* (e.g. BiFC) would prove direct or indirect interaction of Rep with plant protein binding partners (Rao et al., 2014). Due to the complexities of plant-microbe interactions within the plant cell, elucidation of virus and host protein binding complexes will expose the intricacies of a number of processes (Garbutt et al., 2014). Processes involving viral pathogenicity, cell cycle manipulation and host cell modifications can be revealed. Understanding how the plant responds to infection can be further elucidated. Insight into the complex interaction network of plant-virus proteins can lead to novel pathogen control steps being investigated.

Some efforts to curb viral replication by interfering with the functionality of Rep have been carried out to date. Peptide aptamers have been implemented to introduce plant virus resistance (Reyes et al., 2013). Truncated Rep proteins and artificial Rep binding proteins have been used in transgenic *Arabidopsis* studies, functioning by inhibiting Rep-mediated viral replication and establishing viral resistance (Ruhel and Chakraborty, 2018). Further structural studies of Rep would enhance these plant resistance efforts. Comprehensive knowledge of Rep elucidated through quaternary structure analysis and 3D structural investigation would assist in developing and implementing more robust and long lasting biotechnology-based approaches to establish plant resistance. Due to the conserved nature of Rep (Nash et al., 2011), identification of additional geminivirus specific motifs would provide more targets for broad spectrum virus resistance efforts (Reyes et al., 2013).

Implications of this work, in combination with future investigations, will contribute to the knowledge base of all geminiviruses Rep proteins. Due to geminivirus Rep being highly conserved (Nash et al., 2011), any structural and functional information elucidated can assist in the development of plant virus defence strategies for any crops affected by this plant pathogen. Ultimately, expanding the understanding of geminiviruses will aid in combatting

this destructive plant pathogen, influencing social and economic stability in developing regions of the world (Jeske, 2009).

References

Ach, R.A., Durfee, T., Miller, A.B., Taranto, P., Hanley-Bowdoin, L., Zambryski, P.C., and Gruissem, W. (1997). RRB1 and RRB2 encode maize retinoblastoma-related proteins that interact with a plant D-type cyclin and geminivirus replication protein. Mol. Cell. Biol. *17*, 5077–5086.

Achilonu, I., Siganunu, T.P., and Dirr, H.W. (2014). Purification and characterisation of recombinant human eukaryotic elongation factor 1 gamma. Protein Expr. Purif. *99*, 70–77.

Allie, F., Pierce, E.J., Okoniewski, M.J., and Rey, C. (2014). Transcriptional analysis of South African cassava mosaic virus -infected susceptible and tolerant landraces of cassava highlights differences in resistance , basal defense and cell wall associated genes during infection. BMC Genomics *15*, 1–30.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

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

Andrade, M.A., Chacón, P., Merelo, J.J., and Morán, F. (1993). Evaluation of secondary structure of proteins from uv circular dichroism spectra using an unsupervised learning neural network. Protein Eng. Des. Sel. *6*, 383–390.

Arguello-Astorga, G.R., and Ruiz-Medrano, R. (2001). An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. Arch. Virol. *146*, 1465–1485.

Arguello-Astorga, G., Lopez-Ochoa, L., Kong, L.-J., Orozco, B.M., Settlage, S.B., and Hanley-Bowdoin, L. (2004). A novel motif in geminivirus replication proteins interacts with the plant retinoblastoma-related protein. J. Virol. *78*, 4817–4826.

Arguello-Astorga, G.R., Guevara-Gonzalez, R.G., Herrera-Estrella, L.R., and Rivera-Bustamante, R.F. (1994). Geminivirus Replication Origins Have a Group-Specific Organization of Iterative Elements: A model for Replication. Virology *203*, 90–100.

Argüello-Astorga, G., Herrera-Estrella, L., and Rivera-Bustamante, R. (1994). Experimental

and theoretical definition of geminivirus origin of replication. Plant Mol. Biol. *26*, 553–556.

Arnau, J., Lauritzen, C., Petersen, G.E., and Pedersen, J. (2006). Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expr. Purif. *48*, 1–13.

Ascencio-Ibáñez, J.T., Sozzani, R., Lee, T.-J., Chu, T.-M., Wolfinger, R.D., Cella, R., and Hanley-Bowdoin, L. (2008). Global analysis of Arabidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. Plant Physiol. *148*, 436–454.

Auerbach, D., Thaminy, S., Hottiger, M.O., and Stagljar, I. (2002). Review The post-genomic era of interactive proteomics : Proteomics *2*, 611–623.

Bagewadi, B., Chen, S., Lal, S.K., Choudhury, N.R., and Mukherjee, S.K. (2004). PCNA interacts with Indian mung bean yellow mosaic virus rep and downregulates Rep activity. J. Virol. *78*, 11890–11903.

Baguma, Y., Nuwamanya, E., and Rey, M.E.C. (2017). The African Perspective: Developing an African bio-resource based industry: the case for cassava. In Creating Sustainable Bioeconomies: The Bioscience Revolution in Europe and Africa, I. Virgin, and E.J. Morris, eds. (Routledge (Taylor and Francis Group) publishers).

Baneyx, F. (1999). Recombinant protein expression in Escherichia coli. Curr. Opin. Biotechnol. *10*, 411–421.

Baumann, C.G., Matulis, D., Baumann, C.G., Bloomfield, V.A., and Lovrien, R.E. (1998). 1-Anilino-8-naphthalene Sulfonate as a Protein Conformational Tightening Agent Sulfonate as a Protein Conformational Tightening. Biopolymer *49*, 451–458.

Beavis, R.C., and Chait, B.T. (1996). Matrix-assisted laser desorption ionization mass-spectrometry of proteins. Methods Enzymol. *270*, 516–551.

Berrie, L.C., Rybicki, E.P., and Rey, M.E. (2001). Complete nucleotide sequence and host range of South African cassava mosaic virus: further evidence for recombination amongst begomoviruses. J Gen Virol *82*, 53–58.

Berry, S., and Rey, M.E.C. (2001). Molecular evidence for diverse populations of cassava-infecting begomoviruses in southern Africa . Arch. Virol. *146*, 1795–1802.

Bondos, S.E., and Bicknell, A. (2003). Detection and prevention of protein aggregation before, during, and after purification. Anal. Biochem. *316*, 223–231.

Brahms, S., Brahms, J., Spach, G., and Brack, A. (1977). Identification of beta,beta-turns and unordered conformations in polypeptide chains by vacuum ultraviolet circular dichroism. Proc. Natl. Acad. Sci. U. S. A. *74*, 3208–3212.

Brückner, A., Polge, C., Lentze, N., Auerbach, D., and Schlattner, U. (2009). Yeast two-hybrid, a powerful tool for systems biology. Int. J. Mol. Sci. *10*, 2763–2788.

Burns, A., Gleadow, R., Cliff, J., Zacarias, A., and Cavagnaro, T. (2010). Cassava: The Drought, War and Famine Crop in a Changing World. Sustainability *2*, 3572–3607.

Campos-Olivas, R., Louis, J.M., Clerot, D., Gronenborn, B., and Gronenborn, A.M. (2002). The structure of a replication initiator unites diverse aspects of nucleic acid metabolism. Proc. Natl. Acad. Sci. USA *99*, 10310–10315.

Carter, S.E., Fresco, L.O., Jones, P.G., and Fairbairn, J.N. (1997). Introduction and diffusion of cassava in Africa. In IITA Research Guide 49, p. 34.

Castillo, A.G., Collinet, D., Deret, S., Kashoggi, A., and Bejarano, E.R. (2003). Dual interaction of plant PCNA with geminivirus replication accessory protein (Ren) and viral replication protein (Rep). Virology *312*, 381–394.

Castillo, A.G., Kong, L.J., Hanley-Bowdoin, L., and Bejarano, E.R. (2004). Interaction between a geminivirus replication protein and the plant sumoylation system. J. Virol. *78*, 2758–2769.

Cattoni, D.I., Kaufman, S.B., and González Flecha, F.L. (2009). Kinetics and thermodynamics of the interaction of 1-anilino-naphthalene-8-sulfonate with proteins. Biochim. Biophys. Acta *1794*, 1700–1708.

Causier, B., and Davies, B. (2002). Analysing protein-protein interactions with the yeast two-hybrid system. Plant Mol. Biol. *50*, 855–870.

Ceniceros-Ojeda, E.A., Rodríguez-Negrete, E.A., and Rivera-Bustamante, R.F. (2016). Two Populations of Viral Minichromosomes Are Present in a Geminivirus-Infected Plant Showing Symptom Remission (Recovery). J. Virol. *90*, 3828–3838.

Chatterji, A., Chatterji, U., Beachy, R.N., and Fauquet, C.M. (2000a). Sequence Parameters That Determine Specificity of Binding of the Replication-Associated Protein to Its Cognate Site in Two Strains of Tomato Leaf Curl Virus – New Delhi. Virology *350*, 341–350.

Chatterji, A., Chatterji, U., Beachy, R.N., and Fauquet, C.M. (2000b). Sequence parameters that determine specificity of binding of the replication-associated protein to its cognate site in two strains of tomato leaf curl virus-New Delhi. Virology *273*, 341–350.

Chen, D., and Texada, D.E. (2006). Low-usage codons and rare codons of Escherichia coli. Gene Ther. Mol. Biol. *10*, 1–12.

Chen, J.W., Romero, P., Uversky, V.N., and Dunker, A.K. (2008). Conservation of Intrinsic Disorder in Protein Domains and Families: I. A Database of Conserved Predicted Disordered Regions. J Proteome Res. *5*, 879–887.

Choudhury, N.R., Malik, P.S., and Singh, D.K. (2006). The oligomeric Rep protein of Mungbean yellow mosaic India virus (MYMIV) is a likely replicative helicase. Nucleic Acids Res. *34*, 6362–6377.

Chowda-Reddy, R.V., Achenjang, F., Felton, C., Etarock, M.T., Anangfac, M.T., Nugent, P., and Fondong, V.. (2008). Role of a geminivirus AV2 protein putative protein kinase C motif on subcellular localization and pathogenicity. Virus Res. *135*, 115–124.

Clara, S. (1999). Functional genomics. Proc. Natl. Acad. Sci. U. S. A. *96*, 8825–8826.

Clérot, D., and Bernardi, F. (2006). DNA Helicase Activity Is Associated with the Replication Initiator Protein Rep of Tomato Yellow Leaf Curl Geminivirus. J. Virol. *80*, 11322–11330.

Van Criekinge, W., and Beyaert, R. (1999). Yeast Two-Hybrid: State of the Art. Biol. Proced. Online *2*, 1–38.

