

https://dspace.org/

# Geminivirus replication-association protein (Rep) as a target for the development of small molecule inhibitors by

**Kgomokaboya, Maite Codlinne**

**2020**

https://hdl.handle.net/10539/30984

Kgomokaboya, Maite Codlinne (2020) Geminivirus replication-association protein (Rep) as a target for the development of small molecule inhibitors, University of the Witwatersrand, Johannesburg, <http://hdl.handle.net/10539/30984>

Downloaded from a DSpace repository. https://dspace.org/




# Geminivirus replication-association protein (Rep) as a target for the development of small molecule inhibitors

by

**Maite Codlinne Kgomokaboya**

**(1482339)**

**Dissertation**

Submitted in fulfilment of the requirements for the degree

**Master of Science**

in

**Molecular and Cell Biology**

in the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

Supervisor: Prof Yasien Sayed

Co-supervisor: Dr Salerwe Mosebi

August 2020

## Declaration

I, Maite Codlinne (Student number: 1482339), am a student registered for the degree of Master of Science in the academic year 2020.
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 all the work submitted for examination is my own unaided work except where I explicitly indicated otherwise.
• 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.



**……**
**24 Day of, August, 2020**

**Abstract**

Replication-association protein (Rep) is an indispensable protein for the replication of *Geminivirus* single-stranded DNA in its life cycle. *Geminivirus*, in particular *African Cassava Mosaic Virus* (ACMV), affects agricultural crops such as Cassava (*Manihot esculenta* Crantz) that have the potential to bridge food insecurity in sub-Saharan Africa where it is mostly produced. For this virus to replicate, it encode for a Rep protein to initiate the rolling cycle replication (RCR) therefore making the protein indispensable for viral life cycle. Understanding the mechanism of viral replication proteins is essential because of the involvement with replication and spread of the disease. There is an urgent need to develop novel strategies to control the virus. Therefore, the main aim of this study was to investigate ACMV Rep protein functional activity as a target for the development of small molecule inhibitors. To achieve this aim, overexpression of full length ACMV Rep recombinant protein in *Escherichia coli* (*E. coli*) BL21 (DE3) pLysS bacterial cells and purification on nickel affinity chromatography was successfully done. Biochemical and biophysical characterisation is a crucial step to follow in order to understand the role of proteins in the viral life cycle. The structural determination of ACMV Rep protein was first predicted using online bioinformatics tool ExPASy and further determined with Fourier-transform infrared (FTIR) and intrinsic fluorescence spectroscopy. The secondary structure prediction from ExPASy resulted in the prediction of helices, sheets and coils. From FTIR spectroscopy analyses, the Amide I region was detected which was reported to be common in proteins representing the secondary structure. The overlapping peaks of the Amide I region were resolved through deconvolution on OriginPro 8 and this resulted in α-helices, β-sheets, β-turns and coils which correlates with the ExPASy results. Intrinsic fluorescence spectra showed an emission maximum wavelength of 355 nm at both 280 nm and 290 nm excitation wavelengths. The emission maxima shows that the tryptophan is exposed to the solvent. Assessing the activity of proteins is another step to determine proper folding of the protein. This was determined with binding and cleavage assays using Electromobility Shift Assay (EMSA). Results showed that purified ACMV Rep was functional in both the binding and cleavage activity at increasing concentrations (≥ 0.4 μM). Enzyme-linked Immunosorbent Assay (ELISA), a microtiter plate assay for protein functionality, was also developed for the interaction of ACMV Rep and ACMV DNA. This assay is amenable to high throughput screening of potential inhibitors. ELISA results also confirmed the binding activity of the purified Rep protein to ACMV DNA, this assay can be adapted to screen potential small molecule inhibitors of ACMV Rep interaction with ACMV

DNA. The identification of small molecules that inhibit the activity of ACMV Rep protein to viral DNA on EMSA was done. Natural phenolic compounds such as Epigallocatechin gallate (EGCG) and Chicoric acid (CA) have inhibition activity on the binding of ACMV Rep to DNA at a concentration of ≥ 100 µM. In addition, EGCG inhibits the cleavage activity of DNA by ACMV Rep protein with a concentration of ≥ 100 µM while CA showed no inhibition towards cleavage activity of the ACMV Rep on the DNA. From this study, more can be done to understand the crystal structure of the protein and its mechanism of interaction.

## Dedication

To my son, **Masilo Tumiso Kgomokaboya**, this one is for you Bozza!!!

***In loving memory of my Grandmother***
***Kgomokaboya Maropene Florah, Koko,***
***I am almost there!!!***

**Acknowledgment**

First and foremost, I would like to thank the **Almighty God** for giving me the strength and ability to undertake this project. All glory and honour belong to Him!

- To my supervisors, **Dr. Salerwe Mosebi and Prof. Yasien Sayed**, thank you for the unwavering support, guidance, and motivation you have shown me throughout the course of this study.
- To **Dr. Qasim Fish**, Q; I can never imagine anyone who has devoted much time to helping others succeed selflessly so than you did with me. Your patience, kindness and constant guidance have seen me through this project.
- To **Centre for Metal-based Drug Discovery group**. Dr. Mabel Coyanis, Dr. Thompho J. Rashamusi, Reagan M Mohlala, Dr. Zikhona Njengele-Tetyana, and Amukelani Marivate: You have constantly shown support and readily willing to assist when I needed help.
- To **Advanced Materials Division**, it was great meeting all of you and being part of the AMD community has made a difference in so many levels. Thank you for always showing a smile at the corridor, it made a difference where one seemed to lose hope.
- To **Dr. Abongile Tjijana and Raymond Makola** (soon to be Dr), thank you for always availing yourself to assist.
- I would like to thank the **University of the Witwatersrand** and the **South African National Research Foundation** for financial support.
- Thank you **MINTEK** for providing the resources to conduct this research.
- To **my family**, Mom, Stepdad, brother, sister, uncle, nieces and nephews. Thank you for always being my source of emotional well-being. Thank you for always believing in my dreams.
- To the **Sadiki brothers**: Vho Sadiki, you guys have been present since the start of this journey, you supported me at my lowest time in life. God gave me double the blessings.
- To **Dr. Thomas Chauke**, your music kept me going through my lab work, tough times and my writing process. **Hi khensile kokwani Chauke**.
- To **my Son**, all the tears you have cried whenever you see me leaving you behind, all the pain you felt when you had to go through each day without me. To all the questions you used to ask as to what am I doing. And where am I going? **I felt it too child**.

**Thank you!!!**

## Research Outputs

Conference Contributions

**Geminivirus replication-association protein (Rep) as a target for the development of small molecule inhibitors.** M. C. Kgomokaboya[1, 2], Q. Fish[1], S. Mosebi[3,] and Y. Sayed[2].

Oral presentation.

Molecular Biosciences Research Thrust Postgraduate Research Day on the 28 of November 2019 at WITS Adler Museum, Faculty of Health Sciences Building, University of the Witwatersrand. **Fourth prize**

## Table of Contents

Declaration ..... ii

Abstract ..... iii

Dedication ..... v

Acknowledgment ..... vi

Research Outputs ..... vii

Table of Contents ..... viii

List of Figures ..... xii

List of Tables ..... xv

List of Abbreviations ..... xvi

Chapter 1: Introduction ..... 1

1.1 Food security: A global food crisis ..... 1

1.2 Cassava (*Manihot esculenta* Crantz) ..... 1

1.3 Geminiviruses ..... 2

1.3.1 *Geminivirus*: genome organization ..... 2

1.3.2 *Geminivirus* proteins ..... 4

1.3.3 *Begomovirus* life cycle ..... 5

1.3.3.1 Replication initiator protein and its functions ..... 8

1.4 Strategies on controlling Geminiviruses ..... 10

1.4.1 Natural antiviral mechanism of plants ..... 10

1.4.1.1 RNA silencing ..... 10

1.4.1.2 Plant antioxidants ..... 11

1.4.2 Man-made antiviral mechanism in plants ..... 12

1.4.2.1 Pathogen-derived resistance (PDR) ..... 12

1.4.2.2 Gene editing: Non-transgenic Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) system ..... 13

1.4.2.3 Naturally-derived synthetic compounds ..... 14
1.5 Rationale of the study ..... 16
1.6 Hypothesis ..... 16
1.7 Aim ..... 16
1.8 Objectives ..... 16
Chapter 2: Methodology ..... 17
2.1 Overexpression of the recombinant ACMV Rep ..... 17
2.1.1 Nucleotide and amino acid sequences coding for ACMV Rep protein ..... 17
2.1.2 ACMV Rep molecular clone construction ..... 17
2.1.3 Bacterial transformations ..... 18
2.1.4 Glycerol bacterial stock production ..... 19
2.1.5 DNA plasmid purification ..... 19
2.1.6 DNA sequencing of ACMV gene encoding Rep protein ..... 20
2.1.7 Restriction enzyme double digestion ..... 20
2.1.8 Bacterial growth curve determination ..... 20
2.1.9 Gene induction for the over-expression of recombinant ACMV Rep protein ..... 21
2.2 Confirmation of protein induction by Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis (SDS-PAGE) ..... 21
2.3 Validation of ACMV Rep protein by Western blot ..... 21
2.4 Optimisation of Rep protein overexpression ..... 22
2.5 Purification of ACMV Rep protein from BL21 (DE3) pLyS bacterial cells ..... 22
2.5.1 Determination of Rep protein solubility ..... 22
2.5.2 Extraction of Rep protein from inclusion bodies ..... 23
2.5.2.1 Solubilisation using urea ..... 23
2.5.3 Purification and refolding of solubilised Rep protein ..... 23
2.5.4 ACMV Rep concentration determination ..... 24
2.5.5 Optimisation purification of ACMV Rep protein ..... 24

2.5.5.1 Extraction of Rep protein from inclusion bodies using anionic detergent Sodium Dodecyl Sulfate (SDS) ..... 24
2.5.5.2 Purification of Rep protein with N-laurylsarcosine (Sarkosyl) on immobilised affinity chromatography (IMAC) ..... 24
2.5.5.3 Removal of residual Sodium Dodecyl Sulfate (SDS) using Ion-exchange chromatography (IEC) ..... 25
2.6 Structural determination of recombinant Rep proteins ..... 25
2.6.1 Fourier-transform infrared spectroscopy (FTIR) ..... 25
2.6.2 Intrinsic and extrinsic fluorescence ..... 25
2.7 Functional analysis of the purified recombinant Rep proteins ..... 26
2.7.1 Electrophoretic Mobility Shift Assay (EMSA) ..... 26
2.7.1.1 ACMV Rep-DNA binding assay ..... 26
2.7.1.2 ACMV Rep-DNA cleavage assay ..... 26
2.7.2 Development of an Enzyme-Linked Immunosorbent Assay (ELISA) ..... 27
2.7.2.1 Nunc MaxiSorp ™ flat-bottom 96 well plate ..... 27
2.8 Inhibition studies ..... 28
2.8.1 ACMV Rep-DNA binding inhibition assay ..... 28
2.8.2 ACMV Rep-DNA cleavage assay ..... 28
Chapter 3: Results ..... 29
3.1 Nucleotide and amino acid sequences coding for Rep protein ..... 29
3.2 Bacterial transformation ..... 30
3.3 Verification of the ACMV Rep plasmid ..... 31
3.3.1 Enzyme restriction digestion ..... 31
3.3.2 DNA sequencing ..... 32
3.4 Overexpression of ACMV Rep protein ..... 33
3.4.1 Bacterial growth curve ..... 33
3.5 Confirmation of protein induction by Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis (SDS-PAGE) and Western blot ..... 34

3.5.1 Overexpression of target ACMV Rep protein and size determination standard curve ..... 34
3.5.2 Western Blot ..... 35
3.6 Optimisation of ACMV Rep protein Overexpression ..... 36
3.7 Purification, refolding and concentration determination of ACMV recombinant Rep protein ..... 38
3.7.1 Solubility studies of the Rep protein ..... 38
3.7.2 Purification of the ACMV recombinant Rep protein ..... 39
3.7.3 Purification of ACMV recombinant Rep protein solubilised with SDS ..... 41
3.8 Protein determination ..... 45
3.9 Structural determination of recombinant Rep proteins ..... 45
3.9.1 Secondary structure prediction ..... 45
3.9.2 Fourier-transform infrared (FTIR) ..... 47
3.9.3 Intrinsic and extrinsic fluorescence for tertiary structure ..... 51
3.10 Functional analysis of the purified recombinant Rep proteins ..... 52
3.10.1 Electromobility Shift Assay (EMSA) ..... 53
3.10.2 Enzyme-linked Immunosorbent Assay (ELISA) ..... 54
3.11 Inhibition studies ..... 55
3.11.1 Inhibition of protein-DNA binding ..... 56
Chapter 4: Discussion and Conclusion ..... 58
4.1 Verification of the insert cDNA in the pET-15b plasmid DNA ..... 58
4.2 Overexpression and purification of recombinant ACMV Rep protein ..... 59
4.3 Structural determination of recombinant Rep proteins ..... 62
4.3.1 Fourier-transform infrared (FTIR) ..... 62
4.3.2 Fluorescence spectroscopy ..... 63
4.4 Functional studies of the unfolded and refolded ACMV-Rep ..... 64
4.5 Inhibition of ACMV Rep protein with ACMV oligonucleotide ..... 67
Chapter 5: References ..... 70

Appendix.....86

## List of Figures

Figure 1.1: Genomic organisation of different types of Geminiviruses. .....4
Figure 1.2: Transmission of the virus onto cassava by the whitefly.....7
Figure 1.3: A typical demonstration of ACMV DNA origin of replication. .....9
Figure 1.4: RNA silencing mechanism of plants.....11
Figure 1.5: The structure of a phenol group. .....12
Figure 1.6: Chemical structure of the polyphenolic compound, Chicoric acid. .....14
Figure 1.7: Chemical structure of the polyphenolic compound (−) - epigallocatechin gallate (EGCG).....15
Figure 2.1: pET-32a expression vector map showing the multiple cloning sites. .....18
Figure 3.1: African cassava mosaic virus isolate West Kenyan 844 segment DNA1. .....29
Figure 3.2: ACMV 358 amino acid sequence.....30
Figure 3.3: Agar plate showing pET-15b-ACMV-Rep transformed colonies.....31
Figure 3.4: A 1% agarose gel showing the single and double enzyme digestion of plasmid from transformed BL21 (DE3) pLysS. .....32
Figure 3.5: DNA sequencing chromatogram results from Inqaba Biotec analysed using BioEdit tool.....33
Figure 3.6: *E. coli* BL21 (DE3) pLysS growth curve against time.....34
Figure 3.7: Represents a 12% SDS-PAGE analysis of Rep overexpression in BL21 (DE3) pLysS *E. coli* cells and molecular weight standard curve. .....35
Figure 3.8: A 12% SDS-PAGE showing Rep protein expression induced with 1 mM IPTG in BL21 (DE3) pLysS *E. coli* cells (A) and the corresponding western blot (B). .....36
Figure 3.9: A 12% SDS-PAGE showing overexpression of the recombinant ACMV Rep protein. .....37
Figure 3.10: A 12% SDS-PAGE showing Rep protein expression studies in BL21 (DE3) pLysS *E. coli* cells.....38
Figure 3.11: A 12 % SDS-PAGE analysis of fractions from Rep protein solubility studies...39
Figure 3.12: Nickel column elution profile of a sample containing recombinant Rep protein. .....40

Figure 3.13: A 12 % SDS-PAGE analysis of protein purification on a Ni2+ affinity column. .....41
Figure 3.14: Rep protein purification profile from 1 L bacterial culture on a His Trap HP Ni $^{2+}$ affinity column. .....42
Figure 3.15: A 12% SDS-PAGE analysis of fractions from Rep protein purification on the $Ni^{2+}$ affinity column. .....43
Figure 3.16: Western blot analysis of the Rep protein using 1:2000 anti-His Mouse monoclonal antibody and 1:10000 Goat Anti-Mouse HRP (IgG) antibody. .....44
Figure 3.17: A 12 % SDS-PAGE analysis of fractions from Rep protein purification on the Q Sepharose fast flow. .....44
Figure 3.18: Absorbance spectrum of Rep protein at 280 nm using the NanoDrop™ 2000/2000c Spectrophotometer. .....45
Figure 3.19: ACMV Rep protein secondary structure prediction analysis using the online prediction tool CFSSP.....46
Figure 3.20: An original FTIR spectra of both sodium phosphate buffer (Black) and Rep protein in the buffer (Red) in the 4000–1000 cm-1 regions. .....47
Figure 3.21: An original FTIR spectra of both sodium phosphate buffer and Rep protein in buffer in the 4000–1000 cm-1 region. .....48
Figure 3.22: FTIR spectra of the expanded view of the Amide I region (1600 cm -1-1700 cm 1), a broad peak due to the overlapping secondary components peaks. .....50
Figure 3.23: Second derivative FTIR spectra of the Amide I region (Red) resolved into its secondary structure components. .....50
Figure 3.24: Fluorescence emission spectra of 3 µM Rep protein (red line) and buffer (20 mM sodium phosphate, pH 8.00, 5 mM DTT, 10 % glycerol) when excited at 280 nm.....51
Figure 3.25: Fluorescence emission spectra of 3 µM Rep protein (Red line) and buffer (20 mM sodium phosphate, pH 8.00, 5 mM DTT, 10 % glycerol) when excited at 295 nm.....52
Figure 3.26: A 3% agarose gel in Borex buffer (100 mM borate, 150 mM NaCl, pH 7.5) electrophoresed at 150 V. .....53
Figure 3.27: A 3% agarose gel in Borex buffer (100 mM borate, 150 mM NaCl, and pH 7.5) electrophoresed at 150 V. .....53
Figure 3.28: A 3% agarose gel in Borex buffer (100 mM borate, 150 mM NaCl, pH 7.5) electrophoresed at 300 V. .....54
Figure 3.29: ELISA to confirm the binding of ACMV Rep protein with ACMV dsDNA. ....55

Figure 3.30: A 3% agarose gel in Borex buffer (100 mM borate, 150 mM NaCl, and pH 7.5) electrophoresed at 150 V. .....56
Figure 3.31: A 3% agarose gel in Borex buffer (100 mM borate, 150 mM NaCl, and pH 7.5) electrophoresed at 150 V. .....56
Figure 3.32: A 3% agarose gel in Borex buffer (100 mM borate, 150 mM NaCl, pH 7.5) electrophoresed at 150 V. .....57
Figure 3.33: A 3% agarose gel in Borex buffer (100 mM borate, 150 mM NaCl, pH 7.5) electrophoresed at 150 V. .....57

## List of Tables

Table 2.1**:** Primers used for DNA sequencing.....20
Table 2.2: Primers/Substrates used in this study .....27
Table 3.1: Theoretical characterisation of Rep protein using ExPASy bioinformatics tool....30
Table 3.2: FTIR analysis of ACMV recombinant Rep protein.....49

## List of Abbreviations

| Abbreviation | Full name |
|---|---|
| ACMV | African cassava mosaic virus |
| AEP | Associated effector protein |
| AGO | Argonaute family protein |
| ATP | Adenosine triphosphate |
| CA | Chicoric acid |
| CHAPS | (3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate) |
| CP | Coat protein |
| CR | Common region |
| CRISPR | Clustered Regularly Interspaced Short Palindromic Repeat |
| DCL | Dicer-like protein |
| DMSO | Dimethyl sulfoxide |
| DNA | Deoxy ribonucleic acid |
| dsDNA | Double-stranded DNA |
| DTT | Dithiothreitol |
| *E. coli* | Escherichia coli |
| ECL | Enhanced chemiluminescent |
| EDTA | Ethylenediaminetetraacetic acid |
| EGCG | Epigallocatechin gallate |
| ELISA | Enzyme-linked Immunosorbent Assay |
| EMSA | Electrophoretic Mobility Shift Assay |
| FITC | Florescien-5-isothiocynate |
| FTIR | Fourier-transform infrared spectroscopy |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| HP | High performance |
| HRP | Horseradish peroxidase |
| HTS | High throughput screening |
| IMAC | Immobilised metal affinity chromatography |
| IPTG | Isopropyl β-D-1-thiogalactopyranoside |
| kDa | kilodalton |
| LB | Lysogeny broth |
| LIR | Long intergenic region |

| | |
|---|---|
| MDG | Millennium development goal |
| $MgCl_2$ | Magnesium chloride |
| min | Minutes |
| miRNA | MicroRNA (miRNA) |
| MP | Movement protein |
| NaCl | Sodium chloride |
| NSP | Nuclear Shuttle protein |
| OD | Optical density |
| ORF | Open reading frame |
| PBS | Phosphate-buffered saline |
| PDR | Pathogen-derived resistance |
| piMRNA | Piwi-interacting RNA (piRNA) |
| PMSF | Phenylmethylsulfonyl fluoride |
| PTGS | Post-transcriptional gene silencing |
| RCR | Rolling cycle replication |
| REn | Replication enhancement protein |
| Rep | Replication-associated protein |
| RICS | RNA-induced silencing complex |
| RNA | Ribonucleic acid |
| SDS-PAGE | Sodium dodecyl sulfate- Polyacrylamide gel electrophoresis |
| SIR | Short interic region |
| siRNA | Small interfering RNA |
| sRNA | small RNA |
| ssDNA | Single stranded DNA |
| TBS | Tris-buffered saline |
| TMB | 3,3',5,5'-Tetramethylbenzidine |
| TrAP | Transactivation protein |
| TYLCV | Tomato yellow leaf curl virus |
| UK | United Kingdom |
| UV | Ultra-violet |
| vsiRNA | Viral-derived small interfering RNA |
| FAOSTAT | The Food and Agriculture Organization Corporate Statistical Database |
| BVI | |

| | |
|---|---|
| RRB1 | Retinoblastoma related protein |
| VSRs | Viral suppressors of RNA silencing |
| ROS | Reactive oxygen species |
| DSB | Double stranded breaks |
| SOC | Super optimal broth |

# Chapter 1: **Introduction**

## 1.1 Food security: A global food crisis

Food security is a major problem worldwide, especially in African countries. The inability to readily access sufficient nutritional food physically or economically is due to several reasons such as lack of income, increasing population, climate change, and political instability and plant diseases. It is estimated that by the year 2050, the world population will increase to over 9 billion people with sub-Saharan Africa contributing the highest proportion of population increase and thus increasing food insecurity (United Nations, 2013; FAOSTAT, 2016). Therefore, there is an urgent need to come up with strategies to eradicate poverty and hunger. During the United Nations 2013 summit, world leaders agreed that the first-millennium development goal is to eradicate poverty and hunger (MDG, United Nations, 2013). South Africa is one of the countries that committed itself by means of the constitution, sections 26 and 27 of 1996 that states that everyone has a right to access sufficient and nutritional food (DoA, 1997). The mandate was given to the Department of Agriculture, Forestry and Fisheries to ensure that policies and programs are drawn and implemented that enable every South African to meet their basic need and that these policies foster technological development, farm management, trade, and growth. While an increase in population is a dominant factor in food security, plant pathogenesis, which is defined as diseases affecting plants, has also been contributing greatly to food insecurity by degrading plant production and quality. The major groups of pathogens affecting plants are viruses, bacteria, fungi, nematodes, oomycetes and parasitic plants (Strange and Scott, 2005). Plant disease is a huge problem because of its detrimental effects on the productivity and quality of most crucial agricultural crops including but not limited to maize, sweet potatoes, tomatoes, and cassava. Cassava has the potential to close the poverty and hunger gap in the world, especially in sub-Saharan Africa where it is mostly produced because of its nutritional value (Rey and Vanderschuren, 2017; Widido, 2018).

## 1.2 Cassava (*Manihot esculenta* Crantz)

Cassava is one of the most significant starches and protein-rich crops in the world. Cassava is a drought-tolerant crop that is cultivated on marginal soil and high crop yields are produced under ideal conditions (EL-Sharkawy, 2007). It provides over half a billion people with calories and nutrition and is thus a crucial staple food in several continents including Southern America, Asia and Africa (Daniell *et al*., 2008; Bredeson *et al*., 2016). Amongst the starchy staples,

cassava provides a carbohydrate production of 25% more than maize and 40% higher than rice, making it a low-cost source of calories for both animal feed and human nutrition (Tonukari, 2004). It is used for many purposes, including but not limited to making flour, cassava alcohol, pudding, caramel, animal feed thus its advantage in providing income and industrial application.

Cassava is a semi-perennial crop, allowing its roots to remain dormant underground for up to three years, and resistant to carbon dioxide concentration (Allem and Genèticos, 2002). These advantages are recognised by farmers all over the world, making it a very popular crop in the 21st century. This is particularly true in Africa, where over 110 million tonnes of cassava were produced in 2014, which was 53.7% of the world's total production (FAOSTAT, 2016). Cassava is an important crop that can mitigate the effects of famine worldwide (Chavarriaga-Aguirre *et al*., 2016; Tao *et al*., 2019). It is a staple starch that is rated fourth after maize, rice, and wheat but its production is facing high economic challenges due to viral diseases (Rey and Vanderschuren, 2017). Like many agricultural crops today, major economic challenges to cassava production include challenges by diseases such as tobacco mosaic disease, tomato curl leaf disease and cassava mosaic disease. Cassava mosaic disease is the most significant disease affecting cassava in Africa, and it is caused by Geminiviruses (Fondong, 2017).

## 1.3 Geminiviruses

### 1.3.1 *Geminivirus*: genome organization

Geminiviruses are the largest genus belonging in the family Geminiviridae characterised by their circular and single-stranded DNA (ssDNA genome). They infect a wide range of monocotyledon (monocots) and dicotyledon (dicots) crop species such as maize, cassava, tomato and cotton in tropical and temperate climate region (Zerbini *et al*., 2017). Geminiviruses are divided into nine genera based on the genome organisation, host range, vector and sequence relationships by the International Committee on the Taxonomy of Viruses (ICTV) namely to *Begomovirus*, *Curtovirus*, *Mastrevirus*, *Topocuvirus*, *Becurtovirus*, *Capulovirus*, *Eragrovirus*, *Grablovirus*, and *Turncurtovirus* depicted in Figure 1.1 ( Torres-Herrera *et al*., 2019; Ramesh *et al*., 2017; Zerbin *et al*., 2017 and Brown *et al*., 2012). The first four being the main genera identified, the last five are recent additions to the phylogeny. In this study, the main focus is on begomoviruses. Begomoviruses have either monopartite or bipartite genome. The monopartite (see Figure 1.1 G) has DNA-A like genome while bipartite (see Figure 1.1 G) is composed of two

separate genomes designated DNA-A and DNA-B with the size ranging from 2.6 to 3.0 kb. The two genome components share nearly 200 bases of sequences in the long intergenic region (LIR) that includes the replication origin. Transcription of these genes is under the control of one monodirectional promoter and two bidirectional promoters to produce functional proteins (Fondong, 2013). The DNA-A retains four open reading frames (ORFs) in the complementary strand which encode for coat protein (CP), the replication-associated protein (Rep), transactivation protein (TrAP) and enhancement protein (REn); while the DNA-B has only two ORFs coding in the virion strand for the nuclear transport protein (NSP) and cell-to-cell movement protein (MP**) (**Ye *et al*., 2014; Luna *et al*., 2017; Snehi *et al*., 2017).

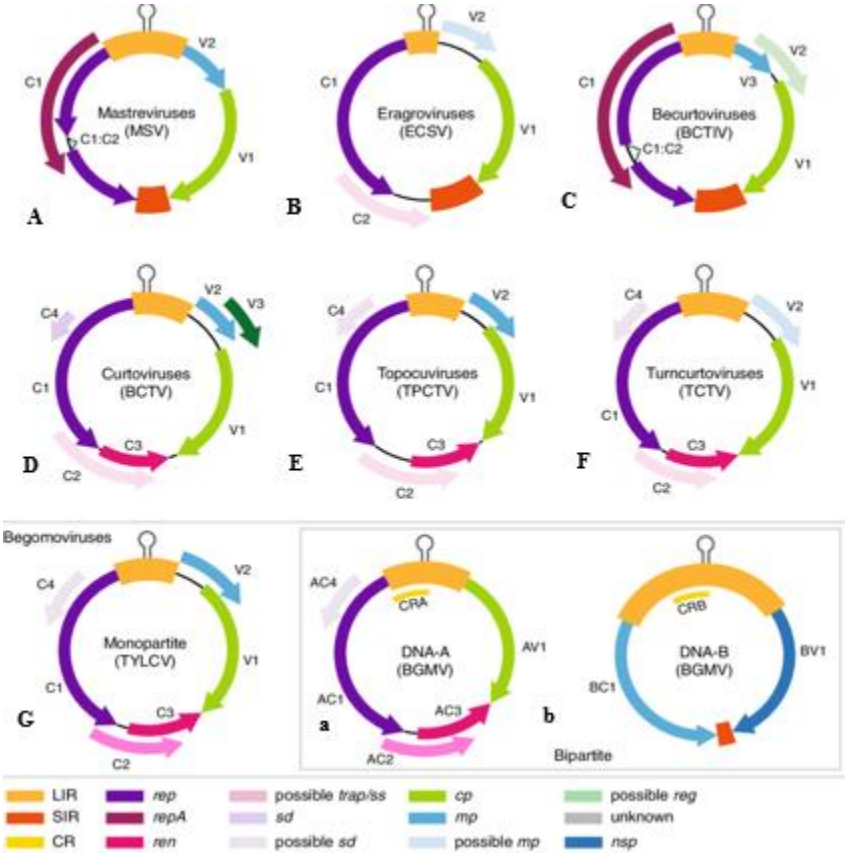



Figure 1.1: Genomic organisation of different types of Geminiviruses.

*Matrevirus* (A), *Eragrovirus* (B), *Becurtovirus* (C), *Curtovirus* (D), *Topocurtovirus* (E), *Turncurtovirus* (F) and *Begomovirus* (G). Among all the above-mentioned Geminiviruses, Begomovirus which is the largest in the Geminiviridae family has two types of viruses: those with a monopartite genome (a) and those with bipartite genome, DNA-A and DNA-B (b). The genome organisation is composed of Open Reading Frames (ORFs) coding for different proteins (*C1*, *C2*, *C3*, and *C4*), Long interic region (LIR), Short interic region (SIR), and Common region (CR) (Zerbini *et al*., 2017).

### 1.3.2 *Geminivirus* proteins

The above-mentioned proteins are transcribed and translated during different stages of the life cycle. *AC2* (also known as *AL2* or *C2*) gene coding for transcriptional activator protein (TrAP) is transcribed at an early stage of the viral life cycle. This gene is present in all *Geminivirus* family and it is a crucial gene because, it acts as a transcriptional factor for the expression of coat protein and *BR1* genes (Hartitz *et al*., 1999; Harrison *et al*., 2002). The TrAP has also been shown to suppress the innate defense mechanism of plants against nucleic acids such as viruses in a transcriptional dependent and non-dependent

mechanism. Results found by Trinks *et al*. (2005) have suggested that TrAP suppresses RNA silencing through activation of cellular proteins that function as regulator of the system in a negative feedback loop. This is as a result of *Arabidopsis* protoplasts transformed with *Mungbean yellow mosaic virus* (MYMV) or *African cassava mosaic virus* (ACMV) during transcriptional profiling. Two of the most important proteins for replication of the viral DNA are the Replication-association protein (Rep) and Replication enhancement protein (REn). *AC1* is also known as *AL1* or *C1* gene encoding Rep that functions as an endonuclease upon binding to the double-stranded DNA and as a ligase that initiates and terminates the rolling circle replication (RCR) (Castillo *et al*., 2003). *AC3* also designated *AL3*, or *C3* encodes for the REn which interacts with the Rep and is associated with enhancing viral infection as well as symptoms (Morris *et al*., 1991; Castillo *et al*., 2003). A gene that is expressed at a later stage is the *AV1*, which codes for a coat protein (CP) that functions as a shell for viral DNA to facilitate the movement of viral DNA into the nucleus through the nuclear pores (Priyadarshini *et al*., 2011; Zerbini *et al*., 2017). The last two proteins that facilitate the movement of viral DNA are the NSP and MP. *BVI* gene codes for NSP that facilitates the movement of viral DNA from the nucleus to the cytoplasm, while the MP facilitates the movement of viral DNA from one cell to the adjacent cell in the viral life cycle (Florentino *et al*., 2006). All these protein-coding genes are depicted in Figure 1.1.

### 1.3.3 *Begomovirus* life cycle

*Begomovirus* is transmitted via the whitefly *Bemisia tabaci* (Gennadius) vector in a circulative persistent manner. It transmits the viral particles with its stylets while feeding on the phloem of plants (see Figure 1.2 A). Once the virus enters the plant cytoplasm, the virus is uncoated and releases the ssDNA. The ssDNA then enters the nucleus through nuclear pores (see Figure 1.2 B). The ssDNA acts as a template for the synthesis of a complementary strand using RNA Polymerase II to form dsDNA intermediate which is transcribed in the nucleus and translated for viral proteins within the cytoplasm (Han *et al*., 2007). The transcribed proteins such as Rep and REn enter the nucleus and form a complex; enhancing Rep mediated ATPase activity when bound to the origin of replication of the dsDNA - initiating rolling cycle replication (RCR) which is not well understood (Pasumarthy *et al*., 2010). Once copies of circular ssDNA are produced by the

RCR, they either enter the replication cycle again or encapsulated by the transcribed coat protein. Transcription of the coat protein is activated by TrAP. After viral ssDNA encapsulation, the virion is transported from the nucleus to the cytoplasm to be ingested by the whitefly while feeding on the phloem of the infected cassava leaves. The whitefly hosting the virus will then infect another cassava, and then the viral cycle continues. Some of the symptoms of the infection include yellow pigmentation of the leaves and stunting of the roots (see Figure 1.2 C). All the proteins involved in the lifecycle of *Begomovirus* are potential targets for antiviral strategies. Therefore, Rep protein which is indispensable for replication of the virus has been pursued further as a potential target (Rizvi *et al*., 2015).

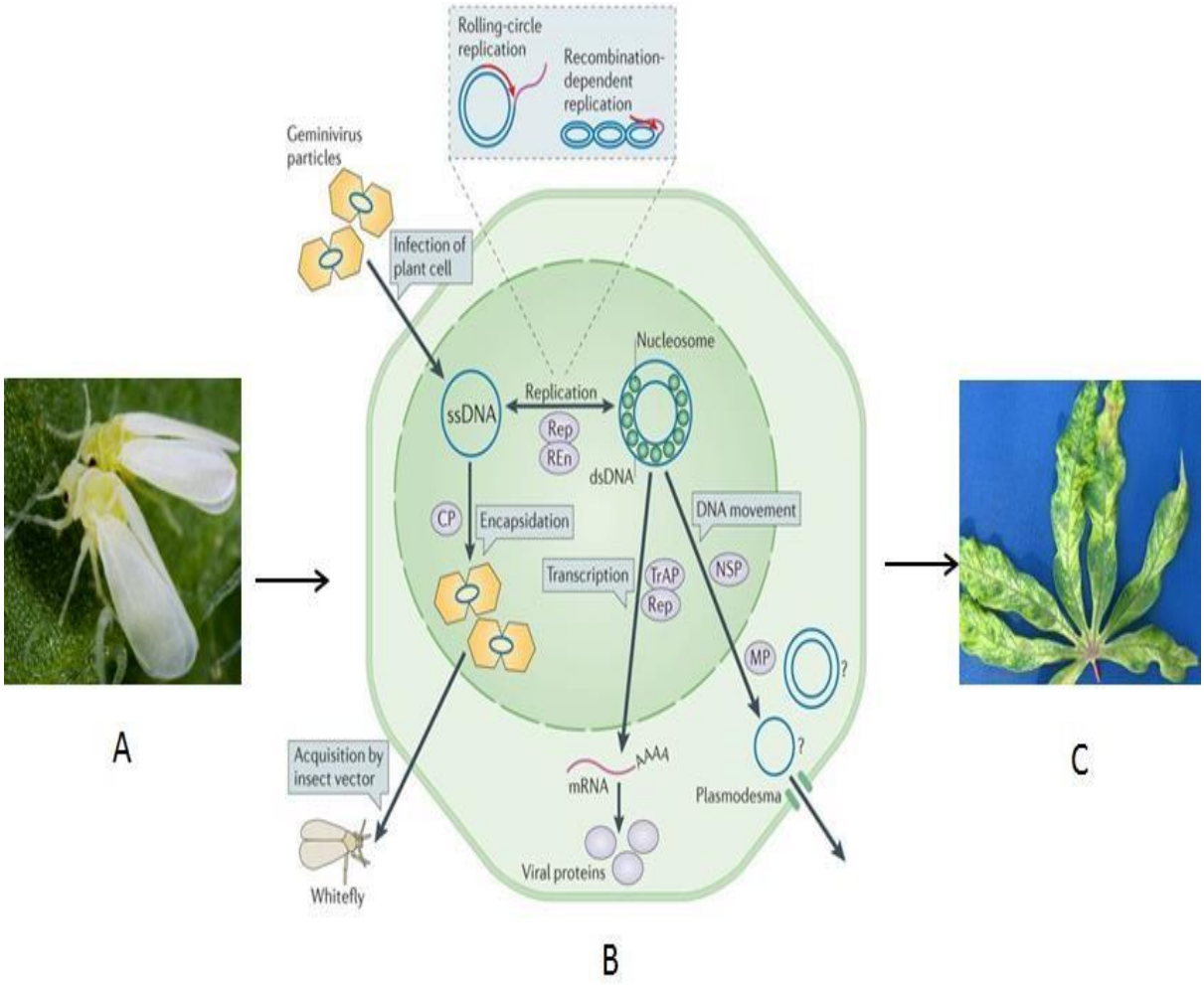



Figure 1.2: Transmission of the virus onto cassava by the whitefly.

The whitefly transmits the viral particle using its stylets while feeding on the plant phloem (A). The virus particle raptures and releases the ssDNA and its functional protein such as Rep. Rep is responsible for the replication of the virus by forming a complex with the REn and binding to the DNA thereby recruiting the host cell factors to assist with replication via the rolling cycle mechanism. It also forms a complex with the transactivation protein to assist in the transcription and translation of viral proteins while the CP coats the ssDNA to be ingested by the whitefly and then reinfection will only when in contact with another cassava plant (B). (C) Represents the African cassava plant infected by the virus and the symptoms after being infected. Adapted from (Adapted from Hanley-Bowdoin *et al*., 2013).

*1.3.3.1 Replication initiator protein and its functions*

Replication-association protein (Rep) encoded by the AC1 open reading frame of DNA-A region is the only viral factor that is essential for viral replication by the RCR mechanism. It is a 40 kDa multifunctional protein that has DNA binding and cleavage activity (Wegrzyn *et al*., 2014). The binding and cleavage site of Rep lie in the N-terminal domain of the protein. Rep protein interacts with other viral proteins like REn, host proteins such as RRB1 and with itself (Orozco, 1997). Rep protein binds to double-stranded DNA at the origin of replication (TAATATT/AC sequence) of the plasmid and nicks the (+) strand leaving the 5' adduct to form a covalent bond with the hydroxyl group of $tyrosine_{103}$ residue identified (see Figure 1.3) (Laufs *et al*., 1995; Orozco *et al*., 1998). After nicking, Rep recruits host factors (pRBR, RF-C, PCNA and histone H3) that assists in the completion of the replication process (Pant *et al*., 2001; Tara *et al*., 2010). Host DNA polymerase III starts the synthesis of a new (+) strand by extending the 3' end simultaneously replacing the old (+) strand, and then Rep protein ligates the gaps. After completion of the (+) strand, Rep cleaves and ligates the ends of the displaced DNA to produce ssDNA. To complete the replication cycle, the (-) strand is synthesised by host cell factors using a separate single-stranded origin of the plasmid (Orozco *et al*., 2000). In general, information about the structure of Rep protein is quite limited and only the N-terminus of TYLCV was resolved (Campos-Olivas *et al*., 2002). One study tried to resolve the crystal structure of CaLCuV $Rep_{1-121,}$ CaLCuV $Rep1_{1-177}$ and CaLCuV $Rep1_{1-349}$ proteins by Nuclear Magnetic Resonance (NMR) without success (Tara *et al*., 2010). Oliveira *et al*. (2017), studied viral proteins such as NS3 and NS4 which are essential for viral DNA replication as potential targets for the development of antiviral compounds for Dengue and Chikungunya viruses, respectively. Consequently, this provides a confirmation of using Rep as a potential target for the development of antiviral agents and further characterisation of ACMV Rep needs to be explored to better understand the Rep interaction with other molecules (DNA, proteins and small molecule inhibitors).

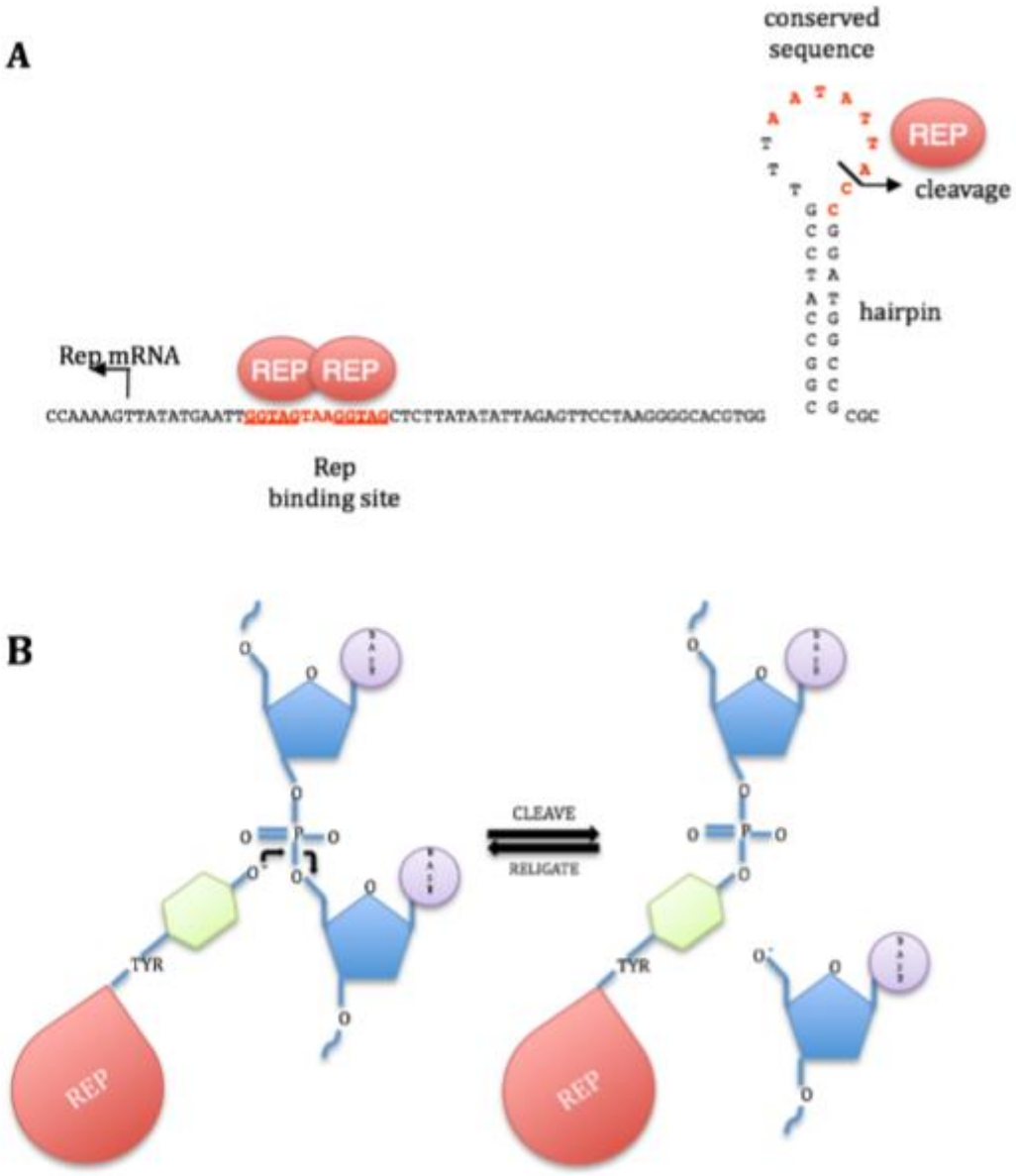



Figure 1.3: A typical demonstration of ACMV DNA origin of replication.

The Rep protein and its recognition site (GGTAGAAGGTAG) for binding are shown in red. The hairpin loop with the cleavage site (AATATTA CC) that is conserved throughout the *Geminivirus* family is indicated in red (A). The 5' end of the cleaved ssDNA fragment forms a covalent bond with tyrosine hydroxyl group of the Rep protein (Tara Nash, PhD thesis, 2010).

## 1.4 Strategies on controlling Geminiviruses

### 1.4.1 Natural antiviral mechanism of plants

#### *1.4.1.1 RNA silencing*

RNA silencing is a natural mechanism of plants to defend themselves against attacks by foreign nucleic acids such as viruses. This was discovered in the early 1990s where a plant transformation experiment was done to introduce transgenes into the genome; this resulted in the silencing of both the homologous host genes and transgenes (Van de Kroll *et al*., 1990). The introduction of transgenes leading to RNA silencing was termed post-transcriptional gene silencing (PTGS) (Tabassum *et al*., 2012; Wang *et al*., 2011). The main role-players in RNA silencing are the virus-derived small RNA (sRNA) and associated effector proteins (AEP). These work together to inactivate viruses in a sequence-specific manner by degrading its genome using a number of proteins such as ribonuclease III (an enzyme that targets RNA and breaks it down) via a cascade of events (Pantaleo *et al*., 2007).

The sRNA is classified into three main classes namely: small interfering RNA (siRNA), microRNA (miRNA) and Piwi-interacting RNA (piRNA) according to their origin, structure, associated effector protein (AEP) and their biological functions (Castel and Martienssen, 2013). Upon attack of a plant by the virus, a long double-stranded RNA (complementary strands and passenger strands) is produced from the replication intermediates of viral RNA (Vanitharani *et* al., 2004). The double-stranded small RNA acts as a substrate for an enzyme (ribonuclease) called Dicer-like protein (DCL) that comes from the cytoplasm to produce 21-24 nucleotide (nt) viral-derived small interfering RNA (vsiRNA) (see Figure 1.4). The interaction of Argonaute family protein (AGO) and the vsiRNA form a multicomplex called RNA-induced silencing complex (RICS). Argonaute family protein (AGO) which is a catalytic component, is recruited and incorporated into the RISC complex. This formation of the RISC-AGO complex will then induce either DNA methylation, repression or degradation (Tabassum *et al*., 2012; Zotti *et al*., 2018).

However, viruses have developed a mechanism that counteracts the RNA silencing defense mechanisms of plants by encoding viral RNA silencing suppressor proteins that interrupts various steps of RNA silencing pathways. These viral suppressors of RNA silencing (VSRs) suppress by seizing of virus-derived siRNAs and inactivation of AGO proteins mainly AGO1 (see Figure 1.4) (Chang and Wang, 2016). Some of the suppressor proteins are the 2b protein of *cassava mosaic virus* (CMV), P21 protein of *Beet yellows virus* (BYV), P19 protein of

*Tombusvirus*, V2 protein of *Tomato yellow leaf curl virus* (TYLCV) and *AC2* protein and *AC4* protein of *Geminivirus* to mention a few (Vanitharan *et al*., 2004; Martínez-Turiño and Hernández, 2009; Csorba *et al*., 2009). The fact that the virus encodes for viral suppressors of RNA silencing mechanisms put forward the basis of searching for a new effective method to control these diseases.

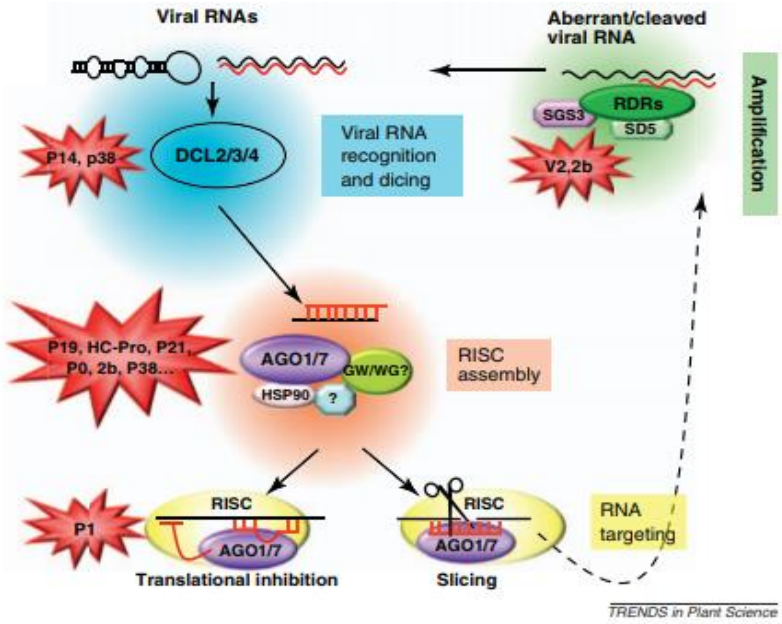



Figure 1.4: RNA silencing mechanism of plants.

The RISC-AGO complex is formed for the methylation, degradation or repression of genetic molecules. The different RNA silencing repressors (P14, P13, P19, P21, 2b, P38, P0, HC-Pro and P1) encoded by the virus interrupt the process of RNA silencing at various stages (Red) (Burgyán and Havelda, 2011).

*1.4.1.2 Plant antioxidants*

Plants have developed an antioxidant system to protect themselves from pathogens. These are enzymatic [decomposition of reactive oxygen species (ROS)] and endogenous non-enzymatic systems [tocopherols, carotenoids, ascorbic acid (vitamin C), several phenolic compounds and tocotrienols (vitamin E)] (Kulbat, 2016). A special group of compounds named phenolics has attracted scientists' attention as they exhibit a protective effect against mechanical damage in

purple-flesh potatoes (Reyes and Cisneros-Zevallos, 2003). These phenolics act as shielding agents, and pesticides against pathogenic organisms.

Phenolics are a heterogeneous group of compounds containing one or more aromatic rings bearing hydroxyl groups with over 8,000 structural variants (see Figure 1.5). They are secondary metabolites found mostly in plants such as fruits and vegetables and are produced through the shikimic acid and phenylpropanoid pathways (Saranraj *et al*., 2019). They contribute to the environmental and physiological functions of the plant as well as to produce flavour, colour and astringency. Additionally, the antioxidant activities of these compounds are associated with the structure of phenolic compounds, largely depending on the number and location of hydroxyl groups and glycosylation (Huang *et al*., 2009). The bioactivity of these compounds has been explored in the inhibition of pathogen infection. Patzke and Schieber (2018) have reported successful growth inhibition of different phytopathogenic fungi by a natural phenolic compound (ferulic acid). This compound was used as one of the active ingredients for the preparation of emulsified pesticides. Compounds extracted from S. *mombin* and S. *tuberosa* have been found to inhibit Type-2 *Dengue virus*. The study suggested that rutin and quercetin have the potential to be developed as an anti-DNV agent (Silva *et al*., 2011).

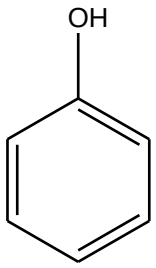



Figure 1.5: The structure of a phenol group.

### 1.4.2 Man-made antiviral mechanism in plants

#### *1.4.2.1 Pathogen-derived resistance (PDR)*

Strategies in controlling either the vector (whitefly) or the virus (*Begomovirus*) are still ongoing. The breeding of wild type cassava, *Manihot glaziovii* Muell.-Arg into cultivated cassava to confer resistance is by far the most successful strategy (Fondong, 2017). Sanford and Johnston (1985) came up with transgenic strategies where cassava was genetically engineered for resistance which was named pathogen-derived resistance (PDR); however, this still has to go through the approval by the government and broad public acceptance (Stanford and Johnston, 1985; Orozco, 2018; Dong and Ronal, 2019). Further studies on PDR also

revealed that plants use RNA silencing as an antiviral defense mechanism (Vance, 2001). *Tomato yellow leaf curl virus* (TYLCV) and *African cassava mosaic virus* (ACMV) encode *AC2* proteins that reverse the silencing of RNA viral defense mechanisms already set by the plant. Therefore, more research is focusing on RNA silencing, transgenic and non-transgenic Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9 systems to confer plant resistance to Geminiviruses.

### *1.4.2.2 Gene editing: Non-transgenic Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) system*

A new technique has recently emerged as an alternative to control plant disease other than plant breeding and transgenic methods. This technique is one of the highly favoured in plant biotechnology because it is easy to design and reagents are easily accessible, more versatile, high success rate and less expensive hence speeding up the generation of cassava cultivars with better-quality traits (Borrelli *et al*., 2018; Lentz *et al*., 2018; Gomez *et al*., 2019). It is called the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9 (CRISPR associated protein 9). CRISPR/Cas9 edits the target plant genome in a site-specific manner using a recruited site-specific nuclease called Cas9 (Xing *et al*., 2014). Upon binding of Cas9 to the target genome, a double-stranded break (DSB) is created within the targeted DNA. This DSB is then repaired through the error-prone, non-homologous end joining ((NHEJ) or homology-directed recombination (HDR) pathways causing either insertion or deletion (INDEL) mutations (Jaganathan *et al*., 2018).

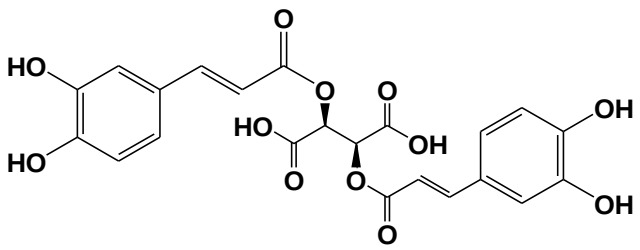
The CRISPR/Cas9 technology has been applied to *Arabidopsis thaliana* and *Nicotiana benthamiana* to confer resistance to single-stranded DNA (ssDNA) *Geminivirus* (Li *et al*., 2013; Upadhyay *et al*., 2013; Chandrasekaran *et al*., 2016). CRISPR/Cas9 was recently applied on transgenic cassava plant that was transformed to express the Cas9 protein together with sgRNA1 and the Cas 9 (control). This showed a failure to confer effective resistance to *African cassava mosaic virus* (ACMV). It also led to the emergence of a new, conserved mutant virus that cannot be cleaved by the catalytic CRISPR/Cas9. It was advised by the authors based on their findings that caution on the use of CRISPR/Cas9 technology to engineer resistance should be considered in both greenhouse and field settings. This should be done in order to avoid the evolution resistance of viruses (Mehta *et al*., 2019).

*1.4.2.3 Naturally-derived synthetic compounds*

Natural compounds have formed the basis of product development for the pharmaceutical, food, cosmetic and insecticides industries. Natural compounds and their derivatives are synthesised and modified to enhance the activity of these compounds for the health benefits of humans, animals, and plants. These compounds are assessed for activity in vitro as potential inhibitors or enhancers. Chicoric acid (CA) and Epigallocatechin gallate (EGCG) will be described in detail below because they are the main focus of this study as potential inhibitors of ACMV Rep interaction with viral DNA.