Czosnek, H., Ghanim, M., Morin, S., Rubinstein, G., Fridman, V., and Zeidan, M. (2001). Whiteflies: Vector, and Victims of Geminiviruses. Adv. Virus Res. *57*, 291–322.

Demain, A.L., and Vaishnav, P. (2009). Production of recombinant proteins by microbes and higher organisms. Biotechnol. Adv. *27*, 297–306.

Desbiez, C., David, C., Mettouchi, A., Laufs, J., and Gronenborn, B. (1995). Rep protein of

tomato yellow leaf curl geminivirus has an ATPase activity required for viral DNA replication. Proc. Natl. Acad. Sci. USA *92*, 5640–5644.

Dewitte, W., and Murray, J.A. (2003). The plant cell cycle. Annu. Rev. Plant Biol. *54*, 235–264.

Dewitte, W., Scofield, S., Alcasabas, A.A., Maughan, S.C., Menges, M., Braun, N., Collins, C., Nieuwland, J., Prinsen, E., Sundaresan, V., et al. (2007). Arabidopsis CYCD3 D-type cyclins link cell proliferation and endocycles and are rate-limiting for cytokinin responses. Proc. Natl. Acad. Sci. *104*, 14537–14542.

Egelkrout, E.M., Robertson, D., and Hanley-bowdoin, L. (2001). Proliferating Cell Nuclear Antigen Transcription Is Repressed through an E2F Consensus Element and Activated by Geminivirus Infection in Mature Leaves. Plant Cell *13*, 1437–1452.

El-Sharkawy, M.A. (2004). Cassava biology and physiology. Plant Mol. Biol. *56*, 481–501.

Enemark, E.J., and Joshua-Tor, L. (2006). Mechanism of DNA translocation in a replicative hexameric helicase. Nature *442*, 270–275.

Fairbanks, G., Steck, T.L., and Wallach, D.F.H. (1971). Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane. Biochemistry *10*, 2606–2617.

FAOSTAT (2014). Rome, Italy: Food and Agriculture Organisation of the United Nations.

Fenn, J.B., Mann, M., Meng, C.K.A.I., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for Mass Spectrometry of Large Biomolecules. Science. *246*, 64–71.

Fields, S. (2005). High-throughput two-hybrid analysis The promise and the peril. FEBS J. *272*, 5391–5399.

Fields, S., and Song, O.K. (1989). A novel genetic system to detect protein–protein interactions. Nature *340*, 245–246.

Flood, J. (2010). The importance of plant health to food security. Food Secur. *2*, 215–231.

Fondong, V.N. (2013). Geminivirus protein structure and function. Mol. Plant Pathol. *14*, 635–649.

Fondong, V.N. (2017). The Search for Resistance to Cassava Mosaic Geminiviruses : How

Much We Have Accomplished , and What Lies Ahead. Front. Plant Sci. *8*, 1–19.

Fonseca, G.J., Cohen, M.J., Nichols, A.C., Barrett, J.W., and Mymryk, J.S. (2013). Viral Retasking of hBre1 / RNF20 to Recruit hPaf1 for Transcriptional Activation. PLOS Pathog. *9*, 1–11.

Fontes, E.P., Luckow, V.A., and Hanley-Bowdoin, L. (1992). A Geminivirus Replication Protein Is a Sequence-Specific DNA Binding Protein. Plant Cell *4*, 597–608.

Fontes, E.P.B., Gladfelter, H.J., Schaffer, R.L., Lan, B., Petty, T.D., and Hanley-Bowdoin, L. (1994a). Geminivirus Replication Origins Have a Modular Organization. Plant Cell *6*, 405–416.

Fontes, E.P.B., Eagle, P.A., Sipe, P.S., Luckow, V.A., and Hanley-Bowdoin, L. (1994b). Interaction between a geminivirus replication protein and origin DNA is essential for viral replication. J. Biol. Chem. *269*, 8459–8465.

Freire, E., Mayorga, O.L., and Straume, M. (1990). Isothermal Titration. Anal. Chem. *62*, 950A–959A.

Garbutt, C.C., Bangalore, P. V, Kannar, P., and Mukhtar, M.S. (2014). Getting to the edge: protein dynamical networks as a new frontier in plant-microbe interactions. Front. Plant Sci. *5*, 1–6.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocol Handbook, J.M. Walker, ed. (Totowa, NJ: Humana Press), pp. 571–607.

Gasymov, O.K., and Glasgow, B.J. (2007). ANS fluorescence: potential to augment the identification of the external binding sites of proteins. Biochim. Biophys. Acta *1774*, 403–411.

George, B., Ruhel, R., Mazumder, M., Sharma, V.K., Jain, S.K., Gourinath, S., and Chakraborty, S. (2014). Mutational analysis of the helicase domain of a replication initiator protein reveals critical roles of Lys 272 of the B’ motif and Lys 289 of the β-hairpin loop in geminivirus replication. J. Gen. Virol. *95*, 1591–1602.

Ghisaidoobe, A., and Chung, S. (2014). Intrinsic Tryptophan Fluorescence in the Detection and Analysis of Proteins: A Focus on Förster Resonance Energy Transfer Techniques. Int. J.

Mol. Sci. *15*, 22518–22538.

Gill, S.C., and von Hippel, P.H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem. *182*, 319–326.

Goldstein, L.S.B., and Philp, A. V. (1999). The Road Less Traveled: Emergin Principles of Kinesin Motor Utilization. Annu. Rev. Cell Dev. Biol. *15*, 141–183.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., et al. (2012). Phytozome: A comparative platform for green plant genomics. Nucleic Acids Res. *40*, 1178–1186.

Gorbalenya, A.E., and Koonin, E. V (1993). Helicases : amino acid sequence comparisons and structure-function relationships. Curr. Opin. Struct. Biol. *3*, 419–429.

Gorbalenya, A.E., Koonin, E. V., and Wolf, Y.I. (1990). A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. FEBS Lett. *262*, 145–148.

Gräslund, S., Nordlund, P., Weigelt, J., Hallberg, B.M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., et al. (2008). Protein production and purification. Nat. Methods *5*, 135–146.

Gutierrez, C. (1999). Geminivirus DNA replication. Cell. Mol. Life Sci. *56*, 313–329.

Gutzat, R., Borghi, L., and Gruissem, W. (2012). Emerging roles of RETINOBLASTOMA-RELATED proteins in evolution and plant development. Trends Plant Sci. *17*, 139–148.

Hanley-Bowdoin, L., Settlage, S.B., Orozco, B.M., Nagar, S., and Robertson, D. (1999). Geminiviruses: models for plant DNA replication, transcription, and cell cycle regulation. Critical Reviews in Plant Science. *18*, 71-106.

Hanley-Bowdoin, L., Settlage, S.B., and Robertson, D. (2004). Reprogramming plant gene expression: a prerequisite to geminivirus DNA replication. Mol. Plant Pathol. *5*, 149–156.

Hanley-Bowdoin, L., Bejarano, E.R., Robertson, D., and Mansoor, S. (2013). Geminiviruses: masters at redirecting and reprogramming plant processes. Nat. Rev. Microbiol. *11*, 777–788.

Hanssen, I.M., Lapidot, M., and Thomma, B.P.H.J. (2010). Emerging Viral Diseases of Tomato Crops. Am. Phytopathol. Soc. *23*, 539–548.

Hartley, J.L., Temple, G.F., and Brasch, M.A. (2000). DNA Cloning Using In Vitro Site-Specific Recombination. Genome Res. *10*, 1788–1795.

Hawe, A., Sutter, M., and Jiskoot, W. (2008). Extrinsic fluorescent dyes as tools for protein characterization. Pharm. Res. *25*, 1487–1499.

Herrmann, C., Horn, G., Spaargaren, M., and Wittinghofer, A. (1996). Differential Interaction of the Ras Family GTP-binding Proteins H-Ras , Rap1A , and R-Ras with the Putative Effector Molecules Raf Kinase and Ral-Guanine Nucleotide Exchange Factor *. J. Biol. Chem. *271*, 6794–6800.

Hillenkamp, F., Karas, M., Beavis, R.C., and Chait, B.T. (1991). Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers. Anal. Chem. *63*, 1193–1201.

Hipp, K., Rau, P., Schäfer, B., Gronenborn, B., and Jeske, H. (2014). The RXL motif of the African cassava mosaic virus Rep protein is necessary for rereplication of yeast DNA and viral infection in plants. Virology *462–463*, 189–198.

Hipp, K., Schäfer, B., Kepp, G., and Jeske, H. (2016). Properties of African Cassava Mosaic Virus Capsid Protein Expressed in Fission Yeast. Viruses *8*, 1–14.

Holtorf, H., Guitton, M.-C., and Reski, R. (2002). Plant functional genomics. Naturwissenschaften *89*, 235–249.

Honoré, B., Østergaard, M., and Vorum, H. (2004). Functional genomics studied by proteomics. BioEssays *26*, 901–915.

Ikai, A. (1980). Thermostability and aliphatic index of globular proteins. J. Biochem. *1898*, 1895–1898.

Iyer, L.M., Leipe, D.D., Koonin, E. V, and Aravind, L. (2004). Evolutionary history and higher order classification of AAA + ATPases. J. Struct. Biol. *146*, 11–31.

James, J.A., Escalante, C.R., Yoon-robarts, M., Edwards, T.A., Linden, R.M., and Aggarwal, A.K. (2003). Crystal Structure of the SF3 Helicase from Adeno-Associated Virus Type 2. Structure *11*, 1025–1035.

Jeske, H. (2009). Geminiviruses. In TT Viruses: The Still Elusive Human Pathogens, E.-M. de Villiers, and H. zur Hausen, eds. (Heidelberg: Springer Verlag), pp. 185–226.

Jones, D.T. (1999). Protein Secondary Structure Prediction Based on Position-specific Scoring Matrices. J. Mol. Biol. *292*, 195–202.

Kaboord, B., and Perr, M. (2008). Isolation of Proteins and Protein Complexes by Immunoprecipitation. Methods Mol. Biol. *424*, 349–364.

Kazlauskas, D., Varsani, A., and Krupovic, M. (2018). Pervasive Chimerism in the Replication-Associated Proteins of Uncultured Single-Stranded DNA Viruses. Viruses *10*, 1–11.

Kelley, L.A., and Sternberg, M.J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. *4*, 363–371.

Kelly, S.M., Jess, T.J., and Price, N.C. (2005). How to study proteins by circular dichroism. Biochim. Biophys. Acta *1751*, 119–139.