Chicoric acid named 2, 3-bis ({[(2Z)-3-(3, 4-dihydroxyphenyl) prop-2-enoyl] oxy}) butanedioic acid according to the International Union of Pure and Applied Chemistry (IUPAC) nomenclature is a polyphenol compound, a derivative of caffeic acid and tartaric acid (see Figure 1.6). It was discovered from the aerial parts, the roots and seeds of Chicory (*Chicorium intybus* L.) by Scarpati and Oriente in 1958. Since its discovery, it has been found in other plants such as the aerial part of *Echinacea purpurea* L (Thomsen *et al*., 2012; Wills and Stuat, 1999) and leaves of *Lactuca sativa* L (Abu-Reidah *et al*., 2013). It has been reported that CA assists plants to protect itself from viral, bacterial and fungal infections (Bauer, 1998; Nishimura and Satoh, 2006; Charvat *et al*., 2006).

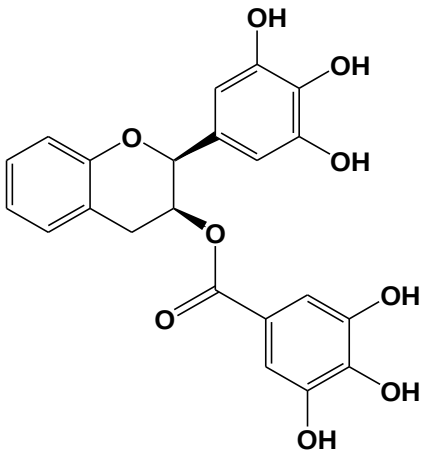
It is also reported that CA has biological activity in inhibiting HIV integrase, resulting in the possibility of hindering the replication of the virus genome (Charvat *et al*., 2006). Studies have shown that, CA exhibits fluorescence quenching of the lactoferrin protein that was used as the basis for the construction of LF and CA nanoparticles (LF-CA-NP) (Li *et al*., 2018). In the current study, we determined the inhibitory properties of CA against the ACMV target protein Rep through biochemical studies.




**Chicoric acid**

Figure 1.6: Chemical structure of the polyphenolic compound, Chicoric acid.

Epigallocatechin gallate (EGCG) is the most abundant antioxidant in green tea from a plant scientifically known as *Camellia sinensis* L. of the Theaceae family (see Figure 1.7). Green tea is an ancient tea that is widely consumed, especially in East Asia, because of its powerful antioxidant activity (Du *et al*., 2012; Chowdhury *et al*., 2016). It is a polyphenolic compound formed by a flavanol core (flavan-3-ols) structure fused with a gallocatechol group and a gallate ester. It constitutes 65 % of the total catechin contents (Islam, 2012). In a study, 10 polyphenol compounds were tested for antiproliferative effects, EGCG showed the most potent activity against colorectal cancer. Compared to the antiproliferative effect of 5-fluorouracil, which is a chemotherapy drug on colorectal cancer, EGCG is even more potent (Du *et al*., 2012).




**Epigallocatechin gallate**

Figure 1.7: Chemical structure of the polyphenolic compound (−) - epigallocatechin gallate (EGCG).

## 1.5 Rationale of the study

ACMV is an important concern with regards to food security in sub-Saharan Africa, as it diminishes a staple food crop in the region, namely cassava. The attack of cassava by the mosaic virus also has economic implications. Understanding the mechanism of replication of the virus is essential so as to interfere with its replication and spread of the disease (Rizvi *et al*., 2015). There is an urgent need for the development of novel strategies to control the virus. Most DNA viruses use either one or more viral enzymes for replication but in some instances, host cellular enzymes are recruited for the successful completion of replication. These enzymes (host and/or viral) are, therefore, potential targets for the development of effective antiviral agents. Rep protein is an obvious target as it is indispensable for viral replication. The identification of small molecule inhibitors can be achieved by understanding the structural and biochemical functions of Rep in the life cycle. This study will add to the body of knowledge that already exists on the Replication-association protein in *Geminivirus*.

## 1.6 Hypothesis

The interaction of Rep protein with viral DNA can be disrupted by small molecule inhibitors.

## 1.7 Aim

To study *Geminivirus*' replication-association protein (Rep) as a target for the development of small molecule inhibitors.

## 1.8 Objectives

- Overexpression and purification of the recombinant Rep protein
- Biochemical and biophysical characterisation of the Rep protein
- Development of high-throughput screening (HTS) assay(s)

# Chapter 2: **Methodology**

## 2.1 Overexpression of the recombinant ACMV Rep

### 2.1.1 Nucleotide and amino acid sequences coding for ACMV Rep protein

For overexpression of ACMV Rep protein, ACMV DNA sequence was obtained from NCBI (https://www.ncbi.nlm.nih.gov/nucleotide/) with an accession number >NC_001467.1:c2756-1680. The DNA sequence was then translated to the amino acid (aa) sequence using the ExPASy translating tool (https://web.expasy.org/translate/) and NCBI ORF finder (https://www.ncbi.nlm.nih.gov/orffinder/ ). From the aa sequence, ExPASy was then used to predict the theoretical molecular weight, pI and extinction coefficient of the Rep protein.

### 2.1.2 ACMV Rep molecular clone construction

ACMV DNA sequence was sent to GenScript for synthesis and cloning into pET-15b (Novagen, Darmstadt, Germany). The ACMV DNA encoding for Rep protein was cloned between *NdeI* and *BamHI* restriction sites of the pET-15b vector. Upon arrival to our lab, the DNA was stored at -20ºC.

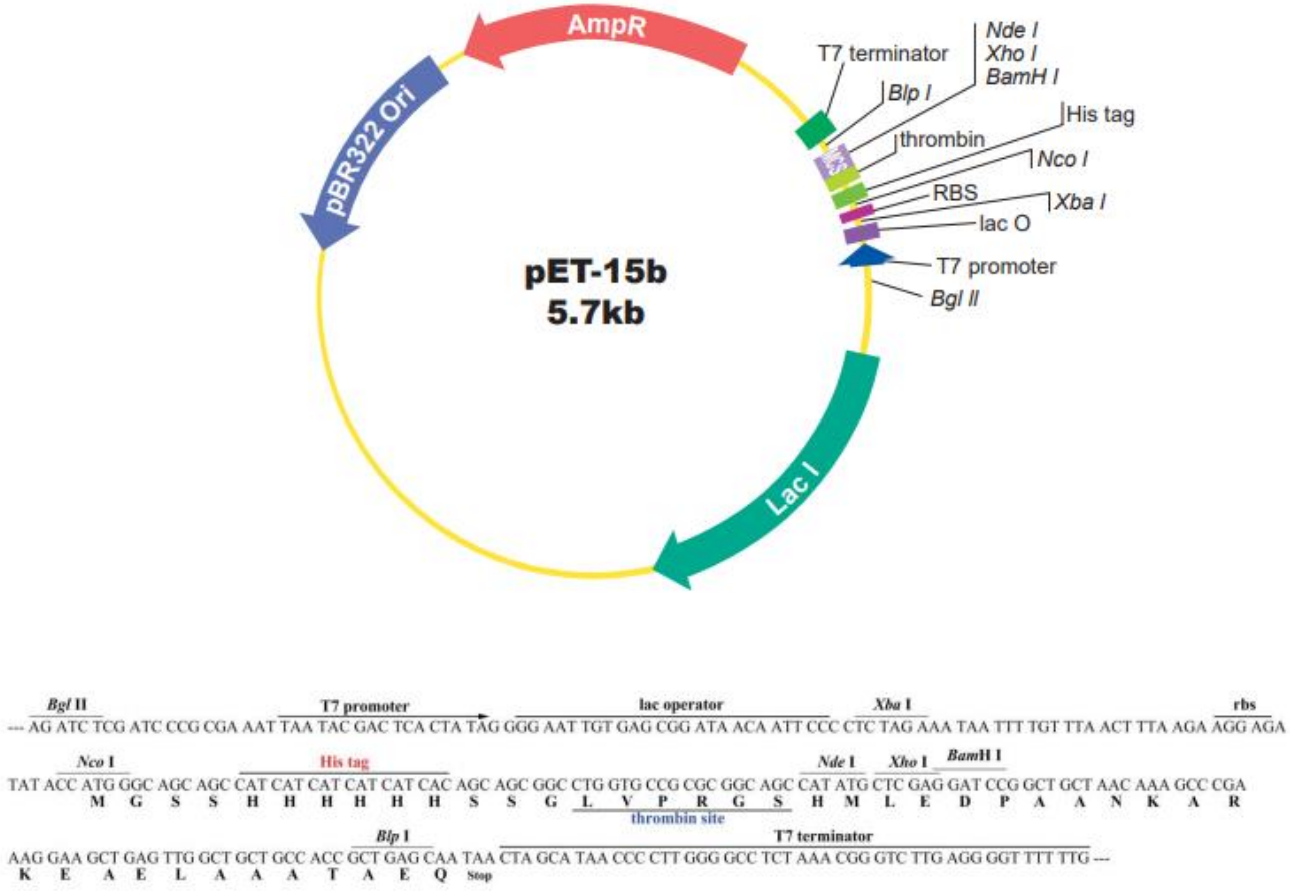



Figure 2.1: pET-32a expression vector map showing the multiple cloning sites.

### 2.1.3 Bacterial transformations

BL21 (DE3) pLysS *E. coli* competent cells (Novagen, Darmstadt, Germany) were transformed with pET-15b vector harbouring Rep recombinant DNA insert (GenScript, USA) using the heat shock method. Briefly, competent cells were thawed on ice. The thawed cells were then gently mixed and 50 µL of the cells pipetted in a transformation tube containing 1 µL (200 ng) pET-15b Rep DNA plasmid and incubated on ice for 30 minutes. The transformation tube with the mixture was heat-shocked at 42 ºC for exactly 45 seconds and the tube was placed on ice for 2 minutes. The mixture was diluted 1:20 with SOC media (New England Biolabs Inc, USA) and incubated for 1 hour at 37 ºC. The diluted mixture was then spread on lysogeny broth (LB) agar (Laboratorios Conda, Spain) plates with appropriate antibiotics [100 µg/mL ampicillin (Melford Laboratories Ltd) and 35 µg/mL chloramphenicol (Calbiochem, Darmstadt, Germany)] and incubated overnight at 37 ºC.

### 2.1.4 Glycerol bacterial stock production

A single colony of the transformed bacterial cells was picked from the LB agar plate and inoculated into 100 mL of LB media containing appropriate antibiotics (100 µg/mL and 35 µg/mL chloramphenicol). The solution was incubated in a shaking incubator at 250 rpm for 5 hours at 30 ºC. The bacterial culture was then aliquoted to a final 1 mL glycerol (Sigma-Aldrich, USA) stock by adding 300 µL of 60% glycerol to 700 µL of bacterial cells. These stocks were stored in a -80 ºC freezer for long term storage.

### 2.1.5 DNA plasmid purification

For subsequent DNA sequencing and restriction enzyme digestion, DNA plasmid purification was performed. This is according to StrataPrep Plasmid Miniprep manufacturer's kit (Agilent Technologies, USA). Briefly, LB culture medium was inoculated with a single bacterial colony of BL21 (DE3) pLysS cells, the appropriate concentration of antibiotics was added and cells were incubated overnight at 37 ºC with vigorous shaking. The cell cultures were then aliquoted into a 1.5 mL microcentrifuge tube (Eppendorf, Hamburg, Germany) and centrifuged for 1 minute at 16 000 ×$g$ , the supernatant was discarded. A volume of 100 µL solution 1[50 mM Tris-HCl (pH 7.5), 10 mM EDTA and 50 µg/ml of RNase A] was added to the tube to resuspend the culture pellet and disperse the cells. A volume of 100 µL solution 2 [0.2 M NaOH, 1% (w/v) SDS] was added to the tube and mixed by gently inverting several times; the 125 mL of solution 3 [10 mM Tris HCl (pH 7.5), 100 mM NaCl, 2.5 mM EDTA] was also added and mixed gently. The mixed sample was centrifuged for 5 minutes at 16 000 ×$g$, the supernatant which contains the plasmid DNA was then decanted into a microspin column that is inserted into a 2 mL receptacle tube. The supernatant was centrifuged for 30 seconds at 16 000 ×$g$ and the filtrate discarded. A wash step was performed where 750 µL of 1× nuclease removal buffer was added to the microspin column, centrifuged for 30 seconds at 16 000 ×$g$ and filtrate discarded. The sample in a microspin column was centrifuged at 16 000 ×$g$ again and the filtrate discarded to remove all excess ethanol. The microspin column was transferred to a fresh 1.5 mL microncentrifuge tube, a volume of 50 µL elution buffer ($dH_2O$) was added directly on to the top of the fibre matrix of the microspin column and incubated for 5 minutes. The sample was then centrifuged for 30 seconds at 16 000 ×$g$ and the microspin column discarded, the filtrate sample was stored at -20 ºC.

2.1.6 DNA sequencing of ACMV gene encoding Rep protein

Purified plasmid DNA stored at -20 ºC was thawed at room temperature (25 ºC). A volume of 20 µL plasmid DNA was aliquoted in a 500 µL Eppendorf tube and sent for sequencing using the T7 forward primer and T7 terminator universal primer [Inqaba Biotechnical Industries (Pty) Ltd, South Africa] (see Table 2.1).

Table 2.1**:** Primers used for DNA sequencing

| **Universal primer name** | **Sequence (5'→3')** |
|---|---|
| **T7 promoter** | TAATACGACTCACTATAGGG |
| **T7 terminator** | GCTAGTTATTGCTCAGCGG |

2.1.7 Restriction enzyme double digestion

Restriction enzyme digestion assay was performed as a supplementary assay to DNA sequencing for confirmation of DNA insert into the pET-15b vector. All the reagents were thawed on ice briefly. All the components [5 µL of 1× FD buffer (Fermentas, USA), 1 µL plasmid DNA, 1 µL of restriction enzymes [(*BamHI* and *NdeI* ), (Fermentas, USA)] and 43 µL of nuclease-free water] of the reaction were mixed to a final volume of 50 µL in a microcentifuge tube. The reagents were tapped gently to mix and then centrifuged to settle down the reaction components. The reaction was incubated at 37 ºC in a water bath for an hour. To stop the reaction, a volume of 10 µL buffer [2.5% ficol 1-400 (Sigma-Aldrich, USA), 11 mM EDTA (Sigma-Aldrich, USA) (pH 8), 3.3 mM Tris-HCl (Sigma-Aldrich, USA), 0.017% SDS (Sigma-Aldrich, USA), 0.015% bromophenol blue)] was added to the reaction. The reaction samples along with 1 kb DNA ladder (New England Biolabs inc, USA) were analysed on a 1% agarose gel (Sigma-Aldrich, USA) in Tris-Acetate-EDTA buffer (pH 7.0). DNA migration was viewed on the ChemiDoc MP gel imager (Bio-Rad, USA).

2.1.8 Bacterial growth curve determination

An overnight culture of the selected transformants was grown in LB media with appropriate antibiotics (100 µg/ml ampicillin and 35 µg/ml chloramphenicol) at 37 ºC. Briefly, an overnight culture was diluted 100-fold into fresh LB broth with the appropriate

antibiotics. The growth rate was monitored by reading absorbance at 600 nm every hour. A growth curve of time (hours) and absorbance was plotted using Microsoft Excel 2013.

2.1.9 Gene induction for the over-expression of recombinant ACMV Rep protein

An overnight culture of the selected transformants [BL21 (DE3) pLysS] was grown in LB broth with appropriate antibiotics (100 µg/ml ampicillin and 35 µg/ml chloramphenicol) at 37 ºC. Briefly, an overnight culture was diluted 100-fold into fresh LB broth with the appropriate antibiotics. The culture was grown for 2.5 to 3 hours until it reached an optical density (OD) reading of 0.6 at 600 nm. After an OD is reached, the culture was induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Merk, Calbiochem, USA) and further incubated at 37 ºC in a 250 rpm shaking incubator for 3 hours. The cells were harvested in 50 mL tubes by centrifugation on a model 5810 R multipurpose centrifuge with A-4-62 rotor (Eppendorf, Hamburg, Germany) at 3200 ×$g$ for 30 minutes and stored as pellets at -80 ºC.

## 2.2 Confirmation of protein induction by Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis (SDS-PAGE)

A 12% Tris-Glycine SDS-PAGE (Bio-Rad, USA) was used to analyse the success of the expression with Mini-PROTEAN® II cell utilising the Laemmli method following staining with Coomassie Brilliant Blue (Laemmli, 1970). Briefly, protein samples along with protein molecular marker (Bio-Rad, USA) were loaded in the wells on a 12% Tris-Glycine SDS gel and a power of 150 V applied for migration of proteins based on size. Proteins were analysed on the ChemiDoc™ MP gel imager (Bio-Rad, USA) to validate the presence of the Rep protein.

## 2.3 Validation of ACMV Rep protein by Western blot

For Western blot analysis, proteins separated by SDS-PAGE were transferred to a nitrocellulose PVDF membrane (Life Technologies, USA) using an iBlot gel transfer system (Life Technologies, USA). The membrane was blocked with 5% milk (w/v) powder in TBS (Conda Pronadisa Molecular Biology, Spain) for 1 hour at room temperature (24 ºC), followed by overnight incubation with 1:2000 anti-6×His-tag mouse monoclonal antibody (Thermo Fischer, USA) diluted in blocking buffer at 4 ºC. The blot was washed five times for 5 minutes with TBS-T [20 mM Tris (Sigma-Aldrich, USA), 150 mM NaCl (Sigma-Aldrich, USA), 0.1% (w/v) Tween 20 (Sigma-Aldrich, USA)]. The horseradish peroxidase (HRP)-conjugated Goat

Anti-Mouse IgG (Thermo Fischer, USA ) was incubated for an hour as a secondary antibody with strep-actin (Thermo Fischer, USA) at 1:10 000 dilution, followed by washing with TBS-Tween five times for 5 minutes and developed using the 1:1 enhanced chemiluminescent (ECL) substrate (Bio-Rad, USA). The membrane was then viewed on the ChemiDoc™ MP system by exposure for 1 min.

### 2.4 Optimisation of Rep protein overexpression

The concentration of the Rep protein overexpressed was insufficient to perform further functional and structural studies. Therefore, induction studies which included changing parameters such as, bacterial competent strains [BL21 (DE3), BL21 (DE3) pLysS and T7 Express *E. coli* competence cells], expression time, media [LB broth (10 g Tryptone, 5 g Yeast Extract and 5 g NaCl), 2× YT (16 g Tryptone, 10 g, Yeast Extract and 5 g NaCl), and Terrific Broth (12 g Tryptone, 24 g Yeast Extract, 4 mL Glycerol, 9.4 g $K_2HPO_4$, 2.2 g $KH_2PO_4$)], inducer (IPTG) concentration and the addition of glucose (Sigma-Aldrich, USA). Different concentrations of filter sterilised glucose (0.1 %, 0.2 %, 0.3 %, 0.4 %, 0.5 %, 1%, 2 %, 3 %, 4 % and 5 %) were added to the LB media before inoculation of overnight bacterial cells. The bacterial cells were cultured under a controlled environment (37 ºC, 250 rpm on an orbital shaker) until an $OD_{600}$ of 0.6-0.8. Different concentrations of IPTG (0.1 mM, 0.5 mM, and 1 mM) were added to the bacterial culture to determine the optimal IPTG concentration needed to induce protein overexpression. The maximum time needed for optimal protein expression was determined by taking samples every hour after induction with IPTG. SDS-PAGE was used to analyse the effect of parameters (glucose, IPTG and time) on protein overexpression.

### 2.5 Purification of ACMV Rep protein from BL21 (DE3) pLyS bacterial cells

#### 2.5.1 Determination of Rep protein solubility

The solubility of the Rep protein was determined by resuspending bacterial cell pellet in ice-cold lysis buffer (50 mM HEPES (Sigma-Aldrich, USA) pH 7.5, 500 mM NaCl (Sigma-Aldrich, USA), 10% glycerol, 1 mM $MgCl_2$ (Sigma-Aldrich, USA), 1 mM PMSF (Sigma-Aldrich, USA), 4 mM CHAPS (Sigma-Aldrich, USA), 1 µl DNase I (Sigma-Aldrich, USA) and 10 µl lysozyme (Sigma-Aldrich, USA). The bacterial culture was grown at 37 ºC pre-induction ($OD_{600}$ of 0.6); the cultures were then grown at 37 ºC for 5 hours and at 16 ºC overnight post-induction. The bacterial cells were homogenised for 2 minutes using a homogeniser (Bandelin, Berlin) followed by cell lysis through sonication.

Briefly, the sonicator (Labsonic M ultrasonic processor, Sartorius, Germany) was used to break the cell walls for 30 seconds on the ice, 80% with 0.8 cycles. The process was repeated for six cycles and the lysate was centrifuged for 30 minutes, 12000 ×*g* at 4 ºC. The supernatant and pellet were sampled for solubility analysis by SDS-PAGE.

### 2.5.2 Extraction of Rep protein from inclusion bodies

#### *2.5.2.1 Solubilisation using urea*

To solubilise the Rep protein from inclusion bodies, the pellet was resuspended in unfolding buffer A [50 mM HEPES (Sigma-Aldrich, USA) pH 7.5, 500 mM NaCl, 10% glycerol, 8 M urea (Sigma-Aldrich, USA)] overnight at 37 ºC with gentle shaking. The sample was finally centrifuged at 16000 ×*g* for 30 min at 4 ºC to clear debris, and the supernatant was collected.

### 2.5.3 Purification and refolding of solubilised Rep protein

Purification of the target protein from unwanted bacterial proteins was done on immobilised metal affinity chromatography (IMAC) charged with $Ni^{2+}$ resin (GE Healthcare, UK) and pre-equilibrated using buffer A (50 mM Hepes pH 7.5, 1 M NaCl, 10% glycerol and 8 M urea) connected to a Next Generation chromatography system (Bio-Rad, USA, USA). The cleared lysate (supernatant) was applied on 5 mL His-Trap FF (GE Healthcare, UK) column at 1mL/min, the column was washed exhaustively with buffer A to remove unbound proteins. The protein was eluted using a 0-100% linear of 500 mM imidazole (Sigma-Aldrich, USA). Elution of the protein was monitored at 280 nm and fractions containing protein were collected and analysed for the presence of Rep protein on a 12% Tris-Glycine SDS-PAGE. Fractions containing ACMV Rep were pooled and concentrated using an ultrafiltration cell with a 10 kDa cut-off membrane. Refolding of the protein to its native conformational structure was achieved by simple buffer dialysis where the urea is slowly removed at 4 ºC using a 4-mL (10-bed volumes) gradient from 8 to 0 M urea refolding buffer (Glynou *et al*., 2003).

2.5.4 ACMV Rep concentration determination

The concentration of His-Tag-Rep was determined using NanoDrop™ 2000/2000c Spectrophotometer (Thermo Fisher Scientific, USA) at 280 nm. With the absorbance and extinction coefficient substituted in the Beer-Lambert law ($c = A/\varepsilon l$), where c is the molar concentration, A is the absorbance at a given wavelength, ε is the molar extinction coefficient ($M^{-1}cm^{1}$) and l is the path length of the cuvette (cm). The extinction coefficient of ACMV Rep is 53 775 $M^{-1}$ $cm^{-1}$. Aliquots of the purified protein were snap-frozen in liquid nitrogen and stored at −80 ºC.

2.5.5 Optimisation purification of ACMV Rep protein

*2.5.5.1 Extraction of Rep protein from inclusion bodies using anionic detergent Sodium Dodecyl Sulfate (SDS)*

Bacterial culture was resuspended in 25 mL ice-cold lysis buffer [25 mM Tris-HCl, pH 8.0, 178 mM NaCl, 10 mM $MgCl_2$, and 1 mM PMSF]. The cells were disrupted by first homogenising and then sonicating for 6 minutes on ice. DNase I (5 µg/mL) was added after sonication and incubated at 4 ºC for 45 minutes to an hour. The lysate was clarified of cell debris by centrifugation at 12 000 ×*g*, 4 ºC for 30 minutes. The pellet with inclusion bodies were resuspended in binding buffer B [25 mM Tris-HCl, pH 8.0, 180 mM NaCl and 1% SDS (w/v)]. The sample in 1% SDS buffer was first homogenised and sonicated for six minutes, six cycles at 80 Hz, then incubated overnight at 24 ºC-25 ºC on a hula mixer (Thermo Fisher Scientific, USA).

*2.5.5.2 Purification of Rep protein with N-laurylsarcosine (Sarkosyl) on immobilised affinity chromatography (IMAC)*

The SDS was removed from the protein sample by precipitation through incubation on cold-ice for 30 minutes and centrifuged twice at 3600 ×*g*. The clarified supernatant was loaded onto a nickel affinity column that was equilibrated with binding buffer [25 mM Tris-HCl, pH 8.0, 180 mM NaCl and 0.1% Sarkosyl (w/v) (Sigma-Aldrich, USA)], flow-throw collected. The 280 nm peak was equilibrated to baseline with the binding buffer. The elution of the protein was done with a linear gradient of 0-500 mM imidazole. Elution of the protein was monitored at 280 nm and fractions containing protein were collected and analysed for the presence of Rep protein on a 12% Tris-Glycine SDS-PAGE.

*2.5.5.3 Removal of residual Sodium Dodecyl Sulfate (SDS) using Ion-exchange chromatography (IEC)*

Fractions of the collected protein from the purification with Sarkosyl (Sigma-Aldrich, USA) were diluted with buffer (25 mM Tris-HCl, 100 mM NaCl, 1 mM DTT and 8 M urea, pH 7.0) to unfold the protein. The solution was then loaded onto an ion-exchange chromatography Q-Sepharose resin (GE Healthcare, UK) at a flow rate of 1 mL/min, where the flow-through was collected in 3 mL fractions. Elution of the protein was monitored at 280 nm on the Next Generation Chromatography (NGC) system (Bio-Rad, USA, USA) and the fractions were analysed for the presence and purity of the Rep protein on a 12% Tris-Glycine SDS-PAGE. The protein was refolded using 10 kDa cut-off snakeskin dialysis tubing (Pierce Biotechnology, USA) into storage buffer (25 mM Tris-HCl, 500mM NaCl, 1 mM DTT) in a stepwise manner.

## 2.6 Structural determination of recombinant Rep proteins

### 2.6.1 Fourier-transform infrared spectroscopy (FTIR)

Secondary structural information of a protein can also be obtained through analysis using Fourier-transform infrared spectroscopy (Perkin-Elmer Frontier MID/FAR IR instrument, USA) with a HATR (Horizontal attenuated total reflectance) crystal plate. Briefly, the background signal was collected and a volume of 2 µL protein in storage buffer solution (25 mM Tris-HCl, 500mM NaCl, 1 mM DTT) was then dropped on the crystal plate. Characterisation was done by measuring transmittance in percentage against wavenumber ($cm^{-1}$). A total of 10 scans were taken for each interferogram with a resolution of 2 $cm^{-1}$ over a range of 400 $cm^{-1}$ to 4000 $cm^{-1}$. An FTIR spectrum of the buffer without the protein was also taken as a reference. The data were analysed with OriginPro 8 software (OriginLab).

### 2.6.2 Intrinsic and extrinsic fluorescence

The tertiary structure of Rep protein was assessed using a Perkin-Elmer LS50B luminescence (Perkin-Elmer, USA) spectrofluorimeter at 20 ºC. All measurements were recorded in triplicate. The tyrosine (Y) and tryptophan (T) residues were excited at 280 nm and the emission spectrum recorded from 280 to 450 nm.