Kerppola, T.K. (2008). Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells. Annu. Rev. Biophys. *37*, 465–487.

Kittelmann, K., Rau, P., Gronenborn, B., and Jeske, H. (2009). Plant geminivirus rep protein induces rereplication in fission yeast. J. Virol. *83*, 6769–6778.

Knight, G.L., Pugh, A.G., Yates, E., Bell, I., Wilson, R., Moody, C.A., Laimins, L.A., and Roberts, S. (2011). A cyclin-binding motif in human papillomavirus type 18 (HPV18) E1 ^ E4 is necessary for association with CDK – cyclin complexes and G2 / M cell cycle arrest of keratinocytes , but is not required for differentiation-dependent viral genome ampli fi catio. Virology *412*, 196–210.

Kong, L., and Hanley-Bowdoin, L. (2002). A Geminivirus Replication Protein Interacts with a Protein Kinase and a Motor Protein That Display Different Expression Patterns during Plant Development and Infection. Plant Cell *14*, 1817–1832.

Kong, L., Orozco, B.M., Roe, J.L., Nagar, S., Ou, S., Feiler, H.S., Durfee, T., Miller, A.B., Gruissem, W., Robertson, D., et al. (2000). A geminivirus replication protein interacts with the retinoblastoma protein through a novel domain to determine symptoms and tissue specificity of infection in plants. EMBO J. *19*, 3485–3495.

Kono, A., Umeda-Hara, C., Adachi, S., Nagata, N., Konomi, M., Nakagawa, T., Uchimiya, H., and Umeda, M. (2007). The Arabidopsis D-Type Cyclin CYCD4 Controls Cell Division

in the Stomatal Lineage of the Hypocotyl Epidermis. Plant Cell Online *19*, 1265–1277.

Koonin, E. V, and Ilyina, T. V (1992). Geminivirus replication proteins are related to prokaryotic plasmid rolling circle DNA replication initiator proteins. J. Gen. Virol. *73*, 2763–2766.

Kostanski, L.K., Keller, D.M., and Hamielec, A.E. (2004). Size-exclusion chromatography-a review of calibration methodologies. J. Biochem. Biophys. Methods *58*, 159–186.

Kushwaha, N.K., Bhardwaj, M., and Chakraborty, S. (2017). The replication initiator protein of a geminivirus interacts with host monoubiquitination machinery and stimulates transcription of the viral genome. PLOS Pathogens, *13*, 1-41.

Kyratsous, C. a., Silverstein, S.J., DeLong, C.R., and Panagiotidis, C. a. (2009). Chaperone-fusion expression plasmid vectors for improved solubility of recombinant proteins in Escherichia coli. Gene *440*, 9–15.

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. *157*, 105–132.

Laconte, L.E.W., Srivastava, S., and Mukherjee, K. (2017). Probing Protein Kinase-ATP Interactions Using a Fluorescent ATP Analog. Methods Mol. Biol. *1647*, 171–183.

Lakowicz, J.R. (2006). Protein Fluorescence. In Principles of Fluorescence Spectroscopy, (Baltimore, USA: Springer), pp. 529–569.

Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S.G., Schell, J., and Gronenborn, B. (1995a). In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. Proc. Natl. Acad. Sci. USA *92*, 3879–3883.

Laufs, J., Schumacher, S., Geisler, N., Jupin, I., and Gronenborn, B. (1995b). Identification of the nicking tyrosine of geminivirus Rep protein. FEBS Lett. *377*, 258–262.

Laufs, J., Jupin, I., David, C., Schumacher, S., Heyraud-Nitschke, F., and Gronenborn, B. (1995c). Geminivirus replication: genetic and biochemical characterization of Rep protein function, a review. Biochimie *77*, 765–773.

Lazarowitz, S.G. (1987). The Molecular Characterization of Geminiviruses. Plant Mol. Biol. Report. *4*, 177–192.

Lazarowitz, G., Wu, L.C., Rogerqb, S.G., and Elmerb, J.S. (1992). Sequence-Specific lnteraction with the Viral ALI Protein ldentifies a Geminivirus DNA Replication Origin. Plant Cell *4*, 799–809.

Lees, J.G., Miles, A.J., Wien, F., and Wallace, B.A. (2006). A reference database for circular dichroism spectroscopy covering fold and secondary structure space. Bioinformatics *22*, 1955–1962.

Legg, J.P., and Fauquet, C.M. (2004). Cassava mosaic geminiviruses in Africa. Plant Mol. Biol. *56*, 585–599.

Legg, J., Somado, E.A., Barker, I., Beach, L., Ceballos, H., Cuellar, W., Elkhoury, W., Gerling, D., Helsen, J., Hershey, C., et al. (2014). A global alliance declaring war on cassava viruses in Africa. Food Secur. *6*, 231–248.

Legg, J.P., Lava Kumar, P., Makeshkumar, T., Tripathi, L., Ferguson, M., Kanju, E., Ntawuruhunga, P., and Cuellar, W. (2015). Cassava Virus Diseases: Biology, Epidemiology, and Management (Elsevier Inc.). Advance in Virus Research. *91*, 85-141.

Li, D., Zhao, R., Lilyestrom, W., Gai, D., Zhang, R., DeCaprio, J.A., Fanning, E., Jochimiak, A., Szakonyi, G., and Chen, X.. (2003). Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. Nature *423*, 512–518.

Liang, X., Peng, L., Baek, C.-H., and Katzen, F. (2013). Single step BP/LR combined Gatewat reactions. Biotechniques *55*, 265–267.

Lichtman, J.W., and Conchello, J.A. (2005). Fluorescence microscopy. Nat. Methods *2*, 910–919.

Liu, H., Boulton, M.I., and Davies, J.W. (1997). Maize streak virus coat protein binds single- and double- stranded DNA in vitro. J. Gen. Virol. *78*, 1265–1270.

Liu, L., Saunders, K., Thomas, C.L., Davies, J.W., and Stanley, J. (1999). Bean Yellow Dwarf Virus RepA, but Not Rep, Binds to Maize Retinoblastoma Protein , and the Virus Tolerates Mutations in the Consensus Binding Motif. Virology *279*, 270–279.

Luque, A., Sanz-Burgos, A.P., Ramirez-Parra, E., Castellano, M.M., and Gutierrez, C. (2002). Interaction of Geminivirus Rep Protein with Replication Factor C and Its Potential Role during Geminivirus DNA Replication. Virology *302*, 83–94.

Ma, T., Zou, N., Yuan Lin, B., Chow, L.T., and Wade Harper, J. (1999). Interaction between cyclin-dependent kinases and human papillomavirus replication-initiation protein E1 is required for efficient viral replication. Proc. Natl. Acad. Sci. USA *96*, 382–387.

Malik, P.S., Kumar, V., Bagewadi, B., and Mukherjee, S.K. (2005). Interaction between coat protein and replication initiation protein of Mung bean yellow mosaic India virus might lead to control of viral DNA replication. Virology *337*, 273–283.

Mariño-Ramírez, L., Kann, M.G., Shoemaker, B.A., and Landsman, D. (2005). Histone structure and nucleosome stability. Expert Rev. Proteomics *2*, 719–729.

Matic, S., Pegoraro, M., and Noris, E. (2016). The C2 protein of tomato yellow leaf curl Sardinia virus acts as a pathogenicity determinant and a 16-amino acid domain is responsible for inducing a hypersensitive response in plants. Virus Res. *215*, 12–19.

Matulis, D., and Lovrien, R. (1998). 1-Anilino-8-naphthalene sulfonate anion-protein binding depends primarily on ion pair formation. Biophys. J. *74*, 422–429.

Menges, M. (2006). The D-Type Cyclin CYCD3;1 Is Limiting for the G1-to-S-Phase Transition in Arabidopsis. Plant Cell Online *18*, 893–906.

Menges, M., de Jager, S.M., Gruissem, W., and Murray, J. a H. (2005). Global analysis of the core cell cycle regulators of Arabidopsis identifies novel genes, reveals multiple and highly specific profiles of expression and provides a coherent model for plant cell cycle control. Plant J. *41*, 546–566.

Moore, K.J., and Lohman, T.M. (1994). Kinetic mechanism of adenine nucleotide binding to and hydrolysis by the Escherichia coli Rep monomer. 2. Application of a kinetic competition approach. Biochemistry *33*, 14565–14578.

Mwaba, I.I. (2011). Development of a VIGS vector based on SACMV for genomic studies in tobacco and cassava.

Nagar, S., Pedersen, T.J., Carrick, K.M., Hanley-Bowdoin, L., and Robertson, D. (1995). A Geminivirus Iduces Expression of a Host DNA Synthesis Protein in Terminally Differentiated Plant Cells. Plant Cell *7*, 705–719.

Nagar, S., Hanley-bowdoin, L., and Robertson, D. (2002). Host DNA Replication Is Induced by Geminivirus Infection of Differentiated Plant Cells. Plant Cell *14*, 2995–3007.

Nagy, P.D. (2008). Yeast as a model host to explore plant virus-host interactions. Annu. Rev. Phytopathol. *46*, 217–242.

Naicker, P., Achilonu, I., Fanucchi, S., Fernandes, M., Ibrahim, M. a a, Dirr, H.W., Soliman, M.E.S., and Sayed, Y. (2013). Structural insights into the South African HIV-1 subtype C protease: impact of hinge region dynamics and flap flexibility in drug resistance. J. Biomol. Struct. Dyn. *31*, 1370–1380.

Nash, T.E., Dallas, M.B., Reyes, M.I., Buhrman, G.K., Ascencio-Ibañez, J.T., and Hanley-Bowdoin, L. (2011). Functional analysis of a novel motif conserved across geminivirus Rep proteins. J. Virol. *85*, 1182–1192.

Nawaz-ul-Rehman, M.S., and Fauquet, C.M. (2009). Evolution of geminiviruses and their satellites. FEBS Lett. *583*, 1825–1832.

Ni, D.Q., Shaffer, J., and Adams, J.A. (2000). Insights into nucleotide binding in protein kinase A using fluorescent adenosine derivatives. Protein Sci. *9*, 1818–1827.

Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H., and Yura, T. (1998). Chaperone coexpression plasmids: Differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in Escherichia coli. Appl. Environ. Microbiol. *64*, 1694–1699.

Oakenfull, E.A., Riou-Khamlichi, C., and Murray, J.A.H. (2002). Plant D-type cyclins and the control of G1 progression. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *357*, 749–760.

Orozco, B.M., and Hanley-Bowdoin, L. (1996). A DNA Structure Is Required for Geminivirus Replication Origin Function. J. Virol. *70*, 148–158.