## 2.7 Functional analysis of the purified recombinant Rep proteins

### 2.7.1 Electrophoretic Mobility Shift Assay (EMSA)

#### *2.7.1.1 ACMV Rep-DNA binding assay*

DNA binding activity of the Rep protein was determined by incubating 0.5 µM dsDNA substrate (Table 2.2) mixed with Rep protein in a total volume of 25 µL with DNA binding buffer (25 mM Tris-HCl, pH 7.5, 75 mM NaCl, 5 mM $MgCl_2$, 2.5 mM DTT, and 0.5 mM EDTA) for 1 hour at room temperature on a shaker for the binding reaction. The reaction was stopped by adding 5 µL of 2× loading buffer (30% glycerol, 0.25% bromophenol blue). The samples were electrophoresed on 3% agarose gel in sodium borate buffer (100 mM boric acid and 75 mM NaCl, pH 7.5) and visualised using the ChemiDoc system from Bio-Rad, USA.

#### *2.7.1.2 ACMV Rep-DNA cleavage assay*

Rep protein cleavage activity was determined by incubating Rep (400 nM) in a total volume of 25 µL with 0.5 µM substrate ss-DNA (Table 2.2) for 1 hour at room temperature on a shaker in cleavage buffer (25 mM Tris-HCl, pH 7.5, 75 mM NaCl, 5mM $MgCl_2$, 2.5 mM DTT, 0.5 mM EDTA) (Kittelmann *et al*., 2009). The reaction was stopped by the addition of 5 µL of 2× agarose gel loading buffer (30% glycerol, 0.25% bromophenol blue). Samples were electrophoresed on a 3% agarose in sodium borate buffer [100 mM boric acid (Sigma-Aldrich, USA) and 75 mM NaCl, pH 7.5]. The gel was viewed on the ChemiDoc™ MP gel imager (Bio-Rad, USA, USA) to determine the cleavage activity of ACMV Rep protein. Optimisation of this assay was done by changing one parameter at a time: running buffer (1× Tris-acetic acid-EDTA/ 1× Tris-borate-EDTA), matrix (agarose/ polyacrylamide), matrix concentration (0.8%, 1%, and 3%) and the different voltage (80 V, 150 V and 300 V) used.

Table 2.2: Primers/Substrates used in this study

| Primer | Sequence | Used in |
|---|---|---|
| **Wild type ssDNA oligonucleotide** | 5′GCAATAATATTACCGGATGGCC3′ | DNA Cleavage assay |
| **Wild type dsDNA oligo 1** | 5′TGTTTTTGGTGTCTTGGTGTCCAATATA3′-FITC | DNA binding assay |
| **Wild type dsDNA oligonucleotide 2** | 3′ACAAAAACCACAGAACCACAGGTTATAT5′-FITC | |

2.7.2 Development of an Enzyme-Linked Immunosorbent Assay (ELISA)

*2.7.2.1 Nunc MaxiSorp ™ flat-bottom 96 well plate*

Rep protein was diluted in a coating buffer [100 mM bicarbonate-carbonate (Sigma-Aldrich, USA) pH 9.6] to 0.5 µM and added to a Maxisorp ™ flat-bottom 96 well plates (Thermo Fisher Scientific, USA ) in 100 µL volume each well. Plates were incubated at 4 ºC overnight. The protein was decanted and blocking buffer [Phosphate-buffered saline (PBS), 5% milk powder] was added at 100 µL per well, the plate was incubated at ambient temperature (25 ºC) for 2 hours. The wells were washed once with PBS. The plate was incubated at ambient temperature for 1 hour. Wells were washed with PBS once and dsDNA-FITC diluted in reaction buffer [50 mM HEPES (Sigma-Aldrich, USA) pH 7.4, 5 mM $MgCl_2$, 1 mM DTT, 10 mM NaCl, 0.3% Igepal (Sigma-Aldrich, USA)] to a final concentration of 0.5 µM and added to wells at 100 µL each. The plate was incubated at ambient temperature for 1 hour. After an hour, the plates were washed three times with TBS-Tween and anti-FITC HRP-conjugated antibody (Sigma-Aldrich, USA) was diluted in TBS-tween by 1:2500 and 100 µL was added to each well. The plate was incubated at ambient for 1 hour and washed with TBS-Tween five times. TMB ( 3,3',5,5'-Tetramethylbenzidine) a substrate for horseradish peroxidase (HRP) (Sigma-Aldrich,

USA) was added to wells and the plate was incubated at 37 ºC for 30 minutes and absorbance was determined at 630 nm on a UV-vis spectrophotometer.

## 2.8 Inhibition studies

### 2.8.1 ACMV Rep-DNA binding inhibition assay

DNA binding inhibition activity of compounds [EGCG (Sigma-Aldrich, USA) and CA (Sigma-Aldrich, USA)] against ACMV Rep interacting with ACMV DNA. Briefly, for positive control: dsDNAsubstrate (0.5 µM) was mixed with ACMV Rep (0.4 µM) protein in a total volume of 25 µL with DNA binding buffer (25 mM Tris-HCl, pH 7.5, 75 mM NaCl, 5 mM $MgCl_2$, 2.5 mM DTT, 0.5 mM EDTA) for 1 hour at room temperature on a shaker for the binding reaction. For inhibition reaction: ACMV Rep was incubated with DNA in the presence of an increasing concentration (0.1-100 µM) of the compound. The reaction was stopped by adding 5 µL of 2× loading buffer (30% glycerol, 0.25% bromophenol blue). The samples were electrophoresed on 3% agarose gel in sodium borate buffer (100 mM boric acid and 75 mM NaCl, pH 7.5) and visualised using the ChemiDoc system from Bio-Rad, USA.

### 2.8.2 ACMV Rep-DNA cleavage assay

Inhibition of Rep protein cleavage activity by compounds was determined by incubating Rep (400 nM) in a total volume of 25 µL with 0.5 µM substrate ssDNA for 1 hour at room temperature on a shaker in cleavage buffer [25 mM Tris-HCl, pH 7.5, 75 mM NaCl, 5 mM $MgCl_2$, 2.5 mM DTT, 0.5 mM EDTA] (Kittelmann *et al*., 2009). For inhibition reaction: Rep was incubated with DNA in the presence of an increasing concentration (0.1-100 µM) of the compound. The reaction was stopped by the addition of 5 µL of 2× agarose gel loading buffer (30% glycerol, 0.25% bromophenol blue). Samples were electrophoresed on a 3% agarose in sodium borate buffer [100 mM boric acid (Sigma-Aldrich, USA) and 75 mM NaCl, pH 7.5]. The gel was viewed on the ChemiDoc™ MP gel imager (Bio-Rad, USA, USA) to determine the cleavage activity of ACMV Rep protein. Optimisation of this assay was done by changing one parameter at a time: running buffer (1× Tris-acetic acid-EDTA/ 1× Tris-borate-EDTA), matrix (agarose/ polyacrylamide), matrix concentration (0.8%, 1%, and 3%) and the different voltage (80 V, 150 V and 300 V) used.

# Chapter 3: **Results**

## 3.1 Nucleotide and amino acid sequences coding for Rep protein

For overexpression of the Rep protein, ACMV DNA sequence was obtained from National Centre for Biotechnology Information (https://www.ncbi.nlm.nih.gov/nucleotide/) with an accession number >NC_001467.1:c2756-1680 (see Figure 3.1), the DNA sequence was then used as a template translated to amino acid (aa) sequence (see Figure 3.2) using the NCBI Open Reading Frame (ORF) finder (https://www.ncbi.nlm.nih.gov/orffinder/ ). The ORF 2 was the one coding for ACMV Rep protein with 358 amino acids. From the amino acid sequence, ExPASy was then used to predict the theoretical molecular weight, pI and extinction coefficient of the Rep protein (see Table 3.1).

>NC_001467.1:c27561680

ATGAGAACTCCTCGTTTTCGAATTCAAGCCAAGAATGTCTTTCTCACATACCCAA
AGTGTTCTATACCCAAAGAACACCTACTGTCATTCATTCAAACACTTTCTCTCCA
ATCAAACCCTAAGTTCATTAAAATCTGTAGAGAGCTGCATCAGAATGGGGAACC
TCACTTGCATGCCCTCATTCAATTCGAGGGTAAAATCACGATTACGAACAATCGT
CTCTTCGATTGTGTACACCCAAGCTGTAGCACCAGTTTCCACCCCAACATTCAAG
GTGCCAAATCAAGCTCGGATGTCAAGTCCTATCTGGATAAGGACGGCGACACCG
TCGAATGGGGACAATTTCAAATCGATGGACGATCTGCTAGAGGCGGTCAACAAT
CAGCGAATGATGCTTACGCCAAAGCGCTTAACAGCGGCAGTAAGTCAGAAGCTC
TTAATGTCATTAGGGAATTAGTCCCAAAGGACTTTGTACTTCAGTTTCATAATCTA
AATAGTAATTTAGATAGGATTTTCCAAGAGCCACCAGCTCCTTATGTTTCTCCCTT
CCCATGTTCTTCCTTTGACCAAGTTCCTGTTGAAATTGAAGAATGGGTCGCTGAT
AATGTTAGGGATTCCGCTGCGCGGCCATGGAGACCTAATAGTATTGTCATAGAA
GGTGATAGCAGAACAGGGAAGACGATATGGGCCAGATCTTTAGGCCCACATAAT
TACCTGTGTGGACACCTGGACCTTAGTCCAAAGGTCTTCAATAATGCTGCCTGGT
ACAACGTCATTGATGACGTCGATCCCCACTACCTAAAGCACTTTAAGGAATTCAT
GGGGTCCCAGAGGGACTGGCAGTCCAACACGAAATACGGGAAACCCGTTCAAAT
TAAAGGTGGAATTCCCACTATCTTCCTCTGCAATCCAGGACCTACCTCGTCCTAT
AAAGAGTTCCTAGCCGAGGAAAAGCAAGAAGCGCTAAAGGCCTGGGCATTAAA
GAATGCAATCTTCATCACCCTCACAGAACCACTCTACTCAGGTTCCAATCAAAGT
CACTCACAGACAAGTCAAGAAGCGAGCCATCCGGCGTAG

Figure 3.1: African cassava mosaic virus isolate West Kenyan 844 segment DNA1.

≥VIRT4928
MRTPRFRIQAKNVFLTYPKCSIPKEHLLSFIQTLSLQSNPKFIKICRELHQNGEPHLHAL
IQFEGKITITNNRLFDCVHPSCSTSFHPNIQGAKSSSDVKSYLDKDGDTVEWGQFQID
GRSARGGQQSANDAYAKALNSGSKSEALNVIRELVPKDFVLQFHNLNSNLDRIFQEP
PAPYVSPFPCSSFDQVPVEIEEWVADNVRDSAARPWRPNSIVIEGDSRTGKTIWARSL
GPHNYLCGHLDLSPKVFNNAAWYNVIDDVDPHYLKHFKEFMGSQRDWQSNTKYG
KPVQIKGGIPTIFLCNPGPTSSYKEFLAEEKQEALKAWALKNAIFITLTEPLYSGSNQS
HSQTSQEASHPA

Figure 3.2: ACMV 358 amino acid sequence.

Table 3.1: Theoretical characterisation of Rep protein using ExPASy bioinformatics tool.

| | |
|---|---|
| **Theoretical Molecular weight (kDa)** | **40** |
| **Theoretical pI** | 6.77 |
| **Theoretical Extinction Coefficient ($M^{-1}$ $cm^{-1}$)** | 53775 |
| **Number of Tyrosine residues (Y)** | 10 |
| **Number of Tryptophan residues (W)** | 7 |
| **Number of Phenylalanine residues (F)** | 19 |

### 3.2 Bacterial transformation

The transformation of BL21 (DE3) pLysS *E. coli* competent cells with a plasmid carrying a gene coding for ACMV His-tagged Rep (GenScript) by a heat shock method was successful. These cells were plated on sterilised agar media containing appropriate antibiotics (100 µg/mL ampicillin and 35 µg/chloramphenicol). The cells were incubated overnight at 37 ºC to assess the success of the transformation procedure (see Figure 3.3). The uptake of the plasmid DNA by the bacterial *E. coli* cells was successful. Bacterial stock solution was successfully produced after picking a single colony and growing it overnight.

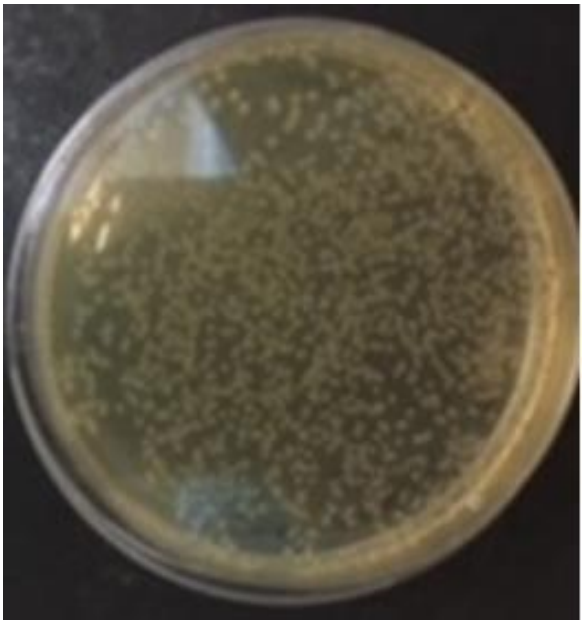


Figure 3.3: Agar plate showing pET-15b-ACMV-Rep transformed colonies.

Bacterial hosts BL21 (DE3) pLysS *E. coli* competent cells from New England Biolabs were used. The (1:100) diluted cells were plated on sterilised agar media with appropriate antibiotics, ampicillin (100 µg/mL) and chloramphenicol 35 µg/mL) and incubated overnight at 37 ºC.

## 3.3 Verification of the ACMV Rep plasmid

### 3.3.1 Enzyme restriction digestion

Restriction digestion was done to confirm the uptake of plasmid DNA with ACMV Rep gene insert. *NdeI* and *BamHI* restriction enzymes that digets at *NdeI* and *BamHI* restriction sites were used. The enzymes used digest maximally at 37 ºC in a water bath for 1 hour and the products were analysed on 1% agarose gel electrophoresis. Figure 3.4 shows the successful restriction digestion of plasmid DNA after incubation. Lane 1 shows undigested plasmid DNA while Lane 2 shows single enzyme digestion with a band approximately above 5700 bp and Lane 3 shows two distinctive bands with sizes that are expected (5700 bp and 1155 bp). This figure indicates the success of the enzyme restriction digestion.

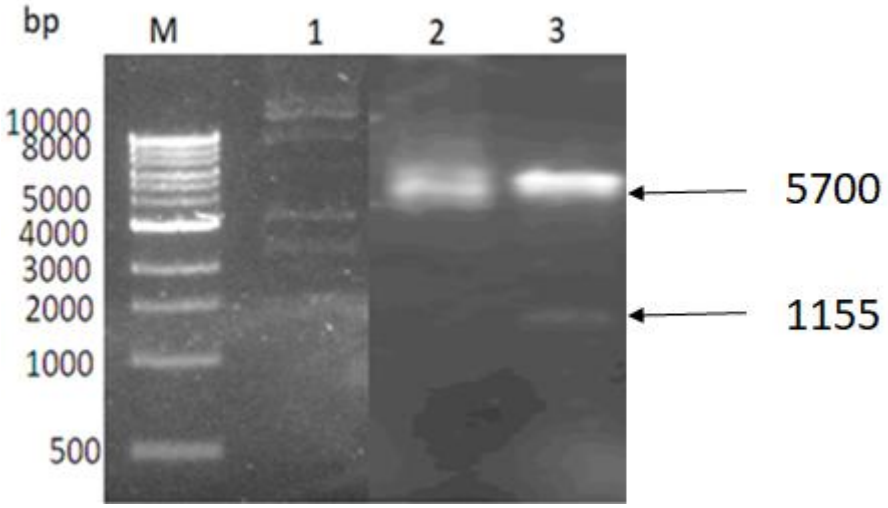



Figure 3.4: A 1% agarose gel showing the single and double enzyme digestion of plasmid from transformed BL21 (DE3) pLysS.

Lane M: 1 kb Molecular DNA marker (Fermentas, USA), Lane 1: Undigested plasmid, Lane 2: Single digest of plasmid DNA with *BamHI* restriction enzyme and Lane 3: Double digest of plasmid DNA with *BamHI* and *NdeI* restriction enzymes.

### 3.3.2 DNA sequencing

To confirm the identity of target gene insert on the plasmid, DNA sequencing was performed at Inqaba Biotec. The purified DNA sample was analysed using the T7 promoter and T7 terminator universal primers (Inqaba Biotec, South Africa) and the sequencing results obtained showed 92.47% identity to Rep sequence from the NCBI depository as a result of codon optimisation applied during the synthesis of Rep gene by Genscript. Results for both the T7 forward and reverse primers were analysed but Figure 3.5 only shows the T7 forward primer results. The arrow depicts a portion of the Rep protein from the start codon (ATG).

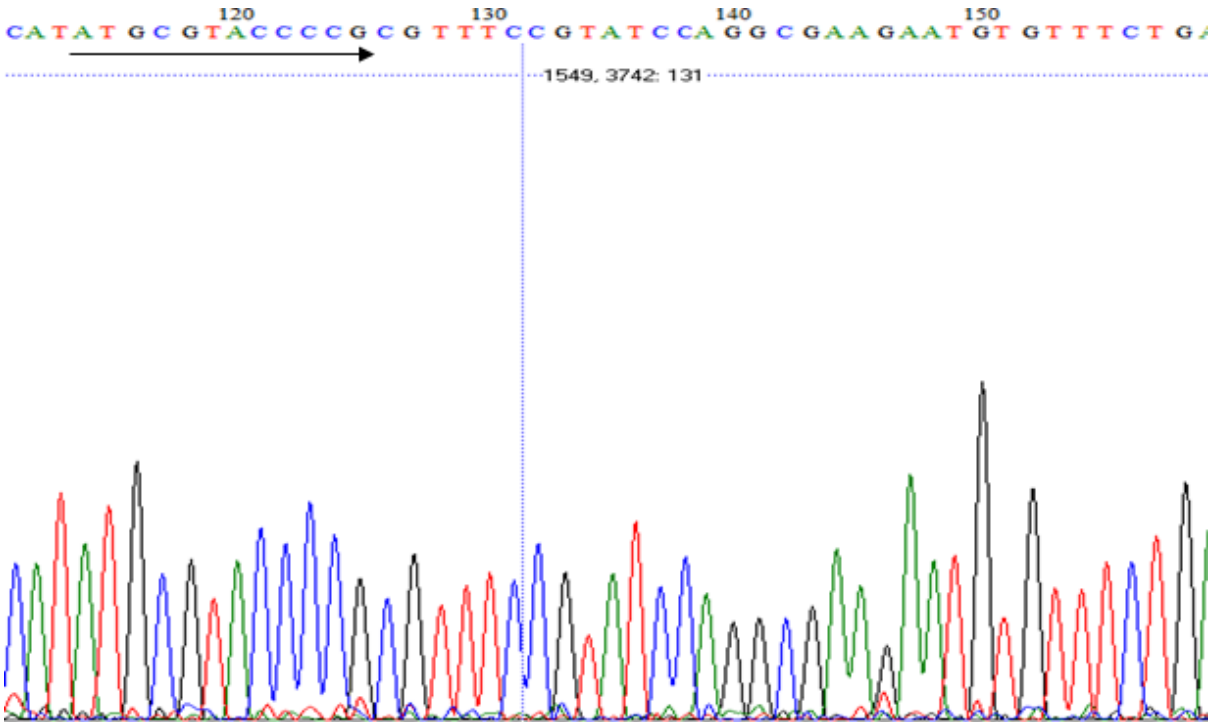



Figure 3.5: DNA sequencing chromatogram results from Inqaba Biotec analysed using BioEdit tool.

Coloured peaks correspond to each unique nucleotide picked, blue: Cytosine (C), black: Guanidine (G), green: Adenine (A) and Red: Thymine (T). Only the T7 forward primer sequencing results are presented in this figure with the arrow showing where the start codon coding for Methionine (M).

## 3.4 Overexpression of ACMV Rep protein

### 3.4.1 Bacterial growth curve

To determine the optimal time to start induction of protein expression by the bacterial cells. The growth of bacteria (see Figure 3.6) in growth media supplemented with 5% glucose was studied. The maximum time the bacteria grew to an optical density of 0.6-0.8 was 2 hours.

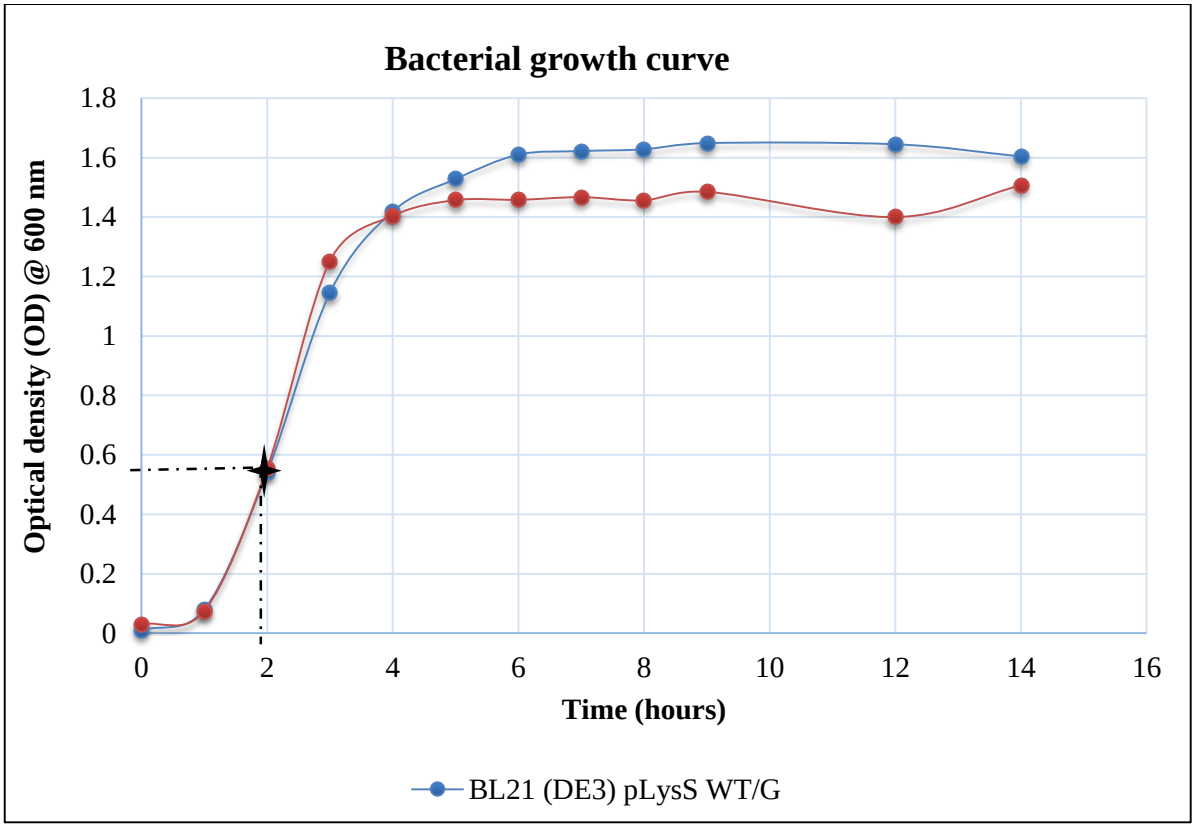



Figure 3.6: *E. coli* BL21 (DE3) pLysS growth curve against time.

The growth of bacteria is determined by the optical density in the growth media at an absorbance of 600 nm. Blue line [BL21 (DE3) pLysS without glucose] represents bacterial growth in media not supplemented with 5% glucose and the red line [BL21 (DE3) pLysS with glucose. Dotted line intercepting at a point (asterisk) is extrapolating the time used by the bacteria to reach an optical density of 0.6 which is 2 hours.

## 3.5 Confirmation of protein induction by Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis (SDS-PAGE) and Western blot

### 3.5.1 Overexpression of target ACMV Rep protein and size determination standard curve

Expression of ACMV recombinant Rep protein by BL21 (DE3) pLysS *E. coli* cells was detected after 5 hours post-induction with 1 mM IPTG, as determined by SDS-PAGE (see Figure 3.7 A). Rep protein with an expected size of 40 kDa was overexpressed in the induced bacterial sample as seen in Figure 3.7 A, Lane 1 compared to the un-induced bacterial sample in Figure 3.7, Lane 2. A molecular marker in kilodalton was electrophoresed alongside with the bacterial samples and used for size determination (see Figure 3.7, Lane M). The migration distance of the molecular marker was used to construct a standard curve (see Figure 3.7 B).

The size of the overexpressed protein was determined by using the migration distance of overexpressed Rep protein on the equation y= -0.031x + 2.6114 of the standard curve. It was then calculated to be 40 kDa (see Figure 3.7 B).

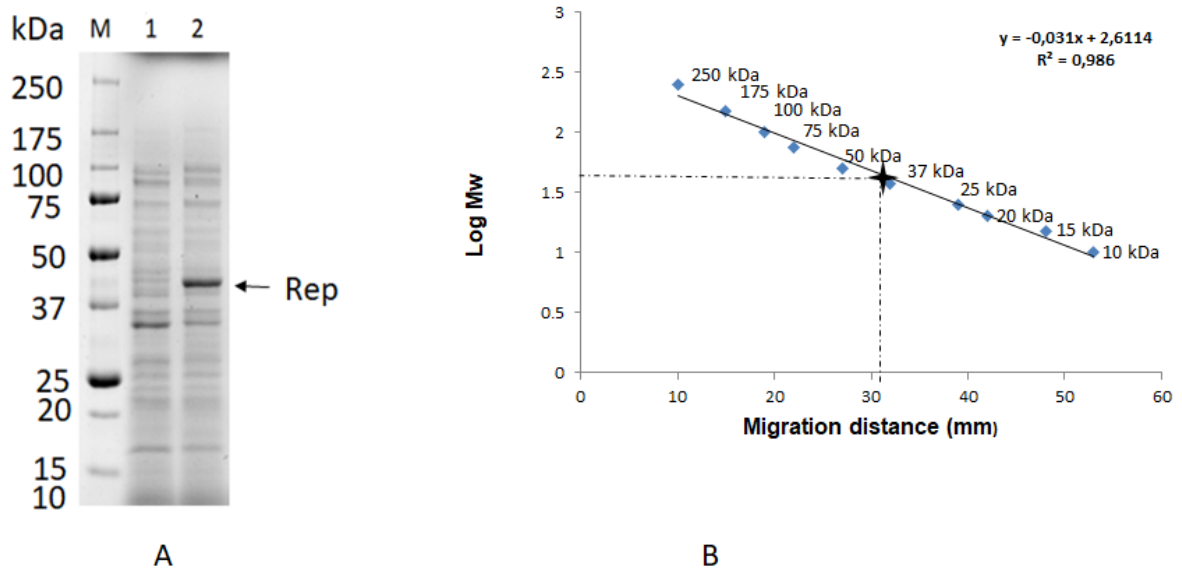



Figure 3.7: Represents a 12% SDS-PAGE analysis of Rep overexpression in BL21 (DE3) pLysS *E. coli* cells and molecular weight standard curve.

Lane M: Molecular marker protein in kilodalton (Precision Plus Protein TM Standard), Lane 1: Un-induced bacteria cells and Lane 2: Induced bacteria cells. Arrow indicates the 40 kDa overexpressed recombinant Rep protein (A). (B) Represents a standard curve relating relative mobility to the log of molecular weight markers (kDa) used in 12.5% reducing SDS-PAGE gels. The equation of the trend line is given by y = -0.031x +2.6114, with a correlation coefficient of 0.986.