Orozco, B.M., and Hanley-Bowdoin, L. (1998). Conserved sequence and structural motifs contribute to the DNA binding and cleavage activities of a geminivirus replication protein. J. Biol. Chem. *273*, 24448–24456.

Orozco, B.M., Miller, A.B., Settlage, S.B., and Hanley-Bowdoin, L. (1997). Functional domains of a geminivirus replication protein. J. Biol. Chem. *272*, 9840–9846.

Orozco, B.M., Kong, L., Ann, L., Elledge, S., Orozco, B.M., Kong, L., Batts, L.A., Elledge, S., and Hanley-bowdoin, L. (2000). The Multifunctional Character of a Geminivirus Replication Protein Is Reflected by Its Complex Oligomerization Properties. J. Biol. Chem.

275, 6114–6122.

Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995). How To Measure and Predict the Molar Absorption-Coefficient of a Protein. Protein Sci. *4*, 2411–2423.

Palomares, L. a, Estrada-Mondaca, S., and Ramírez, O.T. (2004). Production of recombinant proteins: challenges and solutions. Methods Mol. Biol. *267*, 15–52.

Pant, V., Gupta, D., Choudhury, N.R., Malathi, V.G., Varma, a., and Mukherjee, S.K. (2001). Molecular characterization of the Rep protein of the blackgram isolate of Indian mungbean yellow mosaic virus. J. Gen. Virol. *82*, 2559–2567.

Parrish, J.R., Gulyas, K.D., and Finley, R.L. (2006). Yeast two-hybrid contributions to interactome mapping. Curr. Opin. Biotechnol. *14*, 387–393.

Perczel, A., and Fasman, G.D. (1992). Quantitative analysis of cyclic beta-turn models. Protein Sci. *1*, 378–395.

Perczel, A., and Hollosi, M. (1996). Turns. In Circular Dichroism and the Conformational Analysis of Biomolecules, G.D. Fasman, ed. (New York: Plenum Press), p. 307.

Perczel, A., Hollósi, M., Foxman, B.M., Fasman, G.D., and Fasman, G.D. (1991). Conformational Analysis of Pseudocyclic Hexapeptides Based on Quantitative Circular Dichroism (CD), NOE, and X-ray Data. The Pure CD Spectra of Type I and Type II β-Turns. J. Am. Chem. Soc. *113*, 9772–9784.

Pierce, E.J., and Rey, M.E.C. (2013). Assessing Global Transcriptome Changes in Response to South African Cassava Mosaic Virus [ZA-99] Infection in Susceptible Arabidopsis thaliana. PLoS One *8*, 1–21.

Pilartz, M., and Jeske, H. (1992). Abutilon mosaic geminivirus double-stranded DNA is packed into minichromosomes. Virology *189*, 800–802.

Pilartz, M., and Jeske, H. (2003). Mapping of Abutilon Mosaic Geminivirus Minichromosomes. J. Virol. *77*, 10808–10818.

Pitt-Rivers, R., and Impiombato, F.S.A. (1968). The Binding of Sodium Dodecyl Sulphate to Various Proteins. Biochem. J. *109*, 825–830.

Pooggin, M.M. (2013). How Can Plant DNA Viruses Evade siRNA-Directed DNA

Methylation and Silencing ? Int. J. Mol. Sci *14*, 15233–15259.

Provencher, S.W., and Gloeckner, J. (1981). Estimation of globular protein secondary structure from circular dichroism. Biochemistry *20*, 33–37.

Raghavan, V., Malik, P.S., Roy, N., Mukherjee, S.K., and Choudhury, N.R. (2004). The DNA-A Component of a Plant Geminivirus (Indian Mung Bean Yellow Mosaic Virus) Replicates in Budding Yeast Cells. J. Virol. *78*, 2405–2413.

Raja, P., Sanville, B.C., Buchmann, R.C., and Bisaro, D.M. (2008). Viral genome methylation as an epigenetic defense against geminiviruses. J. Virol. *82*, 8997–9007.

Ramsperger, U., and Stahl, H. (1995). Unwinding of chromatin by the SV40 large T antigen. EMBO J. *14*, 3215–3225.

Rao, V.S., Srinivas, K., Sujini, G.N., and Kumar, G.N.S. (2014). Protein-Protein Interaction Detection: Methods and Analysis. Int. J. Proteomics *2014*, 1–12.

Rey, C., and Vanderschuren, H.V. (2017). Cassava Mosaic and Brown Streak Diseases: Current Perspectives and Beyond. Annu. Rev. Virol. *4*, 429-452.

Reyes, M.I., Nash, T.E., Dallas, M.M., Ascencio-Ibáñez, J.T., and Hanley-Bowdoin, L. (2013). Peptide aptamers that bind to geminivirus replication proteins confer a resistance phenotype to tomato yellow leaf curl virus and tomato mottle virus infection in tomato. J. Virol. *87*, 9691–9706.

Rizvi, I., Choudhury, N.R., and Tuteja, N. (2015). Insights into the functional characteristics of geminivirus rolling-circle replication initiator protein and its interaction with host factors affecting viral DNA replication. Arch. Virol. *160*, 375–387.

Rodríguez-Negrete, E., Lozano-Durán, R., Piedra-Aguilera, A., Cruzado, L., Bejarano, E.R., and Castillo, A.G. (2013). Geminivirus Rep protein interferes with the plant DNA methylation machinery and suppresses transcriptional gene silencing. New Phytol. *199*, 464–475.

Rosano, G.L., and Ceccarelli, E.A. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges. Front. Microbiol. *5*, 1–17.

Rosario, K., Duffy, S., and Breitbart, M. (2012). A field guide to eukaryotic circular single-

stranded DNA viruses : insights gained from metagenomics. Arch. Virol. *157*, 1851–1871.

Rose, A.S., and Hildebrand, P.W. (2015). NGL Viewer : a web application for molecular visualization. Nucleic Acids Res. *43*, 576–579.

Rose, A.S., Bradley, A.R., Valasatava, Y., Duarte, J.M., Prlić, A., and Rose, P.W. (2016). Web-based molecular graphics for large complexes. In Proceedings of the 21st International Conference on Web3D Technology, pp. 185–186.

Rudolph, R., and Lilie, H. (1996). In vitro folding of inclusion body proteins. FASEB J. *10*, 49–56.

Ruhel, R., and Chakraborty, S. (2018). Multifunctional roles of geminivirus encoded replication initiator protein. VirusDisease 1–8.

Sánchez-Durán, M. a, Dallas, M.B., Ascencio-Ibañez, J.T., Reyes, M.I., Arroyo-Mateos, M., Ruiz-Albert, J., Hanley-Bowdoin, L., and Bejarano, E.R. (2011). Interaction between Geminivirus Replication Protein and the SUMO-Conjugating Enzyme Is Required for Viral Infection. J. Virol. *85*, 9789–9800.

Saunders, K., Lucy, a, and Stanley, J. (1991). DNA forms of the geminivirus African cassava mosaic virus consistent with a rolling circle mechanism of replication. Nucleic Acids Res. *19*, 2325–2330.

Saunders, K., Lucy, a, and Stanley, J. (1992). RNA-primed complementary-sense DNA synthesis of the geminivirus African cassava mosaic virus. Nucleic Acids Res. *20*, 6311–6315.

Serebriiskii, I., Khazak, V., and Golemis, E. a (1999). A Two-hybrid Dual Bait System to Deiscriminate Speificity of Protein Interactions. J. Biol. Chem. *274*, 17080–17087.

Settlage, S.B., Miller, A.B., Gruissem, W., and Hanley-Bowdoin, L. (2001). Dual Interaction of a Geminivirus Replication Accessory Factor with a Viral Replication Protein and a Plant Cell Cycle Regulator. Virology *279*, 570–576.

Shen, J., Gai, D., Patrick, A., Greenleaf, W.B., and Chen, X.S. (2005). The roles of the residues on the channel beta-hairpin and loop structures of simian virus 40 hexameric helicase. Proc. Natl. Acad. Sci. USA *102*, 11248–11253.

Shen, W., Reyes, M.I., and Hanley-Bowdoin, L. (2009). Arabidopsis Protein Kinases GRIK1 and GRIK2 Specifically Activate SnRK1 by Phosphorylating Its Activation Loop. Plant Physiol. *150*, 996–1005.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 1–6.

Singh, D.K., Islam, M.N., and Choudhury, N.R. (2007). The 32 kDa subunit of replication protein A (RPA) participates in the DNA replication of Mung bean yellow mosaic India virus (MYMIV) by interacting with the viral Rep protein. Nucleic Acids Res. *35*, 755–770.

Singh, D.K., Malik, P.S., Choudhury, N.R., and Mukherjee, S.K. (2008). MYMIV replication initiator protein (Rep): Roles at the initiation and elongation steps of MYMIV DNA replication. Virology *380*, 75–83.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98*, 503–517.

Stanley, J. (1995). Analysis of African Cassava Mosaic Virus Recombinants Suggests Strand Nicking Occurs within the Conserved Nonanucleotide Motif during Initiation of Rolling Circle DNA Replication. Virology *206*, 707–712.

Stenger, D.C., Revington, G.N., Stevenson, M.C., and Bisaro, D.M. (1991). Replicational release of geminivirus genomes from tandemly repeated copies: evidence for rolling-circle replication of a plant viral DNA. Proc. Natl. Acad. Sci. U. S. A. *88*, 8029–8033.

van Stokkum, I.H., Spoelder, H.J., Bloemendal, M., van Grondelle, R., and Groen, F.C. (1990). Estimation of protein secondary structure and error analysis from circular dichroism spectra. Anal. Biochem. *191*, 110–118.

Van Stokkum, I.H.M., Spoelder, H.J.W., Bloemendal, M., Grondelle, R., and Groen, F.C.A. (1990). Estimation of protein secondary structure and error analysis from CD spectra. Anal. Biochem. *191*, 110–118.

Stothard, P. (2000). Internet On-Ramp Internet On-Ramp. Biotechniques *28*, 1102-1104.

Stryer, L. (1965). The interaction of a naphthalene dye with apomyoglobin and apohemoglobin. A fluorescent probe of non-polar binding sites. J. Mol. Biol. *13*, 482–495.

Stynen, B., Tournu, H., Tavernier, J., and Van Dijck, P. (2012). Diversity in genetic in vivo methods for protein-protein interaction studies: from the yeast two-hybrid system to the mammalian split-luciferase system. Microbiol. Mol. Biol. Rev. *76*, 331–382.

Sundström, J.F., Albihn, A., Boqvist, S., Ljungvall, K., Marstorp, H., Martiin, C., Nyberg, K., Vagsholm, I., Yuen, J., and Magnusson, U. (2014). Future threats to agricultural food production posed by environmental degradation, climate change, and animal and plant diseases - a risk analysis in three economic and climate settings. Food Secur. *6*, 201–215.

Sunter, G., Hartitz, M.D., Hormuzdi, S.G., Brough, C., and Bisaro, D.M. (1990). Genetic analysis of tomato golden mosaic virus: ORF AL2 is required for coat protein accumulation while ORF AL3 is necessary for efficient DNA replication. Virology *179*, 69–77.

Suter, B., Kittanakom, S., and Stagljar, I. (2008). Two-hybrid technologies in proteomics research. Curr. Opin. Biotechnol. *19*, 316–323.

Suyal, G., Mukherjee, S.K., Srivastava, P.S., and Choudhury, N.R. (2013). Arabidopsis thaliana MCM2 plays role(s) in mungbean yellow mosaic India virus (MYMIV) DNA replication. Arch. Virol. *158*, 981–992.

Suyal, G., Rana, V.S., Mukherjee, S.K., Wajid, S., and Choudhury, N.R. (2014). Arabidopsis thaliana NAC083 protein interacts with Mungbean yellow mosaic India virus (MYMIV) Rep protein. Virus Genes *48*, 486–493.

Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T., and Matsuo, T. (1988). Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. Rapid Commun. Mass Spectrom. *2*, 151–153.

Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Appl. Microbiol. Biotechnol. *60*, 523–533.

Tyers, M., and Mann, M. (2003). From genomics to proteomics. Nature *422*, 193–197.

Uetz, P., Dong, Y., Zeretzke, C., Atzler, C., Baiker, A., Berger, B., Rajagopala, S. V, Roupelieva, M., Rose, D., Fossum, E., et al. (2006). Herpesviral Protein Networks and Their Interaction with the Human Proteome. Science. *311*, 239–243.

Vanitharani, R., Chellappan, P., Pita, J.S., and Fauquet, C.M. (2004). Differential roles of AC2 and AC4 of cassava geminiviruses in mediating synergism and suppression of

posttranscriptional gene silencing. J. Virol. *78*, 9487–9498.

Walsh, G. (2002). Proteins Biochemistry and Biotechnology (Chichester: John Wiley and Sons, LTD).

Wang, L., Ding, X., Xiao, J., Jiménez-Góngora, T., Liu, R., and Lozano-Durán, R. (2017). Inference of a Geminivirus−Host Protein−Protein Interaction Network through Affinity Purification and Mass Spectrometry Analysis. Viruses *9*, 1-15.

Waugh, D.S. (2005). Making the most of affinity tags. Trends Biotechnol. *23*, 316–320.

Whitmore, L., and Wallace, B.A. (2008). Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. Biopolymers *89*, 392–400.

Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996). Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. Nature *379*, 466–469.

Wysocka, J. (2006). Identifying novel proteins recognizing histone modifications using peptide pull-down assay. Methods *40*, 339–343.

Yamaguchi, H., and Miyazaki, M. (2014). Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. Biomolecules *4*, 235–251.

Zerbini, F.M., Briddon, R.W., Idris, A., Martin, D.P., Moriones, E., Navas-castillo, J., Rivera-bustamante, R., Roumagnac, P., Varsani, A., and Consortium, I.R. (2017). ICTV ICTV Virus Taxonomy Profile : Geminiviridae. Journal of General Virology *98*, 131–133.

Zininga, T., Achilonu, I., Hoppe, H., Prinsloo, E., and Dirr, H.W. (2015). Overexpression, Purification and Characterisation of the Plasmodium falciparum Hsp70-z (PfHsp70-z) Protein. PLoS One *10*, 1–13.

Conferences to date

Poster presentation:

South African Society for Microbiology - Umhlanga, South Africa, 2016

- The expression and characterisation of the replication-associated protein (Rep) and BC1 of South African cassava mosaic virus

South African Society for Plant Pathology - Champagne Castle, Drakensberg, South Africa, 2017

- The expression and characterisation of the replication-associated protein (Rep) of South African cassava mosaic virus

Oral presentation

Virology Symposium 2016- Sakata, Lanseria, South Africa

- The structural characterization of the replication-association protein from South Africa Cassava Mosaic Virus and determining associated interactions with selected cassava host proteins

Appendix

Super optimum broth (SOB)

2% (w/v) tryptone

0.5% (w/v) yeast extract

10 mM NaCl

2.5 mM KCl

10 mM $MgSO_4$

10 mM MgCl

Super optimum broth with catabolite repression media (SOC)

2% (w/v) tryptone

0.5% (w/v) yeast extract

10 mM NaCl

2.5 mM KCl

10 mM $MgSO_4$

10 mM MgCl

20 mM glucose

Fairbanks staining method

Solution A: 25% Isopropanol, 10% Acetic Acid, 0.05% Coomassie R

Put gel in solution A, microwave until just before boiling, and shake for 5-10 minutes, decant.

Solution B: 10% Isopropanol, 10% Acetic Acid, 0.005% Coomassie R.

Add solution B, microwave and shake 10 min, decant.

Solution C: 10% Acetic Acid, 0.002% Coomassie R

Add solution C, microwave and shake 10 min, decant.

Solution D: 10% Acetic Acid

Add solution D, microwave and shake 10 min, decant.

YPAD plates

1% (w/v) yeast extract

2% (w/v) tryptone

2% (w/v) glucose

2% bacteriological agar

pET-28a/AC1 (Recombinant AC1 sequence is in bold)

TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGT
TACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCT
TTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCG
GGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAA
CTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTC
GCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGG
AACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGA
TTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTT
TAACAAAATATTAACGTTTACAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGC
GGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAA
TTAATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATC
AGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAAC
TCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGA
CTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATC
AAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTT
ATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAA
TCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAA
ATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGC
GCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTC
TAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCA
TCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGC
CAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATG
TTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCA
CCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCA
TGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCAT
AACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGACCAAA
ATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCA
AAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAA
AAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTT
TTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGT
GTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTC
GCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTA
CCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAA
CGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGA

GATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGG
CGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAG
CTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCT
GACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAA
ACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCAC
ATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGA
GTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAG
CGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGT
ATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATA
GTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCC
GACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATC
CGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCA
CCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCG
TGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTC
CAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTT
TCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAA
TGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAAC
ATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGC
GGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATG
TAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAA
TGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGA
AGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCA
CGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAG
CCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCC
ATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTG
ACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCG
ATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCT
GCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCG
ACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTC
AAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTG
CGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTA
ATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGG
TTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCC
CTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATC
CTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCG
TATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCG
CGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACG
ATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGT
CGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCA
GCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGC
GATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCT
TCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAAT
AACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCC
AGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACC

GCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGG
CACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGT
GCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCA
GTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCAC
TTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACG
GTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCA
CATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAA
GGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCC
TGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCA
AGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCT
GCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGA
TCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGC
CGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGC
GAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTA
GAAATAATTTTGTTTAACTTTAAGAAGGAGATATACC**ATGGGCAGCAGCCATCA**
TCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCTAG
CATGACTGGTGGACAGCAAATGGGTCGCGGATCCATGCCGAGGGCTGGTCG
TTTTAGCATAAAAGCCAAAAATTATTTCCTCACGTATCCGAAATGCACTCTC
TCGAAAGAAGCGGCATTAGATCAACTCCGACAACTCCAAACCCCAACAAATA
AATTGTTCATCAAGATCTGCAGAGAACTCCATGAAAATGGGGAACCTCATTT
GCATGCCCTCATTCAGTTCGAGGGCAAGTACAATTGTACCAACCAACGATTC
TTCGACCTCATATCCCCTTCCAGGTCAACACATTTCCATCCAAACATTCAGG
GAGCTAAATCCAGTTCTGACGTCAAGTCCTATTTGGACAAGGACGGAGACA
CCATCCAATGGGGCGAGTTTCAGATCGACGGACGATCTGCTCGCGGCGGAC
AACAATCCGCCAATGACGCTTACGCCAAGGCTCTTAACGCAGCAAGTAAAA
CAGAGGCTCTTAATGTAATCCGGGAACTAGCCCCAAAGGATTTTGTTTTACA
GTTTCATAATTTAAATAGCAATTTAGATAGGATTTTTCAGGAGCCTCCGATT
CCTTATATTTCTCCCTTTCTTTCTTCTTCTTTCACTCATGTTCCTGAGGAACT
TGAAGACTGGGTTTCCGAGAACGTGATGGGTTCCGCTGCGCGGCCATGGAG
ACCGAGTAGTATCGTCATCGAGGGCGATAGTAGGACAGGGAAGACGATGTG
GGCCCGATCTCTGGGACCACACAACTACTTATGTGGACATTTGGATCTCAGT
CCAAAGGTTTACAGCAACGACGCATGGTACAACGTCATTGATGACGTCGAC
CCCCATTACCTCAAGCACTTCAAAGAATTCATGGGGGCCCAAAGGGACTGG
CAAAGCAATACCAAGTACGGGAAGCCGATTCAAATTAAAGGCGGCATTCCC
ACTATCTTCCTATGCAATCCAGGACCGACATCATCATATAAAGAGTTTCTGG
ACGAGGAAAAGAACCAGTCCCTTAAAGCCTGGGCTTTAAAGAATGCAACCT
TCATCACCCTCCACGAGCCATTGTTCTCAAGTGCCCATCAAAGTCCAACACC
GCACCGCGAAGACTAGGCGGCCGCACTCGAGCACCACCACCACCACCACTG
AGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGC
TGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTT
TTGCTGAAAGGAGGAACTATATCCGGAT

pBlueScript nucleotide sequence

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCA
GCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGA
ATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTA
AAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGC
CCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAG
CACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGC
CGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAG
GGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCT
TAATGCGCCGCTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGG
GAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGA
TGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTT
GTAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGG
GTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCCT
GCAGCCCGGGGGATCCACTAGTTCTAGAGCGGCCGCCACCGCGGTGGAGCTCCA
GCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAG
CTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCG
GAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAA
TTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCA
TTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCC
GCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTAT
CAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAG
GAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCC
GCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATC
GACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGT
TTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGA
TACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTG
TAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAA
CCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCA
ACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTA
GCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACT
ACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTAC
CTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAG
CGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAA
GAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCAC
GTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTT
AAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCT
GACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTC
GTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGG
CTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCT
CCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGT
CCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAG
TAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCAT

CGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGAT
CAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGG
TCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATG
GCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGAC
TGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGC
TCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAA
GTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGC
TGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATC
TTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGC
AAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTT
CAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTG
AATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAG
TGCCAC

Histone H3 as cloned by GenScript into pDEST^{TM}22

ATGGCTCGCACAAAGCAAACTGCTAGGAAATCTACTGGAGGCAAGGCTCCAAGG
AAGCAGTTGGCCACCAAGGCAGCAAGGAAATCAGCTCCGGCGACCGGTGGTGTG
AAGAAACCTCATAGGTTCAGGCCTGGAACTGTTGCCCTGAGAGAAATCAGGAAG
TATCAAAAGAGCACAGAGCTCTTGATCAGAAAGCTTCCATTCCAGAGATTGGTG
AGGGAAATCGCTCAAGATTTCAAGACTGATTTGAGATTCCAGAGTAGTGCTGTAG
CAGCTCTACAGGAGGCAGCCGAAGCATATTTGGTTGGGTTGTTTGAGGATACAA
ATCTGTGTGCTATTCATGCTAAGAGGGTTACTATTATGCCCAAGGATATTCAGTT
GGCAAGAAGGATTAGGGGTGAGAGGGCTTAG

Cyclin D3;2 as cloned by GenScript into pDEST^{TM}22

ATGGCTTTGCAAGAAGAGGAGAATCGGCAGCTTCCAAGCTCAGCATTGGTTTTTG
ATGGATTATACTGTGAAGAAGAGGGGATTGAGGAAGATTATGCTTGCGTTGATG
ATCAGACTGTGAGAAAGGAGTCGCCTTTTTCTTGTGTTTTTCTCGAACAAGACCT
GTTTTGGGAAGATGATGAACTAGTTACTCTAATCTCTAAAGAAAAAGAGACCCG
CTTTTGTTTTGATAATGTAGTCTCTGATGGGTCTTTAATGATGGTTCGCAAAGAGG
CTATTGATTGGGTTTTGAGAGTAAAAGCTCACTATGGGTTCACTGCTTTGACTGG
TGTGCTTGCTGTGAACTACTTTGATAGGTTCATTTCTAGTTTAAATTTTCCGAGAG
ACAAGCCTTGGATGGGTCAACTTGCTGTTGTGGCTTGTTTGTCTCTGGCTGCCAA
AGTTGAGGAGACCCAAGTGCCCCTTCTTTTAGACTTGCAAGTGGAGGAATCAAA
GTATGTTTTCGAAGCCAAGACTATAAAGAGAATGGAGCTTCTGGTGCTCTCTACT
CTTCAGTGGAGGATGAATCCTGTGACGCCAATTTCCTTTTTTGATCATATTATAAG
GAGACTTGGACTAAAGAATCAATTGCATTGGGAATTTTTTAGGAGGTCTGAACGG
TTACTTCTCTCTGTCTTTGCTGATTCAAGGTTCATGAGTTATCTTCCTTCTACATTA
GCTACCTCAATTATGCTGCATGTCATTAAGGAGGTTGAGCCGTGTAATCAAGTGG
AACACCAGCATCAGCTTATGTCTGTGATCAAAATCAGTGAGGACAAAGTGAACG
AGTGTTACAAGCTCGTTTTGGAGCTATCAGGCAGTGAGAATAGAAGCTGCAAGA
GAAAGTGTCCCTTGACGCCCAGCAGTCCAAATGGAGTCATTGATGCATACTTCAG
CTGTGATAGCTCAAATGAGTCCTGGGCTGTGGCTTCAACGGTCTCATCACCACCA

GTGACTCGGTTCAAAAGGAATAGAACCCAGGATCAGCAGATGCGGGTGCCTTCA
CTAAATCTTCTTGAGGCGCAACATTTCTGTGTTTCACCGATTATCAATGCATTTCG
GGCGTATAATTTGAACCTTATTGGTACTCAGACAAAGGAAAAAGAAAGAAAGAT
GAGAAATTTTCTATGTGACCTCTTTATTTGA

Cyclin D4;2 as cloned by GenScript into pDEST^{TM}22

ATGGCACAGAATCCTGGCCGAGCAGTCTCGAGCCTTCTGTGTGCAGAGAACACA
ATTTTTGGTGATCTTGATTTCAATGCCTCAGATGAAATTGGGATTCCCCTTTCTTG
GCACCATCAAGATGATCAAAGTCGTAATCAAGACCCCTTTTCTTATAATGATAGC
TGCAAATCCTTGATGGGTTTTACAGTGCATAGTGAAGATAGAATCAAGGAAATG
GTTAAGGGGGAAAAGGAGCATTTGCCCAGAGATGACTATCTAAAGAGATTAAGG
AGTGGAGATTTGGACTTGAGTGTTAGGAGAGAGGCTCTTGATTGGATTTGGAAG
GCTCAAGCCCATTATAATTTCGGACCATTGAGTGTTTGTCTATCTATCAATTACTT
GGATCGATTCCTATCAGTGTATCAATTGCCTAAGGGTAAAGCTTGGGCTGTGCAA
TTGTTAGCTGTGGCTTGCTTATCACTAGCAGCCAAAATGGAAGAGGCTAATGTGC
CTCTCTCTGTAGATTTTCAGGTGGGAGAACCCAAGTTTGTGTTTGAAGCTAAAAC
TATACAAAGAATGGAGCTTTTGGTGTTGAGCACGTTGAAGTGGAGAATGCAAGC
TCTCACTCCCTGCTCCTTCATAGACTATTTCCTTAGCAAAATCAATGGAGATCAA
CATCTATCAACATCATCTATATTTAAATCATTGCAATTAATATTAAGCACACTTAA
AGGTATTGACTTCTTGGAATTCAGGCCTTCAGAGATTGCAGCAGCAGTGGCAATA
TCTGTTTCAGGAGAAGTCCAAGCAGTGGAAATTGATAAGGCAGTACCTTATTTCA
CCCAAGTAGAAAAGGTAAAAAGGGCCAAGAAAAAAAGCAGAGCTCTGTGGTGC
CCACAGAAGAAAGTCCAAAGATAG

Raw sequencing data:

Forward primer sequencing of pET-28a/AC1

TTCCCGTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAG
CAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATAT
GGCTAGCATGACTGGTGGACAGCAAATGGGTCGCGGATCCATGCCGAGGGCTGG
TCGTTTTAGCATAAAAGCCAAAAATTATTTCCTCACGTATCCGAAATGCACTCTC
TCGAAAGAAGCGGCATTAGATCAACTCCGACAACTCCAAACCCCAACAAATAAA
TTGTTCATCAAGATCTGCAGAGAACTCCATGAAAATGGGGAACCTCATTTGCATG
CCCTCATTCAGTTCGAGGGCAAGTACAATTGTACCAACCAACGATTCTTCGACCT
CATATCCCCTTCCAGGTCAACACATTTCCATCCAAACATTCAGGGAGCTAAATCC
AGTTCTGACGTCAAGTCCTATTTGGACAAGGACGGAGACACCATCCAATGGGGC
GAGTTTCAGATCGACGGACGATCTGCTCGCGGCGGACAACAATCCGCCAATGAC
GCTTACGCCAAGGCTCTTAACGCAGCAAGTAAAACAGAGGCTCTTAATGTAATC
CGGGAACTAGCCCCAAAGGATTTTGTTTTACAGTTTCATAATTTAAATAGCAATT
TAGATAGGATTTTTCAGGAGCCTCCGATTCCTTATATTTCTCCCTTTCTTTCTTCTT
CTTTCACTCATGTTCCTGAGGAACTTGAAGACTGGGTTTCCGAGAACGTGATGGG
TTCCGCTGCGCGGCCATGGAGACCGAGTAGTATCGTCATCGAGGGCGATAGTAG

GACAGGGAAGACGATGTGGGCCCGATCTCTGGGACCACACAACTACTTATGTGG
ACATTTGGATCTCAGTCCAAAGGTTTACAGCAACGACGCATGGTACAACGTCATT
GATGACGTCGACCCCCATTACCTCAAGCACTTCAAAGAATTCATGGGGGCCCAA
AGGGACTGGCAAGCAATACCAAGTACGGAGCGATTCAATAAAGCGCATTCCYAC
TATCTCTATGCATCAGACGACATCATCATATAAGAGTTCTGAACACAACGACCAG
TCATAAGCCTGGCTTAAGAAKGCAGCTTCATTA

Reverse primer sequencing of pET-28a/AC1 (Reverse complement)

CRCGACYTGGTCGGTTTAGCATAAAGCCAAATATTCTCACGTATCCGAAATGCAC
CTCTYCTCGAAAGAGCRGCATTAGATCACTCGACACTCCAACCCCACAAATAAAT
GTCATCAGATCTGCAGAGAACTCCATGAAAATGRGGAACTCATTTGCATGCCCTC
ATTCAGTTCGAGGGCAAGTACAATTGTACCAACCAACGATTCTTCGACCTCATAT
CCCTTCCAGGTCAACACATTTCCATCCAAACATTCAGGGAGCTAAATCCAGTTCT
GACGTCAAGTCCTATTTGGACAAGGACGGAGACACCATCCAATGGGGCGAGTTT
CAGATCGACGGACGATCTGCTCGCGGCGGACAACAATCCGCCAATGACGCTTAC
GCCAAGGCTCTTAACGCAGCAAGTAAAACAGAGGCTCTTAATGTAATCCGGGAA
CTAGCCCCAAAGGATTTTGTTTTACAGTTTCATAATTTAAATAGCAATTTAGATA
GGATTTTTCAGGAGCCTCCGATTCCTTATATTTCTCCCTTTCTTTCTTCTTCTTTCA
CTCATGTTCCTGAGGAACTTGAAGACTGGGTTTCCGAGAACGTGATGGGTTCCGC
TGCGCGGCCATGGAGACCGAGTAGTATCGTCATCGAGGGCGATAGTAGGACAGG
GAAGACGATGTGGGCCCGATCTCTGGGACCACACAACTACTTATGTGGACATTTG
GATCTCAGTCCAAAGGTTTACAGCAACGACGCATGGTACAACGTCATTGATGAC
GTCGACCCCCATTACCTCAAGCACTTCAAAGAATTCATGGGGGCCCAAAGGGAC
TGGCAAAGCAATACCAAGTACGGGAAGCCGATTCAAATTAAAGGCGGCATTCCC
ACTATCTTCCTATGCAATCCAGGACCGACATCATCATATAAAGAGTTTCTGGACG
AGGAAAAGAACCAGTCCCTTAAAGCCTGGGCTTTAAAGAATGCAACCTTCATCA
CCCTCCACGAGCCATTGTTCTCAAGTGCCCATCAAAGTCCAACACCGCACCGCGA
AGACGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCYGCTAM
CAAAGCCCGAAAGAAGYRGG