### 3.5.2 Western Blot

To confirm the overexpression and identity of His-tagged Rep recombinant protein, Western Blot was performed by transferring proteins from SDS-PAGE to a PDVF Western Blot membrane using 1:2000 anti-His mouse monoclonal antibody and 1:10000 Goat Anti-Mouse HRP (IgG) antibody to detect target protein. It is indicated in Figure 3.8 showing the overexpression of Rep protein on an SDS-PAGE and a corresponding Western Blot picture that indeed the overexpressed protein indicated with an arrow in Figure 3.8 A is the detected single band on Figure 3.8 B, also indicated with an arrow.

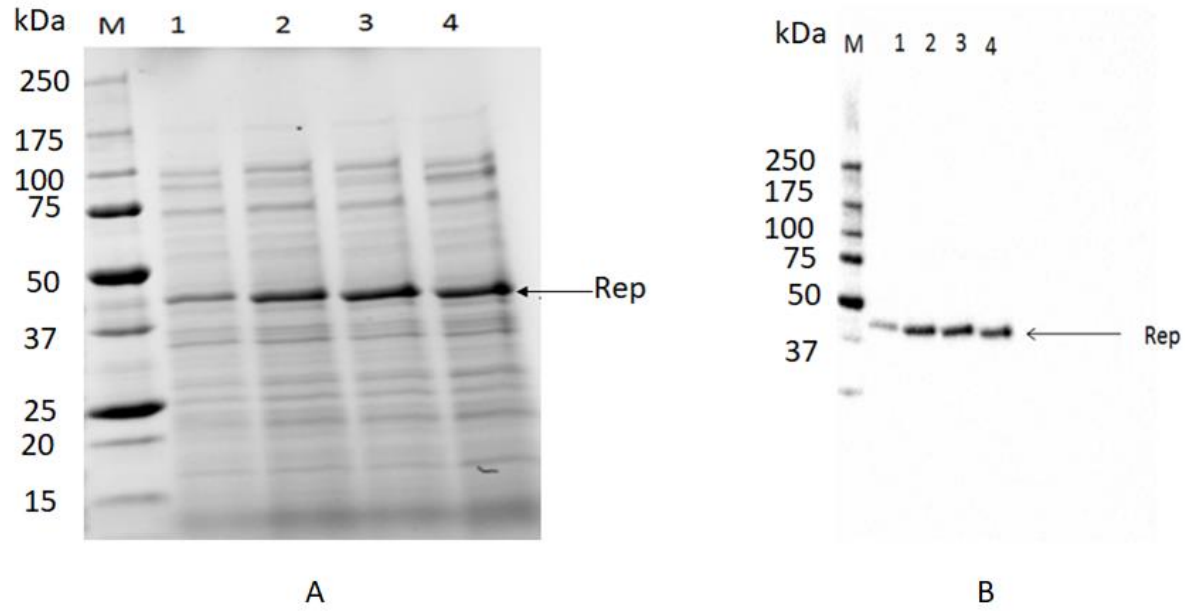



Figure 3.8: A 12% SDS-PAGE showing Rep protein expression induced with 1 mM IPTG in BL21 (DE3) pLysS *E. coli* cells (A) and the corresponding western blot (B).

Rep protein expression induced with 1 mM IPTG and Rep protein analysed on western Blot using 1:2000 anti-His Mouse monoclonal antibody and 1:10000 Goat Anti-Mouse HRP (IgG) antibody (B). Lane M: Molecular marker in kilodaltons, Lane 1: After 1.5 hour induction, Lane 2-4: After 3 hours induction.

### 3.6 Optimisation of ACMV Rep protein Overexpression

To make sure that optimal conditions for high protein production were applied, optimisation of overexpression conditions was explored. The use of different media (Terrific, 2 × YT and LB) as a source of nutrients for the growth of bacteria was compared. The Terrific broth was found to be an optimal growth medium as compared to LB and 2 × YT broth (see Figure 3.9 A). The other condition explored was expression time at an interval of 30 minutes and results showed that expression levels of Rep protein increased over time. Overexpressed Rep protein was detected as early as 30 minutes post-induction but the optimal time for overexpression was three hours post-induction (see Figure 3.9 B, Lane 7). The last two parameters analysed were the use of different IPTG concentrations (0.1 mM, 0.5 mM, and 1 mM) to induce overexpression of the Rep protein. Optimal overexpression was indicated from 0.5 mM-1 mM IPTG (see Figure 3.10 A). Then lastly, different glucose concentrations (1%, 2%, 3%, 4%, and 5%) as a nutrient supplement in LB broth were analysed on SDS-PAGE and resulted in an increased overexpression of recombinant Rep protein as compared to the no glucose sample

(see Figure 3.10 B). As compared to 0% glucose (see Figure 3.10 A) an increase in the percentage concentration of glucose in media has an incremental impact on the overexpression of recombinant Rep protein (see Figure 3.10 B). Lower concentrations of glucose (0.1%, 0.2%, 0.3%, 0.4%, and 0.5%) were also tested and they did not have any effect on the overexpression as compared to higher concentrations in Figure 3.10 B (Data not shown).

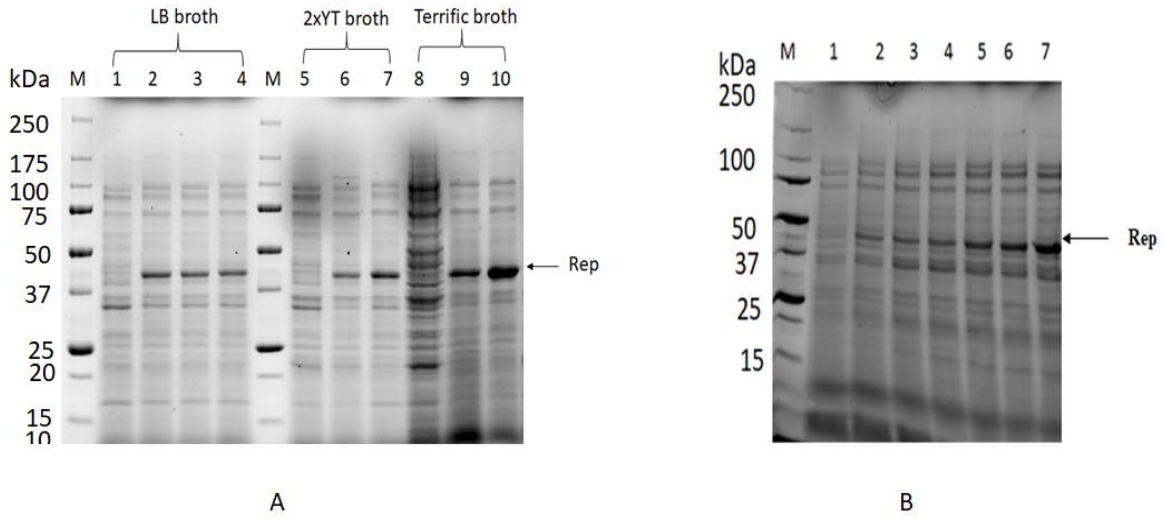



Figure 3.9: A 12% SDS-PAGE showing overexpression of the recombinant ACMV Rep protein.

(A) Represents overexpression in different rich media (LB broth, 2×YT broth, and terrific broth). Lane M: Molecular maker in kilodalton, Lane 1: Uninduced bacteria in LB broth, Lane 2, 3 and 4: Induced bacterias in LB media, Lane 5: Uninduced bacteria in 2×YT media, Lane 6 and 7: Induced bacteria in 2×YT media and Lane 8: Uninduced bacteria in Terrific media, Lane 9, 10: Induced bacteria in Terrific broth. Arrow indicates overexpressed Rep protein at the size of 40 kDa. Terrific broth as a nutrient source produces high yields of recombinant Rep protein compared to 2× YT and LB media under the same growth condition. (B) Represents Rep protein expression studies in BL21 (DE3) pLysS *E. coli* cells at different times, a 30 minutes interval after induction with 1 mM IPTG. Lane M: molecular marker, Lane 1: Uninduced, Lane 2: induction after 30 minutes, Lane 3: induction after 1 hour, Lane 4: induction after 1.5 hours, Lane 5: induction after 2 hours, Lane 6: induction after 2.5 hours and Lane 7: induction after 3 hours. Expression levels of Rep protein increased over time.

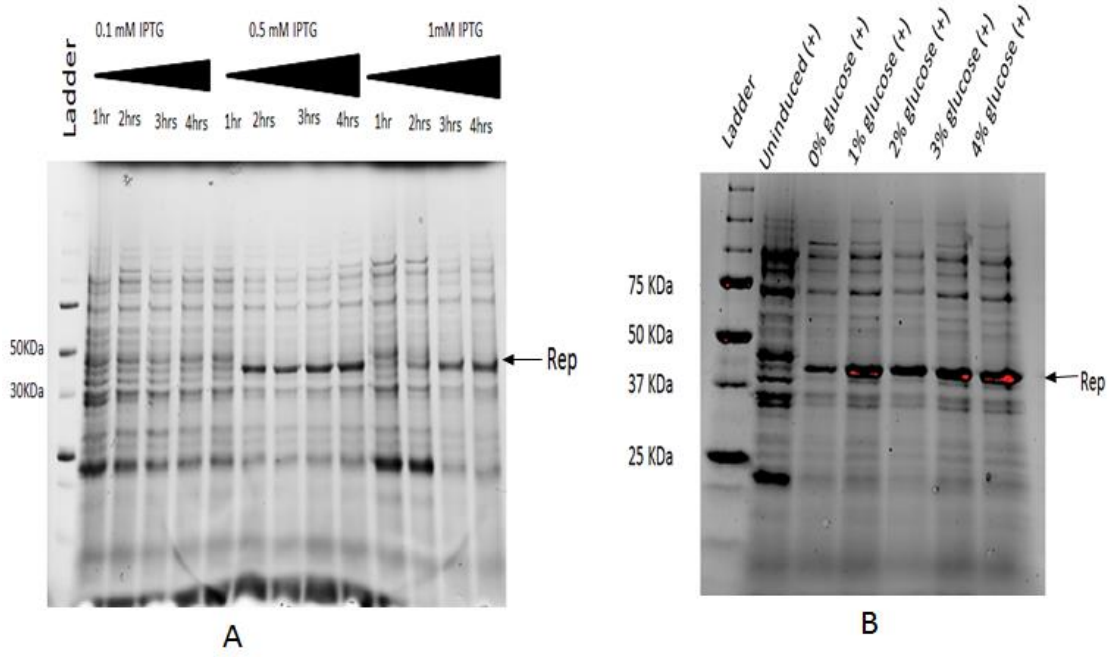



Figure 3.10: A 12% SDS-PAGE showing Rep protein expression studies in BL21 (DE3) pLysS *E. coli* cells.

Ladder: Molecular marker in kilodalton. The arrow indicates overexpressed Rep protein with a size of 40 kDa. (A) Represents the effect of IPTG (0.1 mM 0.5 mM and 1 mM) on overexpression of Rep protein. (B) Represents glucose studies, BL21 (DE3) pLysS *E. coli* cells were allowed to grow in LB medium supplemented with glucose (1%-4%) until an optical density of 0.6 is reached. Then a four hours incubation post-induction with 1 mM IPTG. Overexpression of recombinant Rep protein increases with an increase in percentage concentration of glucose.

## 3.7 Purification, refolding and concentration determination of ACMV recombinant Rep protein

### 3.7.1 Solubility studies of the Rep protein

Bacterial cultures were harvested and processed further by rupturing the cell walls using lysis buffer and sonication. The solubility of the protein was then assessed by analysing the pellet and supernatant fractions on SDS-PAGE after centrifugation (see Figure 3.11). The supernatant showed very low protein concentrations indicated by a band size (see Figure 3.11, Lane 3) and all the protein was located in the lysed pellet fraction (see Figure 3.11, Lane 2). High yields of protein detected in the lysed pellet fraction on SDS-PAGE indicated that the protein was

insoluble (see Figure 3.11). Therefore, attempts were made to optimise the solubility of Rep protein in bacterial cells during the overexpression but no improvements were observed. The insoluble fraction became the focus of subsequent purification procedure for Rep protein.

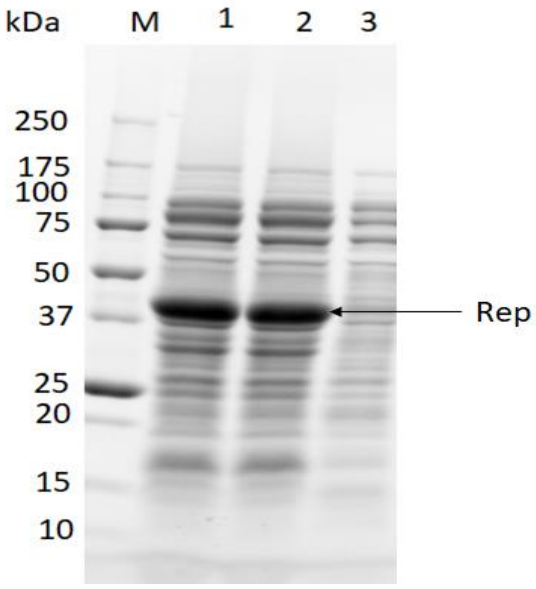



Figure 3.11: A 12 % SDS-PAGE analysis of fractions from Rep protein solubility studies.

Lane M: Molecular marker in kilodaltons, Lane 1: Bacterial lysate, Lane 2: Lysate pellet, Lane 4: Lysate supernatant. Arrow indicates Rep protein at the molecular weight of 40 kDa in the lysate, insoluble and soluble fractions. As compared to the starting lysate (Lane 1) centrifuged, the final insoluble fraction had a band of the same size (Lane 2) and a smaller band in the soluble fraction (Lane 3) suggesting that the protein is in inclusion bodies.

### 3.7.2 Purification of the ACMV recombinant Rep protein

The protein had to be solubilised using 8 M urea to extract it from inclusion bodies. When loaded on an affinity column, the protein did not bind strongly to the His-FF Nickel affinity column as it was observed on SDS-PAGE showing fractions eluted earlier with unwanted bacterial proteins (see Figure 3.12 and Figure 3.13, Lane F4). A protein concentration of $\geq$ 65% purity was purified (see Figure 3.13). This protein was only enough for activity studies but not enough or of good quality for structural studies. The protein was further purified using ion-exchange chromatography columns (HiTrap SP Sepharose HP and HiTrap Q Sepharose HP) but this too did not make any improvements (Data not shown). The purified protein was then refolded by stepwise dialysis and concentrated.

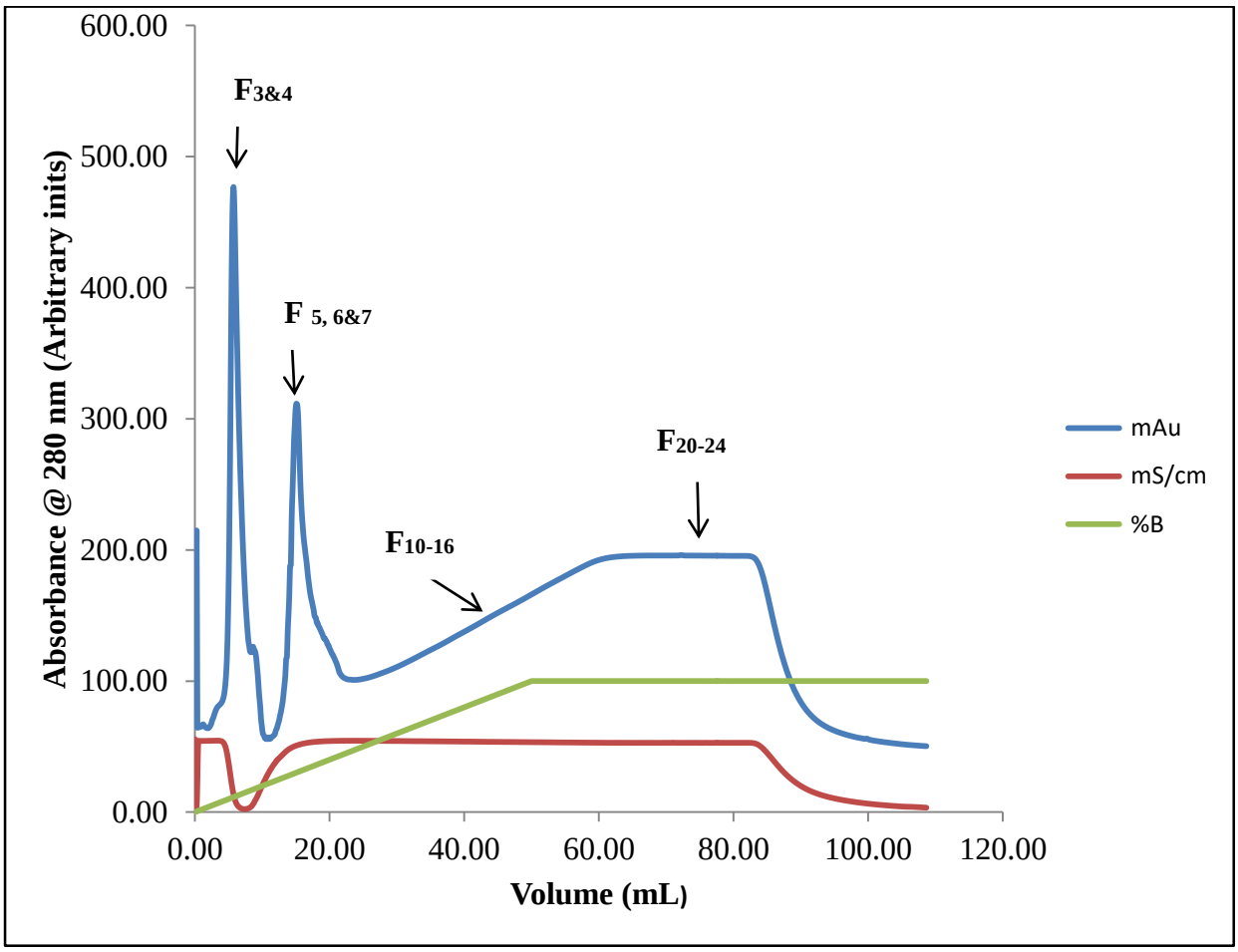



Figure 3.12: Nickel column elution profile of a sample containing recombinant Rep protein.

Conditions: Sample was in binding buffer (50 mM HEPES, pH 8, 1 M NaCl, 10 % glycerol and), flow rate (1 mL/mim), pressure (0.3 mPa), Elution buffer ((50 mM HEPES, pH 7.5, 1 M NaCl, 10 % glycerol and 500 mM imidazole), elution gradient (50 mL/100 % B) and 3 mL fraction size. Arrows indicates peaks at which different fractions were collected. Rep protein was eluted from the second peak (Fraction 5, 6 and 7).

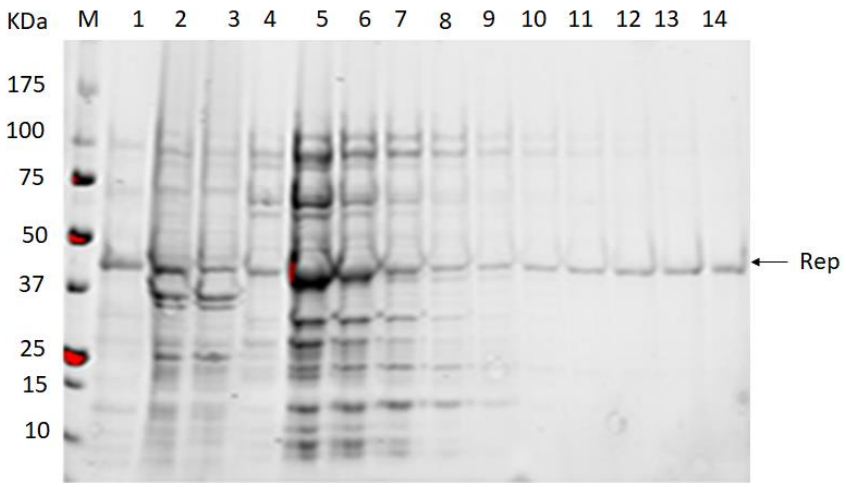



Figure 3.13: A 12 % SDS-PAGE analysis of protein purification on a Ni2+ affinity column.

Lane M: Molecular marker in kilodaltons, Lane 1: Purified Rep protein fraction, Lanes 2: Supernatant fraction from lysis lysate. Lane 3: Protein flow through, Lane F4-14: Eluted protein fraction with a 500 mM gradient imidazole. Arrow indicates the 40 kDa Rep protein. Rep protein was eluted early in fractions 3 and 4 with contaminating bacterial proteins, the protein appears purified from most contaminants later in fractions 10- 14 at an imidazole concentration of 132.5 mM.

### 3.7.3 Purification of ACMV recombinant Rep protein solubilised with SDS

Attempts were also made to totally expose the histidine tag on the protein by using SDS and Sarkosyl as unfolding denaturants and loading it on the His Trap HP Ni $^{2+}$ affinity column (see Figure 3.14). Purification using SDS and Sarkosyl made significant improvements with regards to the purity of the protein. The protein eluted was as pure as $\geq$ 95% and detected on SDS-PAGE (see Figure 3.15). However, it is unfolded/denatured and refolded to its native structure conformation. Western blot detection of the tagged Rep protein using an anti-His monoclonal antibody was performed to confirm the identity of the purified protein (see Figure 3.16). To remove residual SDS from the protein, a HiTrap Q Sepharose column was used to bind the negatively charged SDS and let the SDS-free protein elute in 8 M urea elution buffer (see Figure 3.17).

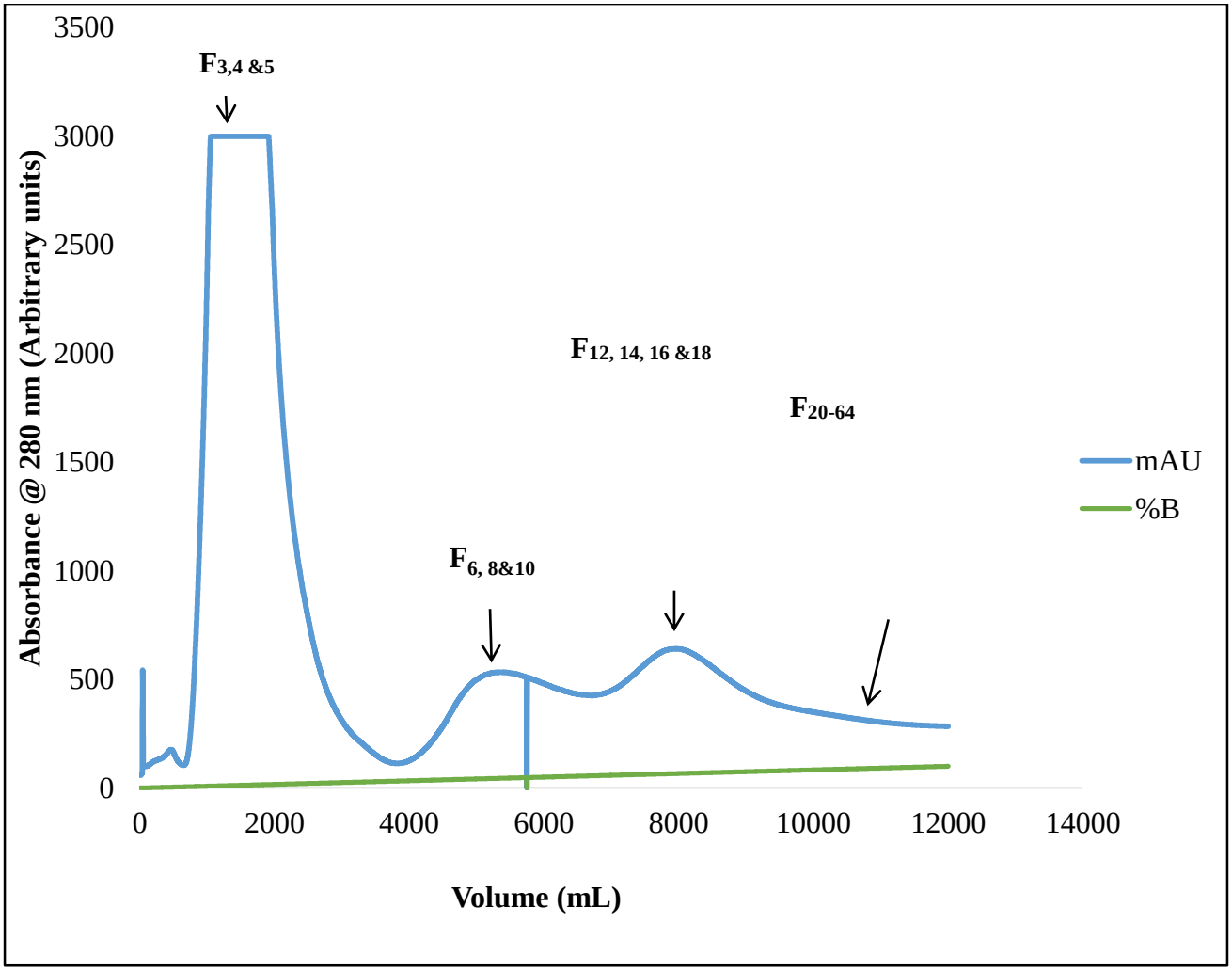



Figure 3.14: Rep protein purification profile from 1 L bacterial culture on a His Trap HP Ni $^{2+}$ affinity column.

Elution was achieved using a gradient of 500 mM imidazole in binding buffer (25 mM Tris-HCl, pH 8.00, 180 mM NaCl, 0.1% Sarkosyl and 1 mM DTT). Rep protein was eluted early in fractions 3 and 4 with contaminating bacterial proteins. The protein appears purified later from fractions 6- 64 at an imidazole concentration of 10 mM.

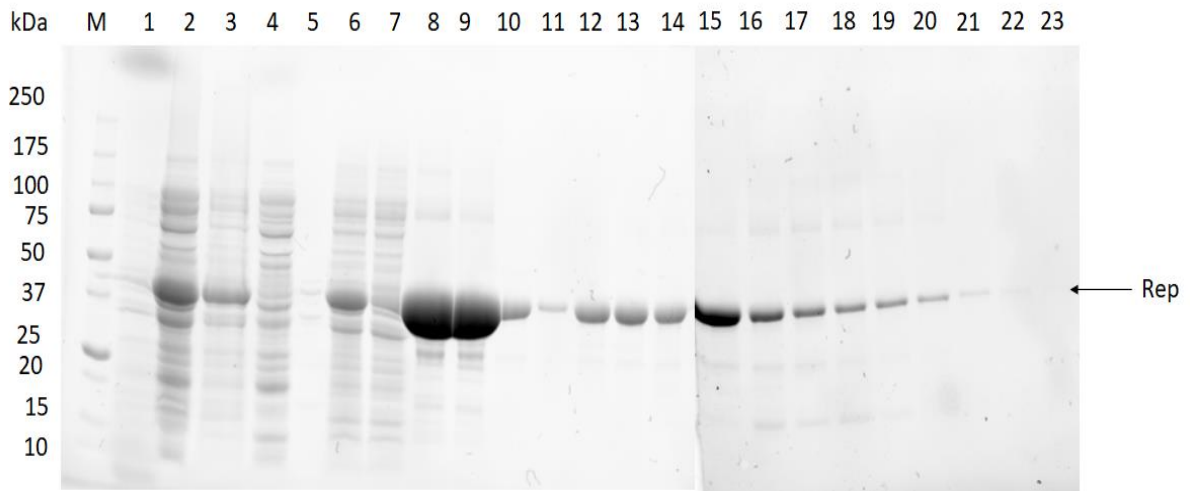



Figure 3.15: A 12% SDS-PAGE analysis of fractions from Rep protein purification on the $Ni^{2+}$ affinity column.