Forward primer sequencing of pEXP^{TM}32/H3

TTYCCCMAAACCMAAAAGGTCTCCGCTGACTAGGGCACATCTGACAGAAGTGGA
ATCAAGGCAARRRRGACTGGAACAGCTATTTCTACTGATTTTTCCTCSAGAAGAC
CTTGACATGATTTTGAAAATGGATTCTTTACAGGATATAAAAGCATTGTTAACAG
GATTATTTGTACAAGATAATGTGAATAAAGATGCCGTCACAGATAGATTGGCTTC
AGTGGAGACTGATATGCCTCTAACATTGAGACAGCATAGAATAAGTGCGACATC
ATCATCGGAAGAGAGTAGTAACAAAGGTCAAAGACAGTTGACTGTATCGTCGAG
GTCGAATCAAACAAGTTTGTACAAAAAAGCAGGCTTCATGGCTCGCACAAAGCA
AACTGCTAGGAAATCTACTGGAGGCAAGGCTCCAAGGAAGCAGTTGGCCACCAA
GGCAGCAAGGAAATCAGCTCCGGCGACCGGTGGTGTGAAGAAACCTCATAGGTT
CAGGCCTGGAACTGTTGCCCTGAGAGAAATCAGGAAGTATCAAAAGAGCACAGA
GCTCTTGATCAGAAAGCTTCCATTCCAGAGATTGGTGAGGGAAATCGCTCAAGAT
TTCAAGACTGATTTGAGATTCCAGAGTAGTGCTGTAGCAGCTCTACAGGAGGCA

GCCGAAGCATATTTGGTTGGGTTGTTTGAGGATACAAATCTGTGTGCTATTCATG
CTAAGAGGGTTACTATTATGCCCAAGGATATTCAGTTGGCAAGAAGGATTAGGG
GTGAGAGGGCTTAGGACCCAGCTTTCTTGTACAAAGTGGTTTGATGGCCGCTAAG
TAAGTAAGACGTCGAGCTCTAAGTAAGTAACGGCCGCCACCGCGGTGGAGCTTT
GGACTTCTTCGCCAGAGGTTTGGTCAAGTCTCCAATCAAGGTTGTCGGCTTGTCT
ACCTTGCCAGAAATTTACGAAAAGATGGAAAAGGTCAAATCGTTGGTAGATACG
TTGTTGACACTTCTAAATAAGCGAATTTCTTATGATTTATGATTTTTTATTATTAA
ATAAGTTATAAAAAAAATAGTGTATACAAATTTTAAAGTGACTCTTAGGTTTTAA
AACGAAATTCCTTATTCCTTGAGTAAACCTCTTTTTC

Reverse primer sequencing of pEXP^{TM}32/H3 (Reverse complement)

CTAGATTGAGATTAAGTTTGCCKTCTTGCATCAGTTAWTGACTGCATAATTATCT
TGTTTCTCGTCATTGTTCTCCGTTCCTTCTTCCTGTTCTTTTTCGCCAATATTTCAG
CTATACAAGCATACATTCACTCAGCTTGAAGCAAGCCTCCTGAAAGATGAAGCT
ACTGTCTTCTATCGAACAAGCATGCGATATTTGCCGACTTAAAAAGCTCAAGTGC
TCCAAAGAAAACCGAAGTGCGCCAAGTGTCTGAAGAACAACTGGGAGTGTCGCT
ACTCTCCCAAAACCAAAAGGTCTCCGCTGACTAGGGCACATCTGACAGAAGTGG
AATCAAGGCTAGAAAGACTGGAACAGCTATTTCTACTGATTTTTCCTCGAGAAGA
CCTTGACATGATTTTGAAAATGGATTCTTTACAGGATATAAAAGCATTGTTAACA
GGATTATTTGTACAAGATAATGTGAATAAAGATGCCGTCACAGATAGATTGGCTT
CAGTGGAGACTGATATGCCTCTAACATTGAGACAGCATAGAATAAGTGCGACAT
CATCATCGGAAGAGAGTAGTAACAAAGGTCAAAGACAGTTGACTGTATCGTCGA
GGTCGAATCAAACAAGTTTGTACAAAAAAGCAGGCTTCATGGCTCGCACAAAGC
AAACTGCTAGGAAATCTACTGGAGGCAAGGCTCCAAGGAAGCAGTTGGCCACCA
AGGCAGCAAGGAAATCAGCTCCGGCGACCGGTGGTGTGAAGAAACCTCATAGGT
TCAGGCCTGGAACTGTTGCCCTGAGAGAAATCAGGAAGTATCAAAAGAGCACAG
AGCTCTTGATCAGAAAGCTTCCATTCCAGAGATTGGTGAGGGAAATCGCTCAAG
ATTTCAAGACTGATTTGAGATTCCAGAGTAGTGCTGTAGCAGCTCTACAGGAGGC
AGCCGAAGCATATTTGGTTGGGTTGTTTGAGGATACAAATCTGTGTGCTATTCAT
GCTAAGAGGGTTACTATTATGCCCAAGGATATTCAGTTGGCAAGAAGGATTAGG
GGTGAGAGGGCTTAGGRCCCAGCTTTCTTGTACAAAGTGGTTTGATGSCCYKYYT
TKTGTAAGTTAAAGAMCCGTC

Forward primer sequencing of pEXP^{TM}32/CYCD3;2

GGCRAKTYCGCTACTCTCCAAAACCAAAAGGTCTCCGCTGACTAGGGCACATCT
GACAGAAGTGGAATCMMSGYTAGAAAGACTGGAACAGCTATTTCTACTGATTTT
TCCTCGAGAAGACCTTGACATGATTTTGAAAATGGATTCTTTACAGGATATAAAA
GCATTGTTAACAGGATTATTTGTACAAGATAATGTGAATAAAGATGCCGTCACAG
ATAGATTGGCTTCAGTGGAGACTGATATGCCTCTAACATTGAGACAGCATAGAAT
AAGTGCGACATCATCATCGGAAGAGAGTAGTAACAAAGGTCAAAGACAGTTGAC
TGTATCGTCGAGGTCGAATCAAACAAGTTTGTACAAAAAAGCAGGCTTCATGGC
TTTGCAAGAAGAGGAGAATCGGCAGCTTCCAAGCTCAGCATTGGTTTTTGATGGA
TTATACTGTGAAGAAGAGGGGATTGAGGAAGATTATGCTTGCGTTGATGATCAG

ACTGTGAGAAAGGAGTCGCCTTTTTCTTGTGTTTTTCTCGAACAAGACCTGTTTTG
GGAAGATGATGAACTAGTTACTCTAATCTCTAAAGAAAAAGAGACCCGCTTTTGT
TTTGATAATGTAGTCTCTGATGGGTCTTTAATGATGGTTCGCAAAGAGGCTATTG
ATTGGGTTTTGAGAGTAAAAGCTCACTATGGGTTCACTGCTTTGACTGGTGTGCT
TGCTGTGAACTACTTTGATAGGTTCATTTCTAGTTTAAATTTTCCGAGAGACAAGC
CTTGGATGGGTCAACTTGCTGTTGTGGCTTGTTTGTCTCTGGCTGCCAAAGTTGAG
GAGACCCAAGTGCCCCTTCTTTTAGACTTGCAAGTGGAGGAATCAAAGTATGTTT
TCGAAGCCAAGACTATAAAGAGAATGGAGCTTCTGGTGCTCTCTACTCTTCAGTG
GAGGATGAATCCTGTGACGCCAATTTCCTTTTTTGATCATATTATAAGGAGACTT
GGACTAAAGAATCAATTGCATTGGGGAATTTTTTTAGGAGGTCTGAACGGGTTAC
TTCTTCTCTGCTCTTTGCTGATTCAAGGGATCATGAGATATCATCTCCTACATTAG
CTACCCTCAATTATGCCKGCCATGG

Reverse primer sequencing of pEXP^{TM}32/CYCD3;2 (Reverse complement)

TGCGGATACAARGTCCTTGTTTTGGGAAGATGATGAGCTAGTACTCTAATCTCTA
AGAAAAAGAGATCGCTTTGTTTTGATAATGTAGTCTCTGATGGGTCTTTAATGAT
GGTTCGCAAAGAGGCTATTGATTGGGTTTTGAGAGTAAAAGCTCACTATGGGTTC
ACTGCTTTGACTGGTGTGCTTGCTGTGAACTACTTTGATAGGTTCATTTCTAGTTT
AAATTTTCCGAGAGACAAGCCTTGGATGGGTCAACTTGCTGTTGTGGCTTGTTTG
TCTCTGGCTGCCAAAGTTGAGGAGACCCAAGTGCCCCTTTCTTTTAGACTTGCAA
GTGGAGGAATCAAAGTATGTTTTCGAAGCCAAGACTATAAAGAGAATGGAGCTT
CTGGTGCTCTCTACTCTTCAGTGGAGGATGAATCCTGTGACGCCAATTTCCTTTTT
TGATCATATTATAAGGAGACTTGGACTAAAGAATCAATTGCATTGGGAATTTTTT
AGGAGGTCTGAACGGTTACTTCTCTCTGTCTTTGCTGATTCAAGGTTCATGAGTTA
TCTTCCTTCTACATTAGCTACCTCAATTATGCTGCATGTCATTAAGGAGGTTGAGC
CGTGTAATCAAGTGGAACACCAGCATCAGCTTATGTCTGTGATCAAAATCAGTGA
GGACAAAGTGAACGAGTGTTACAAGCTCGTTTTGGAGCTATCAGGCAGTGAGAA
TAGAAGCTGCAAGAGAAAGTGTCCCTTGACGCCCAGCAGTCCAAATGGAGTCAT
TGATGCATACTTCAGCTGTGATAGCTCAAATGAGTCCTGGGCTGTGGCTTCAACG
GTCTCATCACCACCAGTGACTCGGTTCAAAAGGAATAGAACCCAGGATCAGCAG
ATGCGGGTGCCTTCACTAAATCTTCTTGAGGCGCAACATTTCTGTGTTTCACCGAT
TATCAATGCATTTCGGGCGTATAATTTGAACCTTATTGGTACTCAGACAAAGGAA
AAAGAAAGAAAGATGAGAAATTTTCTATGTGACCTCTTTATTTGAGACCCAGCTT
TCTTGTACAAAGTGGTTTGATGGCCGCTAAGTAAGTAAGACGTCGAGCTCTAAGT
AAGTAACGGCCGCCACCGCGGTGGAGCTTTGGACTYTCSGCMMT