Lane M: Molecular marker, Lane 1: Un-induced bacterial cells, Lane 2: Bacterial lysate, Lane 3: Lysate pellet, Lane 4: Lysate supernatant, Lane 5: Pellet in 1% SDS, Lane 6: Sample loaded, Lane 7: Flow-through, Lane 8-23: Fraction 3, 4, 5, 6, 8, 10, 12, 14, 16,18, 20, 24, 26, 28, 30, 40, 50, 46, 60, 62 and 64.

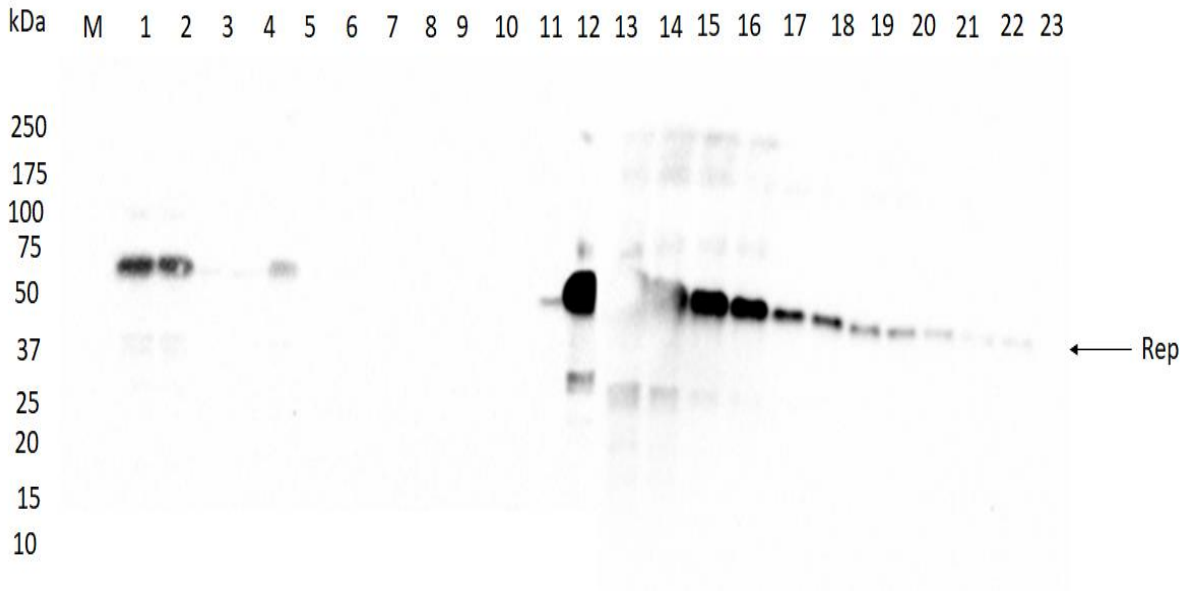



Figure 3.16: Western blot analysis of the Rep protein using 1:2000 anti-His Mouse monoclonal antibody and 1:10000 Goat Anti-Mouse HRP (IgG) antibody.

Lane M: Molecular marker, Lane 1: Un-induced bacterial cells, Lane 2: Bacterial lysate, Lane 3: Lysate pellet, Lane 4: Lysate supernatant, Lane 5: Pellet in 1% SDS, Lane 6: Sample loaded, Lane7: Flow through, Lane8-23: Fraction 3, 4, 5, 6, 8, 10, 12, 14, 16, 18, 20, 24, 26, 28, 30, 40, 50, 46, 60, 62 and 64.

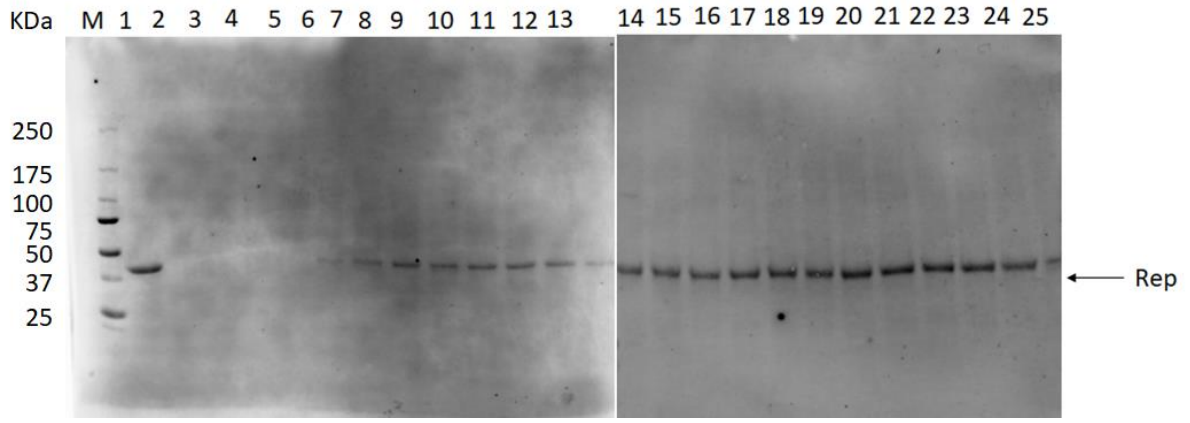



Figure 3.17: A 12 % SDS-PAGE analysis of fractions from Rep protein purification on the Q Sepharose fast flow.

Lane M: Molecular marker in kilodalton, Lane 1: Sample from $Ni^{2+}$ affinity column, Lane 2: Flow-through collected in 3 mL fractions.

### 3.8 Protein determination

The concentration of Rep protein was calculated using the protein absorbance at 280 nm from NanoDrop™ 2000/2000c Spectrophotometers (see Figure 3.18).

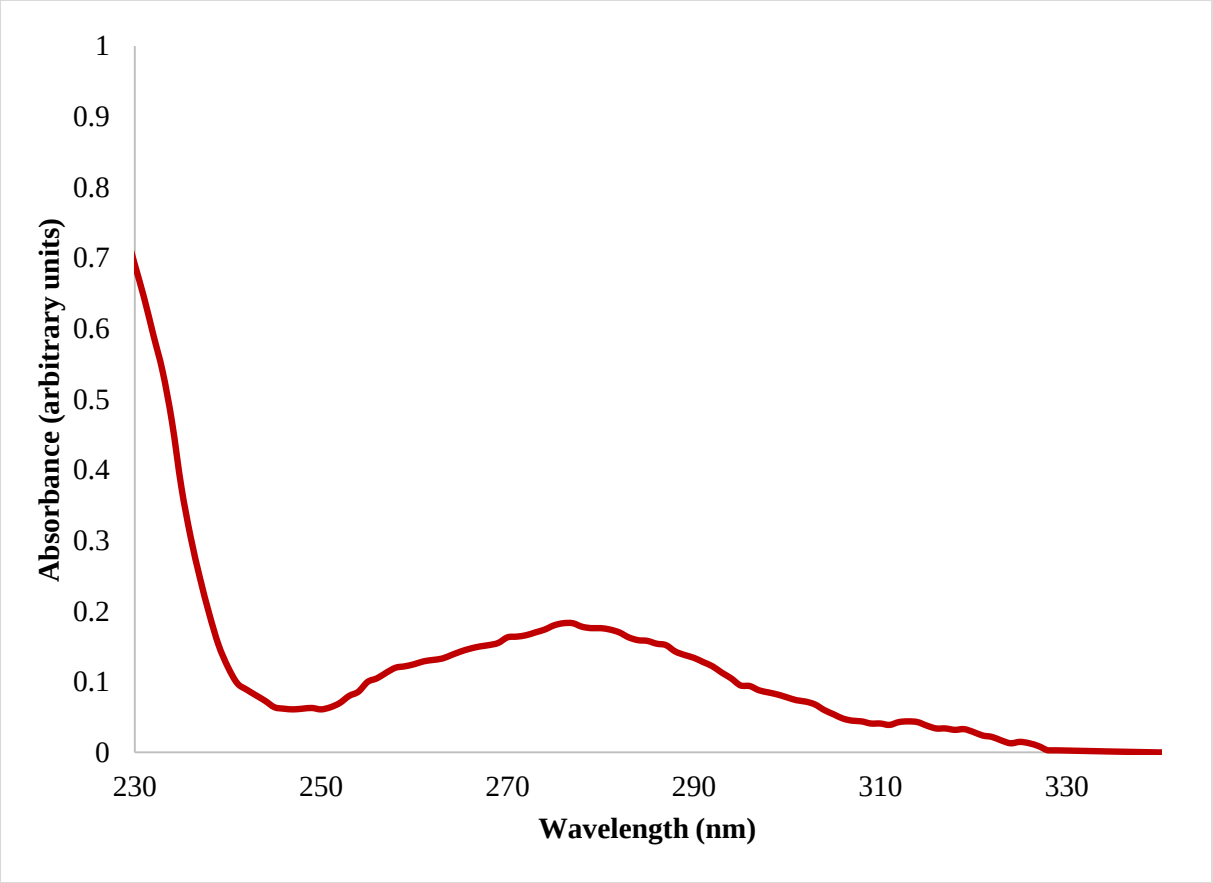



Figure 3.18: Absorbance spectrum of Rep protein at 280 nm using the NanoDrop™ 2000/2000c Spectrophotometer.

### 3.9 Structural determination of recombinant Rep proteins

3.9.1 Secondary structure prediction

The secondary structure composition was predicted using an online program, CFSSP (http://www.biogem.org/tool/chou-fasman/index.php), found on ExPASy. A fasta format amino acid sequence was used as a query template and the output is presented in the figure below (see Figure 3.19). Online secondary structure prediction results in Figure 3.19 show that Rep protein is composed of 59.8% of α-helix (Red), 29. 3% of extended strands (which makes up β-sheet in blue) and 15.6% of turns (Green). According to the report, coils were not calculated as a percentage but the graphic view of the prediction shows that Rep protein contains coils (Yellow).

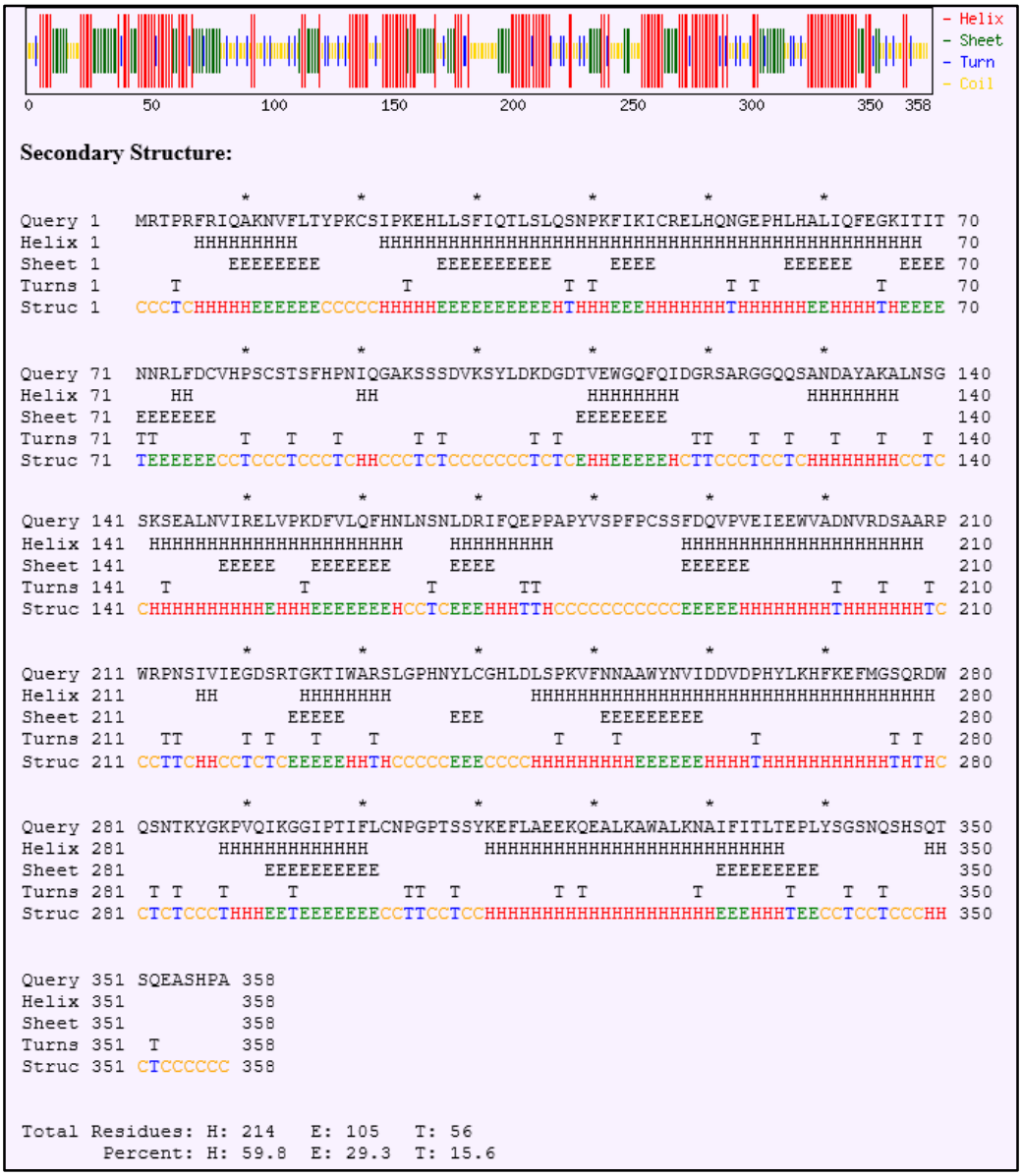



Figure 3.19: ACMV Rep protein secondary structure prediction analysis using the online prediction tool CFSSP.

This server uses the Chou and Fasman algorithm to predict the secondary structure of ACMV Rep protein based on analyses of relative frequencies of each amino acid in beta sheets, turns and alpha helices in comparison with known protein structures resolved with X-ray crystallography. The Rep protein is shown to have helices (H) depicted in red, sheets (E) depicted with green, turns (T) depicted with blue and coils (C) depicted in yellow. This also shows the percentage of each composition except for coils, where the percentage of helices is 59.8%, 29.3% for sheets and lastly for β turns is 15.6 %.

3.9.2 Fourier-transform infrared (FTIR)

The secondary structure was characterised using Fourier-transform infrared. This was done as percentage transmittance (see Figure 3.20) and then converted into absorbance using Beer's law (A= 2- log T %)) (see Figure 3.21 ). The Fourier-transform IR detected only the Amide I region (1600 cm $^{-1}$ - 1700 cm $^{-1}$). This region cannot reveal the secondary composition of a protein because of the overlapping individual peaks. The peaks were identified and described according to their as compared to known frequencies of peaks from proteins (see Table 2.1). The overlapping Amide I region (see Figure 3.22) was then self-deconvoluted after baseline correction under the Gaussian model (see Figure 3.23).

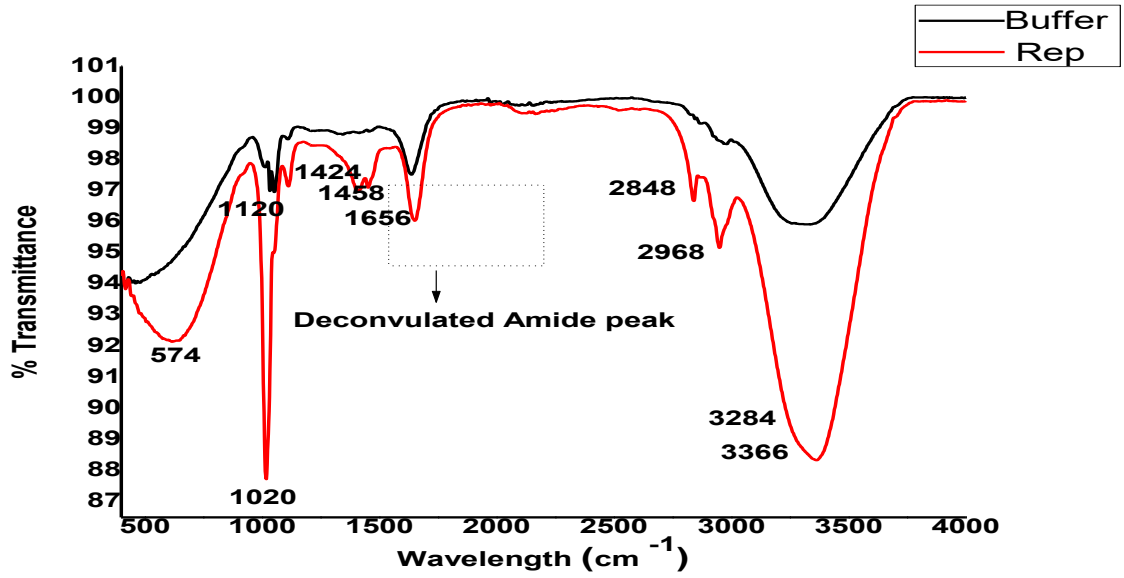



Figure 3.20: An original FTIR spectra of both sodium phosphate buffer (Black) and Rep protein in the buffer (Red) in the 4000–1000 cm$^{-1}$ regions.

The data was measured as percentage transmittance. The graph shows differ rent peaks indicating different functional groups in both the protein and buffer. The peaks at different wavelengths are described in Table 3.2.

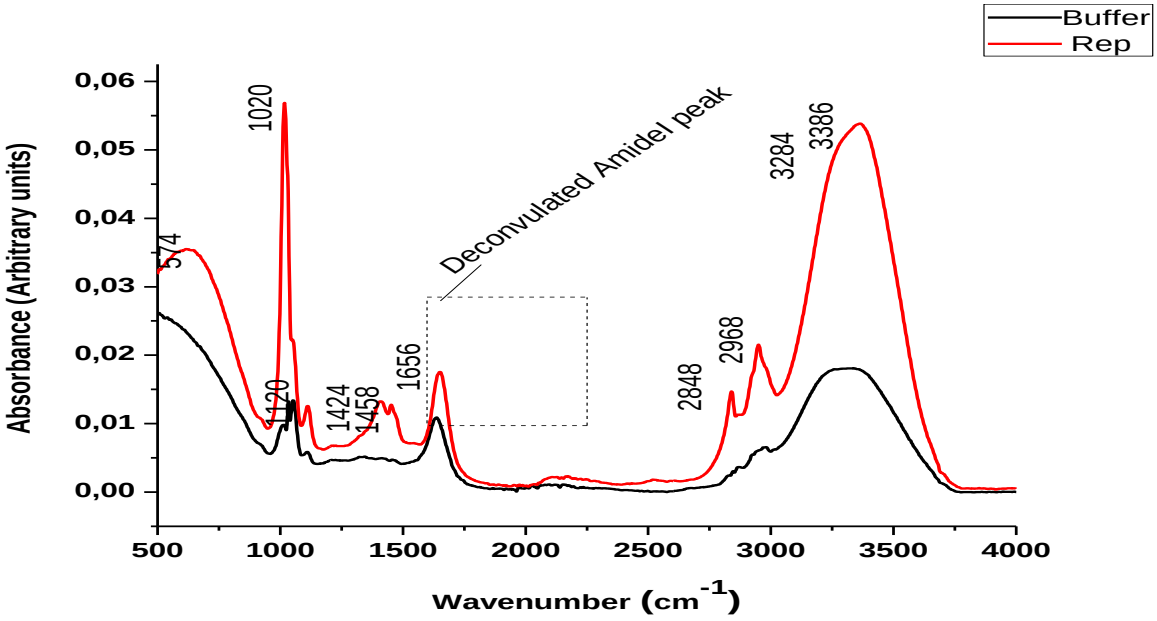



Figure 3.21: An original FTIR spectra of both sodium phosphate buffer and Rep protein in buffer in the 4000–1000 $cm^{-1}$ region.

The data was translated from percentage transmittance to absorbance. The different peaks indicate different functional groups, peak at 574 $cm^{-1}$: -C≡C-H: C-H bend, peak at 1020 $cm^{-1}$: C-O stretch, 1120 $cm^{-1}$: C-N stretch (in-ring), 1424 $cm^{-1}$: C-C stretch, 1656 $cm^{-1}$, 2848 $cm^{-1}$, 2968 $cm^{-1}$: C-H stretch, 3284 $cm^{-1}$: -C≡C-H: C-H stretch, 3386 $cm^{-1}$: O-H stretch, H bonded. The marked deconvoluted peak is one that depicts a secondary structure of the Rep protein.

Table 3.2: FTIR analysis of ACMV recombinant Rep protein

| Approximate frequencies ($cm^{-1}$) | Description | Functional groups | Rep protein frequencies ($cm^{-1}$) |
|---|---|---|---|
| 3500-3200 | O-H stretch, H bonded | Alcohols, phenols | 3386 |
| 3330-3270 | -C≡C-H: C-H stretch | Alkynes (Terminal) | 3284 |
| 3000-2850 | C-H stretch | Alkanes | 2968 |
| | | | 2848 |
| | | | 1656 |
| 1470-1450 | C-H bend | Alkanes | 1458 |
| 1500-1400 | C-C stretch | Aromatics | 1424 |
| 1120 | C-N stretch (in-ring) | Aliphatic amines | 1120 |
| 1320-1000 | C-O stretch | Alcohols, carboxylic acids, esters, ethers | 1020 |
| 700-610 | -C≡C-H: C-H bend | Alkynes | 640 |

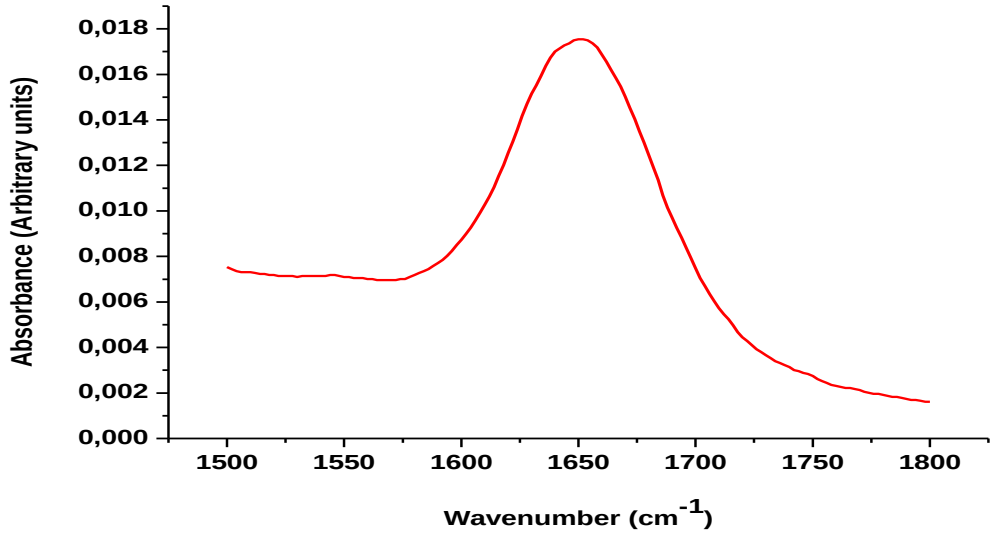



Figure 3.22: FTIR spectra of the expanded view of the Amide I region (1600 cm $^{-1}$-1700 cm $^{-1}$), a broad peak due to the overlapping secondary components peaks.

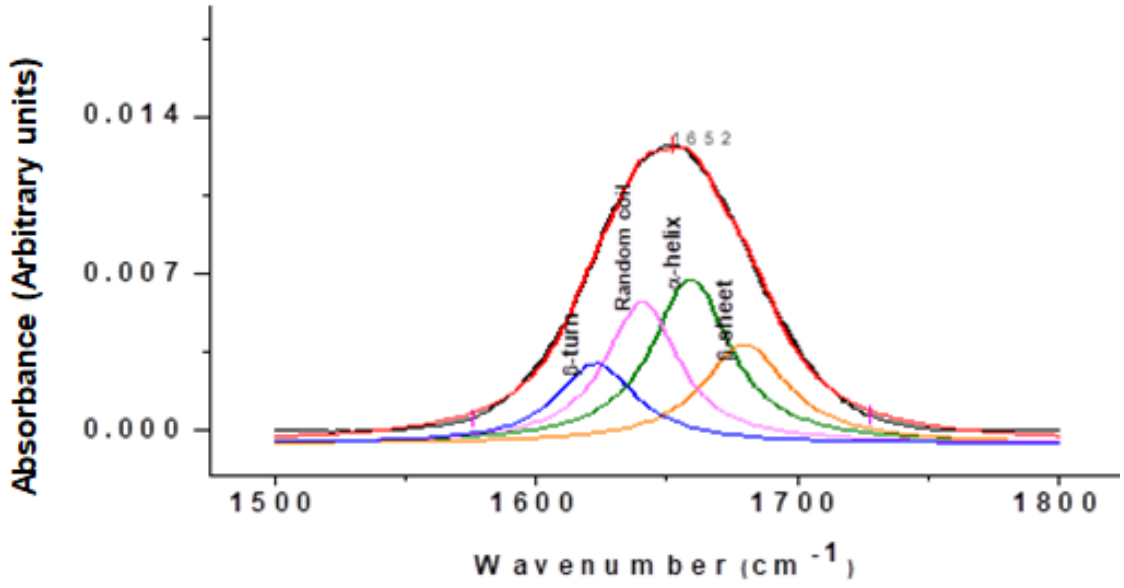



Figure 3.23: Second derivative FTIR spectra of the Amide I region (Red) resolved into its secondary structure components.

Baseline was corrected before deconvolution under the Gaussian model. The Amide 1 region was mathematically resolved under the Gaussian model into α-helices (Green), Random coils (Pink), β-sheets (Orange) and β-turn (Blue).

3.9.3 Intrinsic and extrinsic fluorescence for tertiary structure

The tertiary structure of the Rep protein was determined using a fluorescence spectrophotometer. The indole ring on the seven tryptophan residues of Rep protein was excited at a wavelength of 295 nm (see Figure 3.24) while both the phenyl ring of tyrosine and the indole ring of tryptophan were excited at 280 nm (see Figure 3.25) which resulted in an emission wavelength maxima at 350 nm.

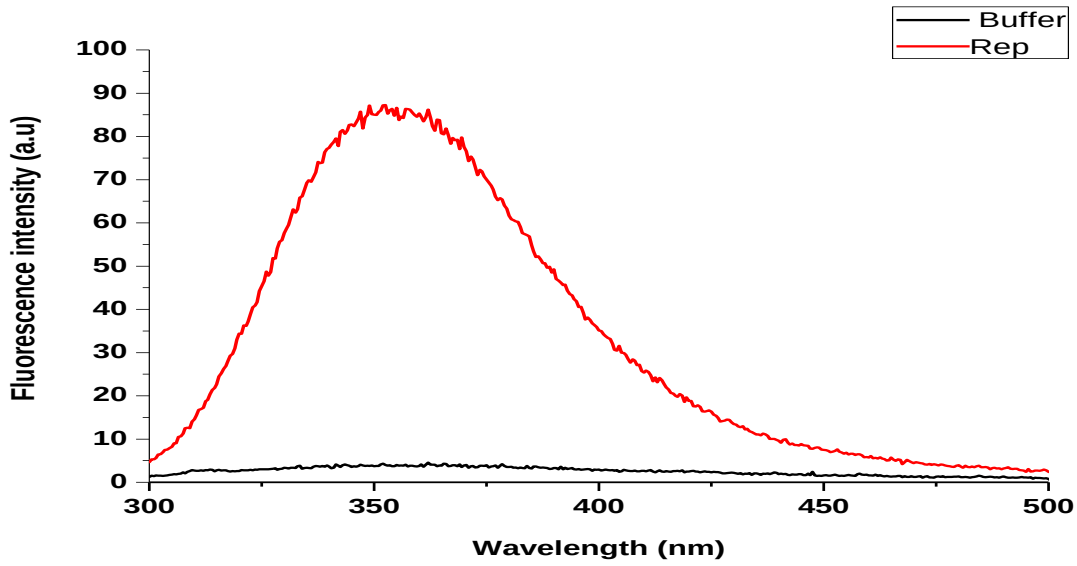



Figure 3.24: Fluorescence emission spectra of 3 µM Rep protein (red line) and buffer (20 mM sodium phosphate, pH 8.00, 5 mM DTT, 10 % glycerol) when excited at 280 nm.