Forward primer sequencing of pEXP^{TM}32/CYCD4;2

TMCCTGGMCAAGAAGCGGAATCAAGGCAAAARRGGRCYGGAACAGCTATTTCT
ACTGMTTTTTCCTCSMKAAGACCTTGACATGATTTTGAAAATGGATTCTTTACAG
GATATAAAAGCATTGTTAACAGGATTATTTGTACAAGATAATGTGAATAAAGAT
GCCGTCACAGATAGATTGGCTTCAGTGGAGACTGATATGCCTCTAACATTGAGAC
AGCATAGAATAAGTGCGACATCATCATCGGAAGAGAGTAGTAACAAAGGTCAAA
GACAGTTGACTGTATCGTCGAGGTCGAATCAAACAAGTTTGTACAAAAAAGCAG

GCTTCATGGCACAGAATCCTGGCCGAGCAGTCTCGAGCCTTCTGTGTGCAGAGAA
CACAATTTTTGGTGATCTTGATTTCAATGCCTCAGATGAAATTGGGATTCCCCTTT
CTTGGCACCATCAAGATGATCAAAGTCGTAATCAAGACCCCTTTTCTTATAATGA
TAGCTGCAAATCCTTGATGGGTTTTACAGTGCATAGTGAAGATAGAATCAAGGA
AATGGTTAAGGGGGAAAAGGAGCATTTGCCCAGAGATGACTATCTAAAGAGATT
AAGGAGTGGAGATTTGGACTTGAGTGTTAGGAGAGAGGCTCTTGATTGGATTTG
GAAGGCTCAAGCCCATTATAATTTCGGACCATTGAGTGTTTGTCTATCTATCAATT
ACTTGGATCGATTCCTATCAGTGTATCAATTGCCTAAGGGTAAAGCTTGGGGCTG
TGCAATTGTTAGCTGTGGCTTGCTTATCACTAGCAGCCAAAATGGAAGAGGCTAA
TGTGCCTCTCTCTGTAGATTTTCAGGTGGGAGAACCAAGTTTGTGTTTGAAGCTA
AAACTATACAAAGAATGGAGCTTTTGGTGTTGAGCACGTTGAAGTGGAGAATGC
AAGCTCTCACTCCTGCTCCTTCATAGACTATTTCTTAGCAAATCAATGGAGATCA
ACATCTATCAACATCATCTATATTTAATCATTGCATTAATATTAGCACACTTTAAG
GCATTGACTTTCTGAATCAGCTTCAGAGATGCAGCAGCATGGCAATTTCTGTTCA
GGAAGTCCAAGCCAGTGGA

Reverse primer sequencing of pEXP^{TM}32/CYCD4;2 (Reverse complement)

TGAAGACYTGAAWTGCCTCTATACTGARGCAGCAWTGGATAGTGCGACATCATC
ATCGGAGAGAGGTAGTACAAGTTCAAGGCAGTGACTGTATCGTCGAGGTCGATC
AACAGTTGTACAAAAAAGCAGCTCATGCACAGAATCCTGCCGAGCAGTCTCGAG
CCTTTCTGTGTGCAGAGACACAATTTTTGGKGATCTTGATTCAATGCCTCAGATG
AAATTGGATTCCCCTTTCTTGGCACCATCAAGATGATCAAAGTCGTAATCAAGAC
CCCTTTTCTTATAATGATAGCTGCAAATCCTTGATGGGTTTTACAGTGCATAGTGA
AGATAGAATCAAGGAAATGGTTAAGGGGGAAAAGGAGCATTTGCCCAGAGATG
ACTATCTAAAGAGATTAAGGAGTGGAGATTTGGACTTGAGTGTTAGGAGAGAGG
CTCTTGATTGGATTTGGAAGGCTCAAGCCCATTATAATTTCGGACCATTGAGTGT
TTGTCTATCTATCAATTACTTGGATCGATTCCTATCAGTGTATCAATTGCCTAAGG
GTAAAGCTTGGGCTGTGCAATTGTTAGCTGTGGCTTGCTTATCACTAGCAGCCAA
AATGGAAGAGGCTAATGTGCCTCTCTCTGTAGATTTTCAGGTGGGAGAACCCAA
GTTTGTGTTTGAAGCTAAAACTATACAAAGAATGGAGCTTTTGGTGTTGAGCACG
TTGAAGTGGAGAATGCAAGCTCTCACTCCCTGCTCCTTCATAGACTATTTCCTTA
GCAAAATCAATGGAGATCAACATCTATCAACATCATCTATATTTAAATCATTGCA
ATTAATATTAAGCACACTTAAAGGTATTGACTTCTTGGAATTCAGGCCTTCAGAG
ATTGCAGCAGCAGTGGCAATATCTGTTTCAGGAGAAGTCCAAGCAGTGGAAATT
GATAAGGCAGTACCTTATTTCACCCAAGTAGAAAAGGTAAAAAGGGCCAAGAAA
AAAAGCAGAGCTCTGTGGTGCCCACAGAAGAAAGTCCAAAGGATAGGGACCCA
GCTTTCTKGTACAAAGTGGTTTGGRTGSCKSYTTTGGTAAAGGTTAAGCASCCGG
KT

GenScript

Make Research Easy
CONFIDENTIAL

Certificate of Analysis

Project ID: U8141BJ070-6

Construct Information:

Gene Name: Histone H3_pDEST 22

Clone ID: M66819 **Gene Length:** 411 bp

Cloning Vector: pDEST 22 **Cloning Strategy:** Gateway

QC Items	Specifications	Results	
Sequencing Alignment	Sequencing results are consistent with the targeted insert sequence.	Pass	Consistent
Vector Sequence	The flanking sequences of the cloning site are correct.	Pass	Correct Shown in the SQD file
Restriction Digests	The size of inserted fragment is correct and free of unexpected bands suggesting contamination.	Pass	Correct Shown in attachment 1
DNA Quality	Miniprep: 4 μg OD260/280=1.8~2.0 Free of contamination	Pass	≥ 4 μg OD260/280=1.89 Pure
Quality grade	Research Grade	Pass	Research Grade
Appearance	Clear and free of foreign particles.	Pass	Clear Free of foreign particles
Additional Test		N/A	

NOTE

Shipping at	Plasmid Storing at	Bacstab Storing at	Glycerol Stock Storing at
Room Temperature	-20℃	4℃	-20℃/-80℃

Morgan

Certified by: **Date:** **11-12-2016**

Thank you for your patronage to our Gene services! To maintain this working relationship, we shall be grateful if you can add our webpage URL into your lab website. As a token of appreciation, you will be rewarded by 1,000 EZcoupon™ points. For more information, please contact us by e-mail at web@genscript.com.

For research use only

860 Centennial Ave., Piscataway, NJ 08854, USA

Toll-Free: 1-877-436-7274 Tel: 1-732-885-9188 Fax: 1-732-210-0262 Email: order@genscript.com Web: www.genscript.com

Page 1 of 4

GenScript

Make Research Easy
CONFIDENTIAL

Certificate of Analysis

Project ID: U8141BJ070-10

Construct Information:

Gene Name: Cyclin D3, 2 pDEST 22

Clone ID: M66564 **Gene Length:** 1185 bp

Cloning Vector: pDEST 22 **Cloning Strategy:** Gateway

QC Items	Specifications	Results	
Sequencing Alignment	Sequencing results are consistent with the targeted insert sequence.	Pass	Consistent
Vector Sequence	The flanking sequences of the cloning site are correct.	Pass	Correct Shown in the SQD file
Restriction Digests	The size of inserted fragment is correct and free of unexpected bands suggesting contamination.	Pass	Correct Shown in attachment 1
DNA Quality	Miniprep: 4 μg OD260/280=1.8~2.0 Free of contamination	Pass	≥ 4 μg OD260/280=1.82 Pure
Quality grade	Research Grade	Pass	Research Grade
Appearance	Clear and free of foreign particles.	Pass	Clear Free of foreign particles
Additional Test		N/A	

NOTE

Shipping at	Plasmid Storing at	Bacstab Storing at	Glycerol Stock Storing at
Room Temperature	-20℃	4℃	-20℃/-80℃

Morgan

Certified by: **Date: 11-11-2016**

Thank you for your patronage to our Gene services! To maintain this working relationship, we shall be grateful if you can add our webpage URL into your lab website. As a token of appreciation, you will be rewarded by 1,000 EZcoupon™ points. For more information, please contact us by e-mail at web@genscript.com.

For research use only

860 Centennial Ave., Piscataway, NJ 08854, USA

Toll-Free: 1-877-436-7274 Tel: 1-732-885-9188 Fax: 1-732-210-0262 Email: order@genscript.com Web: www.genscript.com

GenScript

Make Research Easy
CONFIDENTIAL

Certificate of Analysis

Project ID: U8141BJ070-8

Construct Information:

Gene Name: Cyclin D4, 2 pDEST 22

Clone ID: M66561 **Gene Length:** 900 bp

Cloning Vector: pDEST 22 **Cloning Strategy:** Gateway

QC Items	Specifications	Results	
Sequencing Alignment	Sequencing results are consistent with the targeted insert sequence.	Pass	Consistent
Vector Sequence	The flanking sequences of the cloning site are correct.	Pass	Correct Shown in the SQD file
Restriction Digests	The size of inserted fragment is correct and free of unexpected bands suggesting contamination.	Pass	Correct Shown in attachment 1
DNA Quality	Miniprep: 4 μg OD260/280=1.8~2.0 Free of contamination	Pass	≥ 4 μg OD260/280=1.87 Pure
Quality grade	Research Grade	Pass	Research Grade
Appearance	Clear and free of foreign particles.	Pass	Clear Free of foreign particles
Additional Test		N/A	

NOTE

Shipping at	Plasmid Storing at	Bacstab Storing at	Glycerol Stock Storing at
Room Temperature	-20℃	4℃	-20℃/-80℃

Morgan

Certified by: **Date: 11-11-2016**

Thank you for your patronage to our Gene services! To maintain this working relationship, we shall be grateful if you can add our webpage URL into your lab website. As a token of appreciation, you will be rewarded by 1,000 EZcoupon™ points. For more information, please contact us by e-mail at web@genscript.com.

For research use only

860 Centennial Ave., Piscataway, NJ 08854, USA
Toll-Free: 1-877-436-7274 Tel: 1-732-885-0188 Fax: 1-732-210-0262 Email: order@genscript.com Web: www.genscript.com