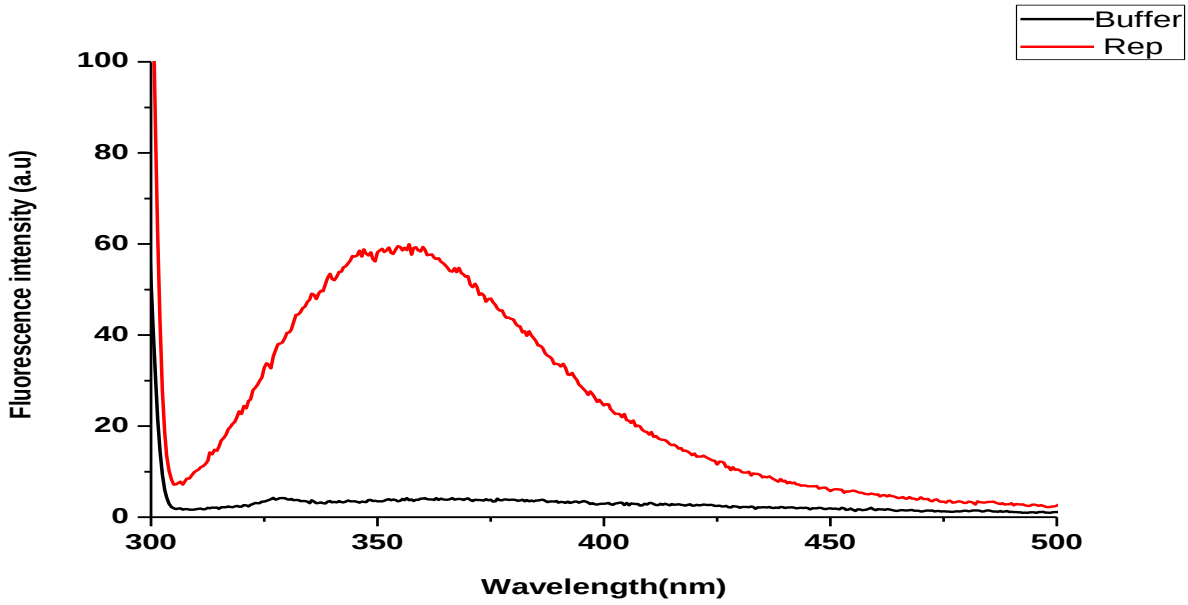



Figure 3.25: Fluorescence emission spectra of 3 µM Rep protein (Red line) and buffer (20 mM sodium phosphate, pH 8.00, 5 mM DTT, 10 % glycerol) when excited at 295 nm.

### 3.10 Functional analysis of the purified recombinant Rep proteins

To determine the integrity and functional activities of the purified and refolded recombinant Rep protein, EMSA for DNA binding and cleavage assays were done. The different concentrations of Rep protein were tested for activity. The assay in Figure 3.26 clearly shows the binding activity of the purified protein, as the band in the negative control well (unbound DNA) migrates at a different speed compared to the band in the experiment wells with due to protein-DNA binding. The cleavage activity of Rep protein to ssDNA substrate was also assessed in an EMSA. Figure 3.27 shows that the ssDNA in the presence of purified Rep protein resulted in two bands compared to the negative control with free ssDNA. Figure 3.28 shows the binding activity of ≥95% purity Rep protein that was SDS-Sarkosyl unfolded and refolded to its native structure. The activity of the protein was concentration-dependent (see Figure 3.28, Lane 4-12) but reaches saturation where not all dsDNA binds to the protein.

3.10.1 Electromobility Shift Assay (EMSA)

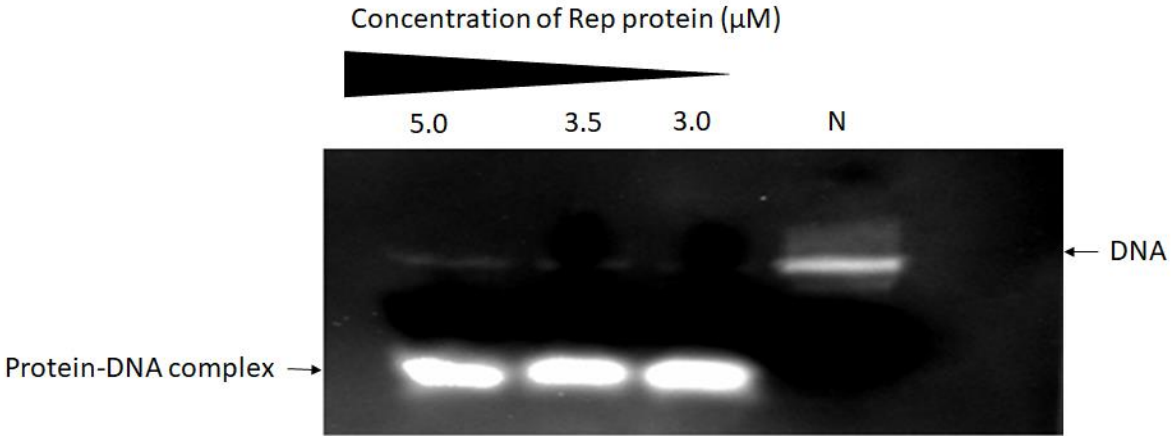



Figure 3.26: A 3% agarose gel in Borex buffer (100 mM borate, 150 mM NaCl, pH 7.5) electrophoresed at 150 V.

It represents DNA binding assay with different protein concentrations (5.0 µM, 3.5 µM and 3.0 µM) in reaction buffer with 0.5 µM dsDNA-FITC. Lane N: DNA only.

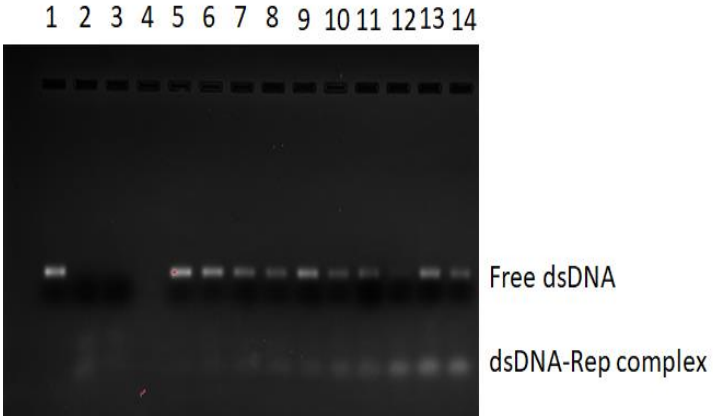



Figure 3.27: A 3% agarose gel in Borex buffer (100 mM borate, 150 mM NaCl, and pH 7.5) electrophoresed at 150 V.

It represents DNA binding assay with different concentrations of Rep protein from Lane 5 to 14 (0.1, 0.2, 0.3 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1) and 0.5 µM dsDNA-FITC. Lane 1: DNA only, Lane 2 and 3: 0.4 µM Protein-DNA complex.

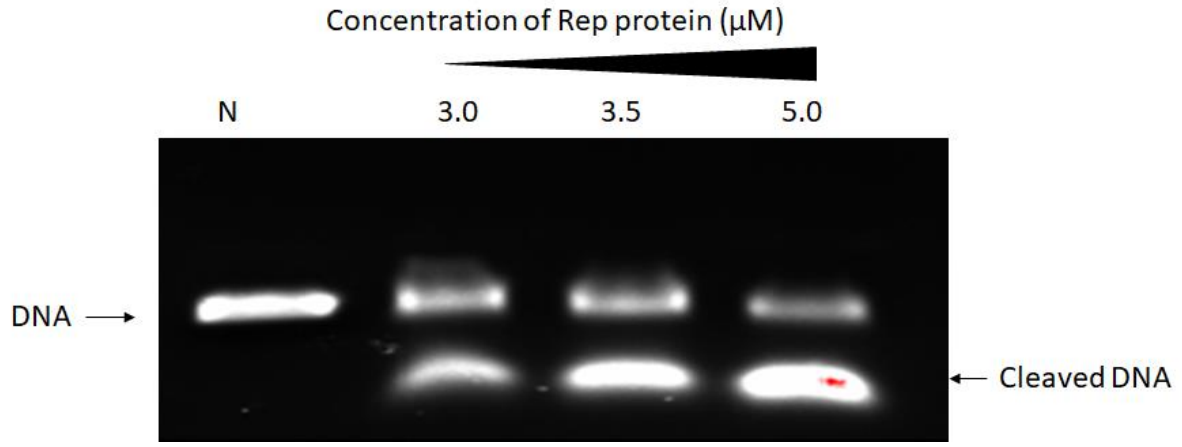



Figure 3.28: A 3% agarose gel in Borex buffer (100 mM borate, 150 mM NaCl, pH 7.5) electrophoresed at 300 V.

It represents DNA cleavage assay with different concentrations of protein (5.0 µM, 3.5 µM, and 3.0 µM) in reaction buffer with 0.5 µM single-stranded DNA. Lane N: DNA only.

### 3.10.2 Enzyme-linked Immunosorbent Assay (ELISA)

To develop a high throughput screening assay for small molecule inhibitors, the binding activity of Rep protein was assessed in a 96-well microtiter plate. The purified protein incubated with FITC-labelled dsDNA formed a complex detected using an anti-FITC antibody that is conjugated to HRP. A negative control reaction was prepared without Rep protein. This activity was analysed by adding TMB substrate for 5 minutes in the dark to develop a blue colour and read at 630 nm (Figure 3.29). The results showed the binding of Rep protein to dsDNA when looking at the absorbance of the positive reaction as compared to the negative reaction. Therefore, this assay is amenable for high throughput screening of compounds.

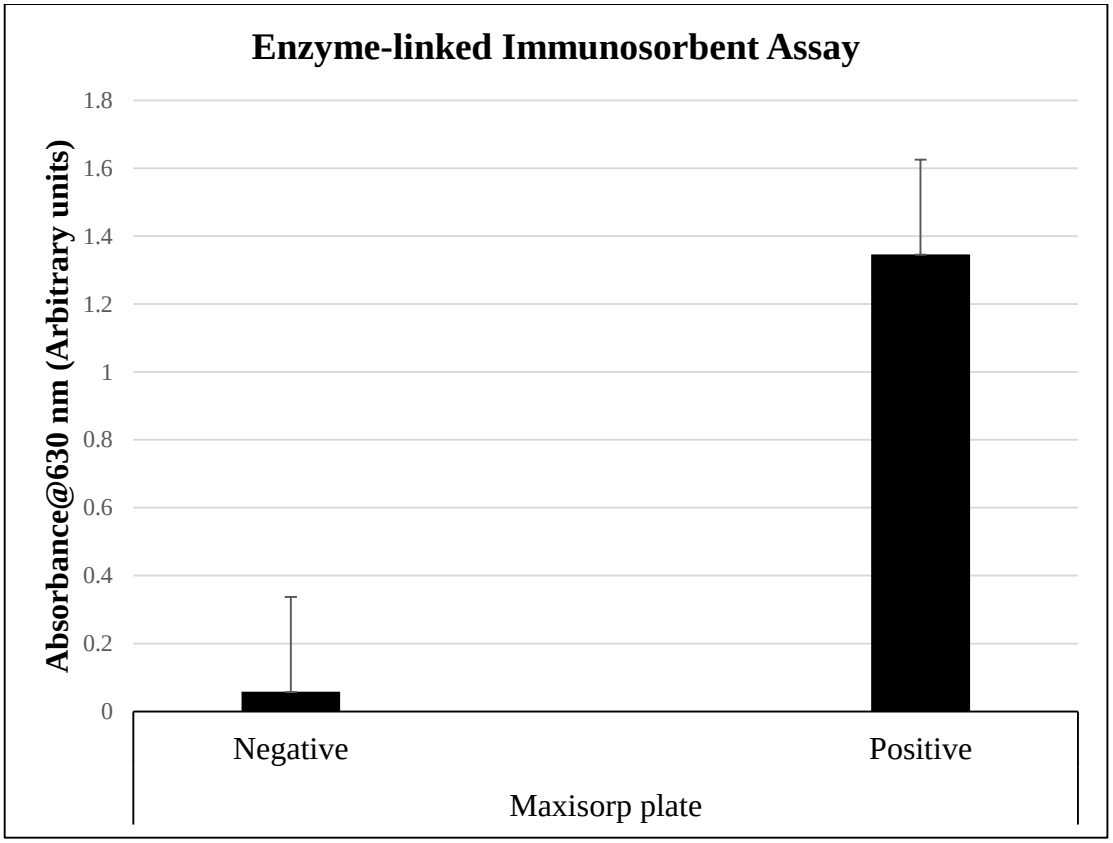



Figure 3.29: ELISA to confirm the binding of ACMV Rep protein with ACMV dsDNA.

The responses indicate binding between Rep and the dsDNA (positive) as compared to the control (negative). The absorbance was measured at 630 nm and the error bars represent standard deviation from two independent experiments.

### 3.11 Inhibition studies

To test the inhibition potential of polyphenolic compounds (CA and EGCG) on the interaction between Rep proteins with ACMV DNA, EMSA for the binding and cleavage assays were performed. A reaction of different concentrations Rep protein with DNA (dsDNA for binding or ssDNA for cleavage) in the presence of CA or EGCG after incubation in a reaction buffer for 1 hour at room temperature (25 °C) was analysed on a 1 % agarose gel. Results shows that inhibition of the binding activity between Rep and its substrate dsDNAstarts at a concentration of ≥ 11µM where a concentration of 100 µM completely inhibited the activity (see Figure 3.30). CA also inhibited the binding interaction of Rep protein with dsDNA at a concentration of 100 µM (see Figure 3.31). On the cleavage activity of Rep protein on ssDNA, inhibition is detected at concentration 33 µM of EGCG (see Figure 3.32 A) with complete inhibition at concentrations ≥ 100 µM (see Figure 3.32 B). CA shows no inhibition of the protein-DNA complex, CA exhibits fluorescence characteristics under UV light. (see Figure 3.33).

3.11.1 Inhibition of protein-DNA binding

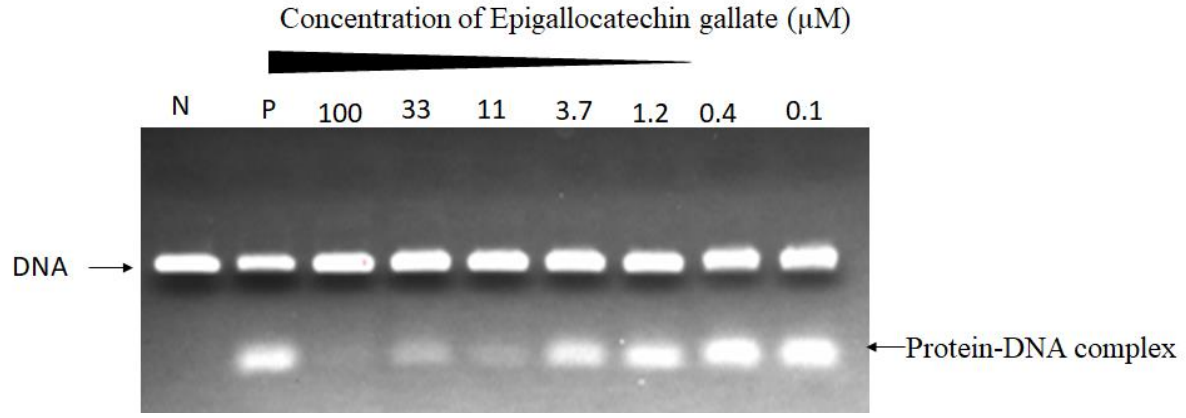



Figure 3.30: A 3% agarose gel in Borex buffer (100 mM borate, 150 mM NaCl, and pH 7.5) electrophoresed at 150 V.

It represents DNA binding assay with different concentrations of EGCG compound in reaction with 0.4 µM Rep protein and 0.5 µM dsDNA-FITC. Lane N: Free dsDNA, Lane P: protein-DNA complex, Lane 0.1, 0.4, 1.2, 3.7, 11, 33, and 100: increasing concentration (µM) of EGCG compound in a reaction with DNA and protein. Inhibition is at concentrations ≥ 11µM.

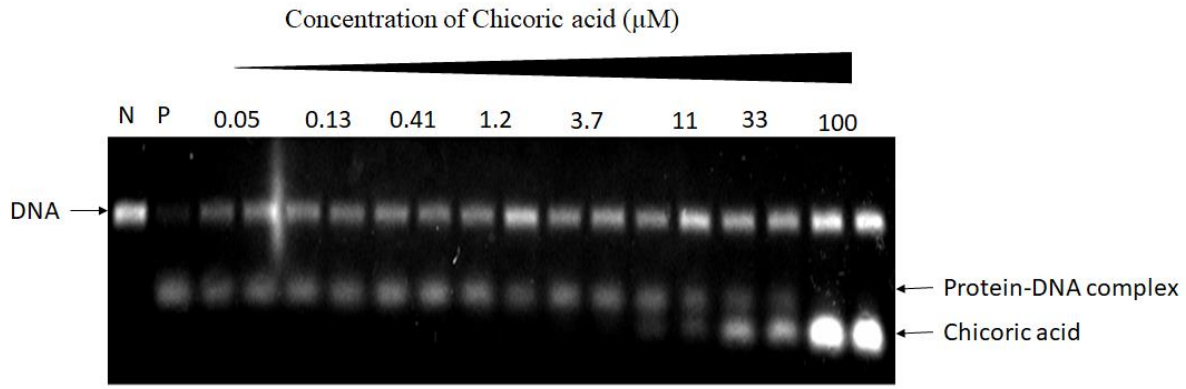



Figure 3.31: A 3% agarose gel in Borex buffer (100 mM borate, 150 mM NaCl, and pH 7.5) electrophoresed at 150 V.

Lane N: Free dsDNA (negative control), Lane P: protein-DNA complex (positive control), Lane 0.05, 0.13, 0.41, 1.2, 3.7, 11, 33 and 100: DNA binding assay with different concentrations of CA compound in reaction with 0.4 µM Rep protein and 0.5 µM dsDNA-FITC. CA exhibits fluorescence characteristics under UV light.

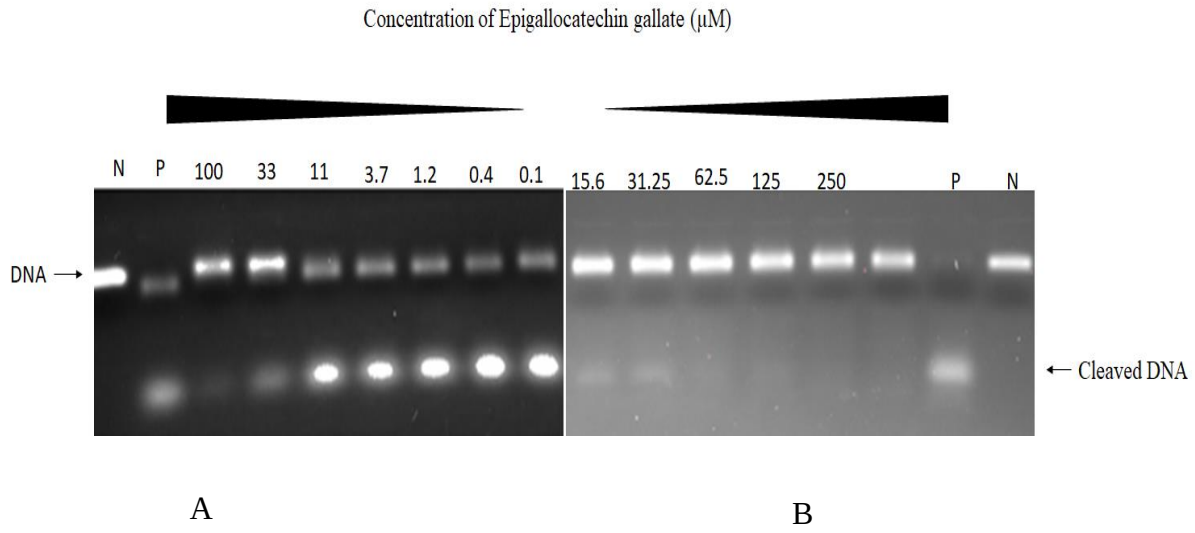



Figure 3.32: A 3% agarose gel in Borex buffer (100 mM borate, 150 mM NaCl, pH 7.5) electrophoresed at 150 V.

(A) Represents DNA cleavage assay with different concentrations of EGCG compound in reaction with 0.4 µM Rep protein and 0.5 µM ssDNA-FITC. (B) represents the same cleavage assay but at higher concentrations of the EGCG compound. Lane N: Free dsDNA (negative control), Lane P: protein-DNA complex (positive control). The inhibition of the cleavage assay by EGCG is at ≥ 15.6 µM on both A and B.

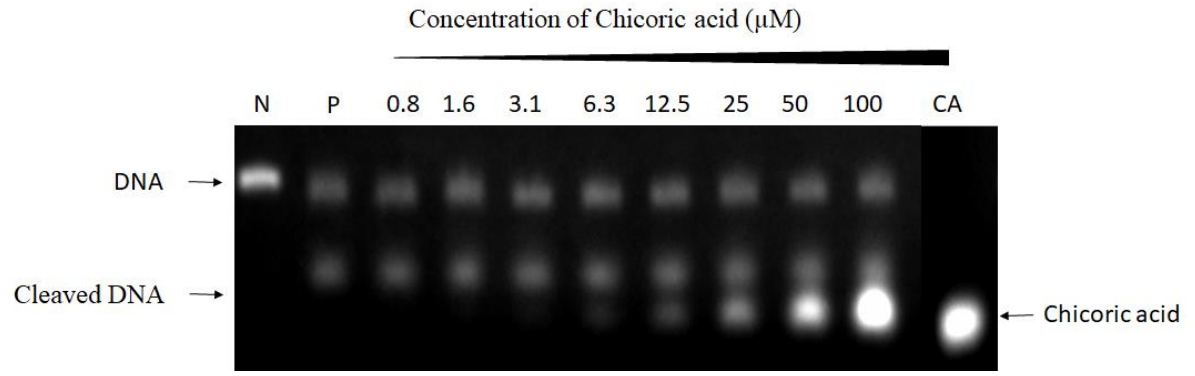



Figure 3.33: A 3% agarose gel in Borex buffer (100 mM borate, 150 mM NaCl, pH 7.5) electrophoresed at 150 V.

It represents DNA cleavage assay with different concentrations of Chicoric acid compound in reaction with 0.4 µM Rep protein and 0.5 µM ssDNA-FITC. Lane N: Free dsDNA (negative control), Lane P: protein-DNA complex (positive control). CA compound does not inhibit the cleavage activity of Rep protein for ssDNA (Lane 0.8, 1.6, 3.1, 6.3, 12.5, 25, 50 and 100 µM). CA exhibits fluorescence characteristics under UV light.

# Chapter 4: **Discussion and Conclusion**

The overall aim of this project was to study *Geminivirus*'s replication-association protein (Rep) as a target for the development of potential small molecule inhibitors. Isolation of target protein from bacterial cell culture is a crucial step towards the biochemical and biophysical characterisation of protein in order to identify potential inhibitors. However, overexpression and purification procedures to obtain high yield and purity are difficult for many studies as it varies for each protein (Lee, 2017). Therefore, common protocols may require optimisation. A common objective is to use a protocol that is easy, fast, cost-effective and versatile (Riek *et al*., 2009). A full-length ACMV Rep protein has never been overexpressed and purified in high yields together with the purity that is enough to perform structural characterisation; hence, there is no structural information for this target protein in literature. This study describes the successful establishment of overexpression and purification methods for a full-length ACMV Rep recombinant protein. The purified protein was then biophysically characterised through FTIR and fluorescence spectroscopy. The activity of the isolated ACMV Rep was then assessed through EMSA, a method commonly used by scientists for analysing the activity of proteins that interact with nucleic acids (Perumal *et al*., 2018). Validation of the functionality of the purified protein was performed through an ELISA. Successful identification of small molecule inhibitors was also pursued in this study through EMSA. The study also demonstrated that CA and EGCG, which are naturally occurring plant phenolic compounds inhibits the interaction between ACMV Rep protein and ACMV DNA. Only these two compounds were tested in this study; therefore, other compounds of the same characteristics can be tested for potential inhibition.

## 4.1 Verification of the insert cDNA in the pET-15b plasmid DNA

Verification of the cDNA inserted into pET-15b plasmid was performed by Sanger DNA sequencing using the universal forward and reverse primers of the T7 promoter (Inqaba Biotechnical Industries, South Africa) shown in Table 2.1. Sequence similarities between the sequencing results and the ACMV Rep codon optimised sequence showed 92.47% identity. Codon optimisation performed at GenScript was a strategy for increasing protein production based on the assumption that rare codons limit the rate of protein synthesis in bacterial systems. It alters DNA sequence with silent mutations that improve the transcriptional and translational

efficiency of the target gene but brings no change to the amino acid sequence of the expressed target protein (Chen and Inouye, 1994; Mauro and Chappell, 2014).

### 4.2 Overexpression and purification of recombinant ACMV Rep protein

Overexpression and purification of recombinant protein form the basis of many studies that are aimed at characterising recombinant proteins in drug discovery. For a protein to be classified as a recombinant protein, its DNA has to be cloned into expression vectors and then subsequently transformed into systems such as eukaryotic systems (mammalian cells, yeast, and insect cells), prokaryotic systems (*Bacillus subtilis* and *Escherichia coli*) and in vitro systems (Jia and Jeon, 2016) using DNA recombinant technology. *E. coli* is commonly used for the production of recombinant protein through recombinant DNA technology because it is easy to grow, inexpensive and provide a fast high-density cultivation (Baneyx, 1999). Despite the advantages of using *E. coli* in overexpression of recombinant proteins, other problems often arise as they result in low expression levels of readily soluble protein but highly insoluble "inclusion bodies" and non-functional protein (Fakruddin *et al*., 2013). Also, expression and purification of proteins using *E. coli* systems are time-consuming and labour-intensive. *E. coli* BL21 (DE3) pLysS competent cells were used in this study containing ACMV Rep vector (see Figure 3.3). Previously, Njengele *et al*. (2016), have successfully used BL21 (DE3) pLysS to overexpress full-length recombinant HIV-1-Vpu protein. Another study has successfully overexpressed Ra-eRF1 in *E. coli* BL21 (DE3) pLysS strain (Karamyshev *et al*., 1999). As for this study, strains such as BL21 (DE3) pLysS, BL21 (DE3) and the T7 *E. coli* cells were tested for overexpression of ACMV Rep protein but only BL21 (DE3) pLysS was better suited in overexpressing the target protein. These results are due to the fact that lysozyme lowers the background expression level of target genes under the control of T7 promoter but does not interfere with the level of expression achieved following induction by IPTG.

Critical factors that affect the rate at which *E. coli* cells produce heterologous protein are factors such as transcriptional promoter strength, which in this study was addressed by using a vector that has a gene of interest immediately downstream of the strong T7 inducible promoter. The promoter is not always on but rather allows to be turned on at a specific growth time by a metabolite such as isopropyl β-d-1-thiogalactopyranoside (IPTG) (Swartz, 2001; Briand *et al*., 2016). The use of IPTG as an inducer has its own limitations such as high cost, it requires cell growth of bacteria to be monitored until they reach optimal cell density before use. In this study, different concentrations of IPTG were analysed for high expression levels of ACMV

Rep protein (See Figure 3.10 A), it was found that a concentration between 0.5 and 1 mM IPTG was optimum. A concentration of 1 mM IPTG is mostly used in literature (Zininga *et al*., 2015; Park *et al*., 2015; Njengele *et al*., 2016).

Furthermore, bacterial cell growth is also a factor that affects expression levels of recombinant enzymes. Growth media can be manipulated by changing several conditions such as the type of nutrient broth used to grow bacteria, pH and temperature of the growth media. Therefore, it is important to accurately select growth media and conditions to achieve high yields of a target recombinant protein that is readily soluble (Sivashanmugam *et al*., 2009; Fazaeli *et al*., 2019). For this study, optimisation of protein expression was done looking at the type of nutrient media used and changing the temperature of the media for the optimal growth of bacteria. The results indicated that terrific broth and 2×YT Broth are the best options as compared to LB for the production of ACMV Rep protein (see Figure 3.9 A). These results are consistent with other studies in the literature that showed rich nutrient media such as 2×YT, Terrific Broth, and Super Broth are a better option for protein production (Kram and Finkel, 2015; Fazaeli *et al*., 2019).

In addition to finding optimum conditions for the overexpression of recombinant ACMV Rep protein, the use of glucose as a carbon source was explored. Terol *et al*. (2019) has reported that glycerol or glucose as a carbon source has a positive effect on the production of protein in acetate metabolism. The study also showed that glucose is the preferred carbon source for *E. coli* cells compared to glycerol. Production of ACMV Rep was improved by exploring different concentrations of glucose for cell growth and it was found that a concentration ≥ 1% increase expression levels in a concentration-dependent manner. Pre-induction and post-induction times are also important factors to be studied for the production of every recombinant protein. Pre-induction time studies, where the time at which bacterial cells reach an optical density of 0.6 was found to be three hours post-inoculation (see Figure 3.6) and the time at which bacterial cells have expressed sufficient protein was found to be three hours post-induction with IPTG. Lastly, it is reported that a decrease in temperature post-induction allows the bacterial cells to grow slowly (overnight) and simultaneously produce a more soluble recombinant protein (Glick and Whitney, 1987; Schein and Mathieu, 1988; De Groot and Ventura, 2006). However, despite reports in literature on the positive effect of decreased temperature on the solubility and production, a decrease in temperature did not make any improvement in this study (data not shown). Therefore, the optimum conditions for overexpression of recombinant ACMV Rep protein from this study are found to be, the use of the *E. coli* bacterial strain BL21 (DE3) pLysS

grown in TB medium that is supplemented with ≥ 1% glucose and induced for protein expression using 0.5 mM IPTG-1 mM IPTG for 3-5 hours post-induction.

The target protein is produced as an insoluble fraction (inclusion bodies) due to improper refolding of eukaryote proteins in *E. coli* cells, where they are either digested quickly by proteases or accumulated as inclusion bodies (Musa *et al*., 2012). After successfully optimising expression yields, the problem was then to express soluble proteins in *E. coli*. As previously mentioned that a decrease in temperature to address this problem was attempted without success. Therefore, the insoluble fraction became the focus of subsequent purification procedure for Rep protein (see Figure 3.10, Lane 2). Inclusion bodies are water-insoluble proteins expressed as aggregates by *E. coli* harbouring target genes from viruses or mammalians. These inclusion bodies are normally solubilised using chaotropic agents such as urea or guanidinium chloride (GnHCl) (Njengele *et al*., 2016; Mohammadian *et al*., 2018). Once the protein is solubilised, different purification techniques are used to separate target protein from bacterial proteins. Some of these techniques are affinity chromatography, ion-exchange chromatography, hydrophobic interaction chromatography, and size exclusion chromatography (Saraswat *et al*., 2013).

Affinity chromatography is a technique used to separate target protein that has a high affinity for the resin. These proteins are normally infused with tags (a sequence of amino acids) that have a high binding affinity for a biological ligand or a chemical group immobilised on the resin. Examples of these tags are Histidine (His) tag, Maltose-binding protein (MBP) tag, Glutathione S-transferase (GST) tag and FLAG (Futatsumori-Sugai *et al*., 2009; Bornhorst *et al*., 2000; Duong-Ly and Gabelli, 2015; Harper and Speicher, 2011). In this study, nickel affinity chromatography was used to separate ACMV Rep protein with a Histidine tag solubilised using an affinity competing for chemical imidazole (pH 8) (see Figure 3.12). Upon analyses of protein purity on SDS-PAGE, most Rep proteins were eluted with unwanted bacterial protein in earlier fractions while low concentrations were eluted at later fractions (see Figure 3.13). Elution of the target protein using chromatography differs for every protein hence the need to optimise the conditions. Optimisation of ACMV Rep protein purification condition was attempted by changing the buffer composition (data not shown).

Finally, purification of a sufficient amount of ACMV Rep protein was detected when using a buffer containing SDS as a solubilising agent and purifying with buffer containing Sarkosyl onto a Nickel affinity column. Schlager *et al*. (2012) published a paper showing the successful

use of anionic denaturing agents to purify insoluble proteins; hence, this provided the basis of using the protocol in this study. Compared to urea that yields a random coil upon interaction with proteins, SDS yields alpha-helical conformations (Neilsen *et al*., 2007; Pace and Tanford, 1968). Unfortunately, Schlager *et al*. (2012) did not report the refolding protocol of proteins purified in that study but had suggested the use of urea to replace SDS and ultimately remove urea in stepwise dialysis. The concentration of SDS was decreased before binding by incubating at 4 ºC (Tsumoto *et al*., 2009; Schlager *et al*., 2012; Tao *et al*., 2018).

To further remove SDS from the protein sample, the sample in urea buffer (pH 7.0) was loaded onto ion-exchange chromatography (HiTrap Q Sepharose HP) (see Figure 3.14). Based on the pI (7.66) of the protein, the protein at pH 7.0 has a positive net charge of +2. As a result, the positively charged protein did not bind to the positively charged Q Sepharose resin and, therefore, was collected as flow-through (see Figure 3.17). The negatively charged SDS was bound to the Q Sepharose resin as a contaminant. Removal of SDS from the resin for maintenance purpose of the column was successfully done using 70% ethanol. The purified protein was successfully refolded in stepwise dialysis while monitoring protein aggregation with a UV- spectrophotometer.

A common technique discovered by Towbin *et al*. (1979) is Western Blot, which is used for the immunodetection and quantitation of target proteins. It is a sensitive, robust and flexible technique that requires low concentration of antibody. Successful Western Blot using a monoclonal anti-His antibody to detect His-tagged ACMV Rep protein was performed (see Figure 3.16).

## 4.3 Structural determination of recombinant Rep proteins

### 4.3.1 Fourier-transform infrared (FTIR)

Characterisation of protein structure is done in so many ways such as the x-ray crystallography, circular dichroism, and FTIR to mention a few. FTIR analysis depicts the vibrational patterns of functional groups that are present in macromolecules and express these molecular structural changes through shifts in wavenumbers (see Figure 3.20 and Figure 3.21). The concentration of the relative functional group that absorbs in a specific infrared region is shown by the intensity of the signal (Long *et al*., 2015). FTIR is particularly sensitive to the secondary structure of proteins. A protein has two prominent features that are depicted in the infrared region of the spectra, these features are the Amide I (1700 $cm^{-1}$-1600 $cm^{-1}$) and Amide II (1600 $cm^{-1}$-1500 $cm^{-1}$). The Amide I region depicted in Figure 3.22 is due to the C═O stretching

while the Amide II region that is not depicted is due to the N-H bending vibrations (Barth, 2019; Miller *et al*., 2013). The Amide I sensitivity is attributed by the composition of secondary structure with α-helix, β-sheet, turn and unordered conformations overlapping on each other (Yang *et al*., 2015; Usoltsev *et al*., 2019). FTIR analysis in this study presents the Amide I region of Rep protein in storage buffer (25 mM sodium phosphate, pH 8, 500 mM sodium chloride, 10% glycerol and 5 mM DTT) with the peak at frequencies 1656 cm-1 (see Figure 3.22). This proves that the purified ACMV Rep protein has refolded to its native secondary structure after using denaturant (urea) to unfold it and allowing refolding to take place while removing the urea stepwise. Unfortunately, it can only provide information on whether there is a secondary structure present or not. The type of secondary structure was further determined by resolving the composition intensities using a mathematical expression self-deconvolution (curve fitting) analysis tool on OriginPro 8 software under the Guass-Lorentzia conditions. Before deconvolution, the Amide I region was first corrected to baseline. The self-deconvolution of the Amide I region to its second derivative (see Figure 3.23) shows that the Amide I is composed of alpha (α) helices (Green), random coils (Pink), beta (β) sheets (Yellow) and beta (β) turns (Blue). These results are consistent and, therefore, corroborate the literature. Pauthe *et al*. (2002) have used FTIR together with deconvolution expression to look at the secondary structure of Fibronectin protein and its changes induced by temperature. The interaction of BSA with nucleic acid was characterised using FTIR with the curve fitting of the amide I region (Tajmir-Riahi *et al*., 2009).

### 4.3.2 Fluorescence spectroscopy

Fluorescence spectroscopy is a technique that provides a method of characterisation for protein conformation by taking advantage of the sensitivity of chromophore in proteins. These chromophores [tyrosine (Y), tryptophan (W), and phenylalanine (F)] have the ability to absorb visible or ultraviolet light by exciting electrons to a higher energy level at a specific wavelength. The absorbed light is then re-emitted as fluorescence at a wavelength that is longer than the excitation wavelength when it loses the energy to return to its ground state (emission wavelength maxima) (Pain, 2004). The maximum emission wavelength (λmax), in the case of tryptophan, is dependent mostly on the polarity of the local environment surrounding the chromophore in a protein. In most proteins, the emission λmax is around 305 nm (less polar environment) to 355 nm (high polar environment) (Lakowicz, 1983; Pain, 2004). ACMV Rep protein contains aromatic amino acids (aa) namely: tyrosine (ten), tryptophan (seven), and phenylalanine (nineteen). Among the three intrinsic fluorophores, tryptophan has the highest

intensity because of its high molar extinction coefficient of 5500 $M^{-1}.cm^{-1}$ at 280 nm than the other chromophores (Pace, 1968; Kozlov *et al*., 2012). The excitation wavelength was at 295 nm for tryptophan residues (see Figure 3.24) and 280 nm for both tyrosine and tryptophan (see Figure 3.25). This resulted in an emission wavelength longer than the excitation wavelength which is at 355 nm; hence, this is called the red-shift. ACMV Rep protein contains a total number of seven tryptophan and ten tyrosine residue whose position is not known because there is no crystal structure of the protein deposited in the Protein Data Bank (PDB). The fluorescence intensity of the protein at 295 nm is lower than that of 280 nm because only tryptophan residues are excited as compared to both tryptophan and tyrosine, respectively. From fluorescence done in this study, it is shown that ACMV Rep protein has a maximum emission wavelength of 355 nm for both 280 nm and 295 nm excitation wavelength. Peter *et al*. (2013) also reported the maximum emission wavelength of chloride intracellular channel protein 1 (CLIC1) to be at 355 nm and this indicates that the tryptophan residues which absorbed light are to be exposed to the solvent. There is no fluorescence spectrum for the protein of the same family to make a comparison with.

### 4.4 Functional studies of the unfolded and refolded ACMV-Rep

After isolating the protein using purification and refolding techniques, the integrity and the activity of the protein was then examined through Electrophoretic Mobility Shift Assay (EMSA). This assay assesses the interaction of target ACMV Rep protein and a specific region on the wild type dsDNA oligonucleotide (see Table 2) corresponding to the 5' intergenic region of *Tomato Leaf Curl New Dehli* Virus (ToLCNDV) (Garner and Revzin, 1981, Chatterji *et al*., 2000). The basic principle of this assay is that it separates unbound DNA fragments from protein-DNA complexes under native conditions on polyacrylamide or agarose gels, free unbound DNA fragments migrate faster than protein-DNA complex. Because of the negatively charged phosphate backbone of the DNA, the separation on agarose occurs from the negative electrode (cathode) to the positively charged electrode (anode). Separation on an agarose gel is determined by the size of the molecule, the percentage concentration of agarose, the conformation of DNA and the electrophoresis buffer used (Lee *et al*., 2012). In this study, the size of DNA was not the determinant of the rate of migration because one primer of the same size and charge was used, only the protein concentration used differs. The migration distance depends on the conformation of DNA upon interacting with protein as compared to the free unbound DNA. From the results of the EMSA in this study, it clearly shows the activity of the protein bound to DNA (see Figure 3.26) when comparing the free unbound DNA migration

distance with the protein-DNA complex. According to literature, the protein-DNA complex will move slower in the gel than an unbound free DNA because of the size (Yakhnin, *et al*., 2012). In general, the rate at which the protein-DNA complex migrates faster than the free DNA is with the exception when the substrate is a circular DNA (Krämer *et al*., 1988). In this study, the electrophoretic mobility of the protein-DNA complex is faster than the unbound free dsDNA (see Figure 3.26). Different concentrations of Rep protein were assessed, with the free dsDNA as a negative control. These results do not correlate with the basic principle of EMSA and what other studies in literature have reported on the migration patterns of protein-DNA complexes (Han *et al*., 2008; Steiner and Pfannschmidt, 2009; Perumal *et al*., 2018). The reason for this migration pattern is not known. This could be due to the Rep protein not binding on the double-stranded oligonucleotide but cleaving, resulting in two identical fragments. The reason for this assumption is that clearly there is an activity of the Rep protein that resulted in the migration pattern of the protein-DNA reaction that differs from the free DNA reaction. Studies such as the fluorescence anisotropy could be performed to monitor the association and dissociation of Rep protein with the FTIC-labelled oligonucleotide. This technique can be used to validate the binding activity of the protein. The use of radioactivity probe to monitor protein-DNA interaction has been predominantly used for EMSA, but fluorescence tags such as FITC on DNA have been explored recently. FITC has an advantage over radioactivity due to its ease at handling, reduced cost, time effective and ultimately improving safety (Hsieh *et al*., 2016). The migration was monitored using the FITC label incorporated at any position within the DNA via the NHS conjugation chemistry. Because the results were not correlating with literature as mentioned above, optimisations of the assay were done by changing the different parameters such running buffer (1× sodium borate/ Tris-borate-EDTA/ Tris-acetic acid-EDTA, pH 7.5), voltage from 150 V to 80 V, concentration of the matrix from 3% to 1% or 0.8%, the type of gel from agarose to polyacrylamide. Regardless of the optimisation conditions tested, the electrophoretic mobility of the protein-DNA complex did not change. However, the main aim which was to show that the binding activity of the ACMV Rep protein was achieved. The best EMSA conditions were chosen to be sodium borate, pH 7.5, 3% agarose, electrophoresed at 150 V for 30 minutes, the rest of the conditions tested were either smearing or little to no difference on the migration rate between the complexed and the un-complexed DNA. These activity studies were done with a protein purified in Figure 3.17. After optimisation of purification and refolding using 1% SDS, 0.1% Sarkosyl and 8 M urea, the protein was checked for activity in a concentration-dependent manner (see Figure 3.27 ). The concentration that showed optimum activity from Figure 3.27 was 0.8 µM.

The same protein was checked for its catalytic activity using ssDNA substrate (Table 2.2) electrophoresed on a 3% agarose gel. ACMV Rep protein cleaves the ACMV viral genome at a conserved specific site in a hairpin loop of the (+) strand origin of *Geminivirus* (Heyraud-Nitschke *et al*., 1995; Hull, 2014). It was first demonstrated that Rep protein of TYLC has cleavage activity by nicking the TAATATTACCGG oligonucleotide that is conserved among all Geminiviruses. After cleavage, Rep protein remains bound to the 5' end of the nicked oligonucleotide and analysed on a gel shift assay (Laufs *et al*., 1995; Kittelmann *et al*., 2006). Hipp *et al*. (2014) reported the cleavage activity of ACMV Rep protein on conserved oligonucleotide analysed using western blots assay. The study also shows the binding of Rep protein to the 5' end of the oligonucleotide, which is consistent with previous studies mentioned. In this study, the results show that the un-cleaved ssDNA fragment appears as a single band (see Figure 3.28) and the cleaved products appear as two distinct bands (see Figure 3.28, Lane 2 and 3). These results are due to ACMV Rep protein nicking the ssDNA (Table 2.2) primer into two fragments, one fragment remains bound to the purified full-length Rep protein using its 5' end of oligonucleotide. These results suggest that the binding of the protein to the 5' end of the cleaved DNA fragment caused the complex to migrate slower on agarose gel due to the overall size of the complex as compared to the free cleaved DNA fragment. Comparing the uncleaved wt. ssDNA to the cleaved (see Figure 3.28), the migration pattern is different suggesting that there was a successful cleavage. EMSA technique is also used to determine the catalytic activity of proteins on nucleic acids (Kim *et al*., 2014; Eskandarian *et al*., 2019).

Enzyme-linked immunosorbent assay (ELISA) was successfully used as a secondary protein-DNA binding technique to assess the binding activity of ACMV Rep protein. This technique is traditionally used to detect a substance through the interaction of an antibody and its antigen (Chen *et al*., 2018). The advantage of ELISA is that it is simple, highly sensitive and offers stability and rapid analysis. However, the disadvantage is that it has a high noise to signal ratio (Sakamoto *et al*., 2018). One of the requirements for a successful ELISA is to reduce the high noise to signal ratio by using blocking agent to prevent non-specific binding of enzyme reagents to the microtiter plate surface. Failure to block the non-specific binding of enzyme reagents or serum antibodies will results in the increase of absorbance readings and causing the results to be analysed as positive, whereas they are a false positive. Different blocking agents such as such as skimmed milk, bovine serum albumin (BSA) and whole serum are commonly used (Xiao and Isaacs, 2012). Douglas *et al*. (1988) concluded that 10% skimmed milk is the

most effective blocking agent as compared to 10% BSA and normal goat serum (NGS). They further recommend the use of skimmed milk as it address the issue of cost, availability and storage. In the current study, a comparison of blocking buffers (BSA and 5% skimmed milk) for optimising sensitivity and minimising background signal was tested (data not shown). The 5% skimmed milk showed most desirable results as compared to BSA. The results presented in Figure 3.29 shows an assay performed using 5% milk as a chosen blocking agent, the difference in optical density ($OD_{600}$) of the experiment (1.346) and the control (0.0582) indicate that the 5% milk made a significant effect on the success of the ELISA. In the present study, ELISA qualitatively confirmed the binding activity of protein to specific DNA primers used in EMSA (see Figure 3.29). By comparing the positive reaction well with the control negative well, it confirms the functional activity of the purified protein. ELISA is mostly developed for high throughput screening of potential inhibitors. Prokoph *et al*. (2016) demonstrated a successful development of novel ELISA for screening of Cdk5-Mediated PPARγ Phosphorylation. Another study has demonstrated the use of ELISA technique for the screening of inhibitors (Yakobov *et al*., 2014). Therefore, an attempt has been made to develop ELISA for high throughput screening of compounds that are potent inhibitors of the ACMV Rep protein interaction.

### 4.5 Inhibition of ACMV Rep protein with ACMV oligonucleotide

As mentioned-above, inhibition of the interaction between ACMV Rep with viral oligonucleotide can be explored. This study focused on assessing the inhibition of protein-DNA interaction by small molecules from plants. Plant phenolics are secondary metabolites that play a crucial role in the defense mechanism of plants against pests, herbivores, and pathogens. There has been an increased demand for these phenolics in pharmaceuticals, cosmetic, food, and pesticides industry (Bourgaud *et al*., 2001) and thus an increase in research focus. Only two phenolic compounds; namely, epigallocatechin gallate and chicoric acid were tested for inhibition in a traditional EMSA used to primarily assess the binding together with the cleavage activity of refolded purified ACMV Rep protein. From the results, EGCG has shown to inhibit both the binding and cleavage activity in a concentration-dependent manner ($\geq 100$ μM) (see Figure 3.30 and Figure 3.32 ) while CA only inhibits the binding activity also in a concentration-dependent manner (see Figure 3.31 ). The use of phenolic compound such as CA was reported to inhibit the activity of tyrosine phosphatase in an allosteric manner (Baskaran *et al*., 2012). Kuban-Jankowska *et al*. (2016) reported the inhibition of CA on YopH bacterial virulence factor in an allosteric manner. The inhibition mechanism of CA on ACMV

Rep interaction with ACMV DNA is not known and was not established in this study. EGCG as a natural derived small molecule compound has also been tested for inhibition of certain enzymes such as the Phosphoglycerate mutase 1 (PGAM1). It was discovered that EGCG inhibits the activity of the enzyme in a concentration depended manner (Li *et al*., 2017).This study is the first study to show the inhibition of ACMV Rep interaction with DNA using naturally derived compounds such as EGCG and CA, the inhibition is accomplished in a concentration dependant manner like most of enzyme inhibition.

In summary, this study reports on a modified and efficient method for the overexpression of a full-length ACMV Rep protein. Because recombinant protein is of importance in the pharmaceutical, food, pesticides and cosmetics industry its production poses a challenge in high scale production. This method proves to be efficient for isolation of recombinant ACMV Rep and it can be used to isolate other proteins that are often produced as inclusion bodies. Of course, optimisation of conditions might have to be performed due to the fact that proteins have different characteristics. Understanding the characteristics such as the pI, pH optimum in terms of stability and activity, the globular and oligomeric status of the enzyme is a crucial step towards the identification of potential inhibitors for their activity. Upon successful purification of sufficient ACMV Rep using the method described in this study, biochemical characterisation of this target protein was successful, where the binding and cleavage activity of the refolded protein was ascertained. A microtiter plate biological assay was established for the interaction of ACMV Rep protein with ACMV oligonucleotide indicating that this assay is amenable to high throughput screening of potential inhibitors. Due to the lack of structural information of this ACMV Rep protein, one of the objectives of this study was to characterise its structure. This method assisted with the isolation of ≥ 95% purified protein suitable for structural characterisation. From this study, the secondary structure of the target protein was determined to be a composition of α-helix, random coils, beta (β) sheet and beta (β) turns. Secondary structure prediction using online software also supports the findings from experimental data in this study. Fluorescence studies that give information about the local changes of the tertiary structure were also performed. The only data obtained from this study was the maximum wavelength of the refolded protein which only suggests that tryptophan was exposed to the solvent. This data cannot identify which tryptophan (position) is exposed since there is no structural information. It could be any of the tryptophan presented in the amino acid sequence.

Due recognition of ACMV Rep protein as an indispensable protein for the replication of ACMV genome, the screening of small-molecule inhibitors was also an objective of this study.

The main focus was to identify potential inhibitors that are of natural origin from plants as they have gained attention as anti-viral agents for many diseases affecting humans, animals, and plants. This study has successfully identified two phenolic compounds (epigallocatechin gallate and chicoric acid) as potential inhibitors for ACMV Rep protein activity.

**For future directions, the following is recommended:**

- The stability study of the protein needs to be done as the protein tends to aggregate in certain buffer conditions that are compatible with other techniques used for characterisation. Circular dichroism, a technique that is commonly used to characterise secondary structure requires low to no chloride buffers. The integrity of Rep protein in low sodium chloride buffer is compromised, therefore, optimisation of buffers that are compatible with the protein integrity should be performed. This will enable characterisation of secondary structure with CD in the future to support the FTIR and prediction data obtained.
- The interaction of ACMV Rep with ACMV DNA is not well-understood. This resulted in the lack of understanding of the migration pattern of protein-DNA complex on agarose when electromobility shift assays were performed. Therefore, fluorescence anisotropy technique to study the interaction is recommended.
- In addition to enzyme-linked immunosorbent assay (ELISA), other high throughput screening assays such as scintillation proximity assay (SPA) may be developed for the screening of small-molecule inhibitors.
- Thermodynamics parameters (stoichiometry, change in enthalpy, binding affinity constant, change in entropy, change in Gibbs free energy and change in heat capacity) can be determined using isothermal titration calorimetry.
- X-ray crystallography can be performed on the purified refolded protein to determine its structure

# Chapter 5: **References**


Abu-Reidah, I. M., Contreras, M. M., Arraez-Roman, D., Segura-Carretero, A and Fernandez-Gutierrez, A. 2013. "Reversed-phase ultra-high-performance liquid chromatography coupled to electrospray ionization-quadrupole-time-of-flight mass spectrometry as a powerful tool for metabolic profiling of vegetables: *Lactuca sativa* as an example of its application". *Journal of Chromatography,* 13. Pp. 212-227.

Allem, A. C and Genèticos, R. 2002. "The origin and taxonomy of cassava, in Cassava Biology, production and utilisation", eds R.J. Hillocks, J.M Thresh, and A. C Bellotti (Willingford:CAB International). Pp. 67-89.

Baneyx, F. 1999. "Recombinant protein expression in *Escherichia coli*". *Current Opinion in Biotechnology,* 10(5). Pp. 411-421.

Barth, A. Usoltsev, D., Sitnikova, V., Nosenko, T., Olekhnovich, R and Uspenskaya, M. 2019. "Comparison of protein secondary structure calculation methods based on infrared spectra deconvolution". *Journal of Mechanical Science and Technology Optics*, 19. Pp. 586-593.

Baskaran, S.K., Goswami, N., Selvaraj S., Muthusamy, V.S and Lakshmi, B.S. 2012 "Molecular dynamics approach to probe the allosteric inhibition of PTP1B by chlorogenic and cichoric acid". *Journal of Chemical Information and Modelling*, 52 (12). 2004-2012.

Bauer R. 1998. "*Echinacea*: biological effects and active principles. In: Lawson L.D., Bauer R, editors. Phytomedicines of Europe: Chemistry and biological activity". *American Chemical Society,* Pp. 140-157.

Bornhorst, J. A and Falke, J. J. 2000. "Purification of proteins using polyhistidine affinity tags". *Methods in Enzymology,* 326. Pp. 245-254.

Borrelli, V., Brambilla, V., Rogowsky, P., Marocco, A and Lanubile, A. 2018. "The Enhancement of Plant Disease Resistance Using CRISPR/Cas9 Technology*". Frontiers in Plant Science,* 9 (1245) Pp. 1-15.

Bourgaud, F., Gravot, A., Milesi, S and Gontier, E. 2001. "Production of plant secondary metabolites: a historical perspective". *Plant Science,* 161(5). Pp. 839-851.

Bredeson, J.V., Lyons, J.B., Prochnik, S.E., Wu, G.A., Ha, C.M., Edsinger-Gonzales, E., Grimwood, J., Schmutz, J., Rabbi, I.Y., Egesi, C and Nauluvula, P. 2016. "Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity". *Nature Biotechnology*, 34(5). Pp.562-570.

Briand, L., Marcion, G., Kriznik, A., Heydel, J. M., Artur, Y., Garrido, C., Seigneuric, R and Neiers, F. 2016. "A self-inducible heterologous protein expression system in *Escherichia coli*". *Scientific reports*, 6.(33037). Pp. 1-11.

Brown, J. K., Fouquet, C. M., Briddon, R. W., Zerbini, M., Moriones, E and NavasCastillo, J. 2012. "Family Geminiviridae. In: Virus Taxonomy: Classification and Nomenclature of Viruses—Ninth Report of the International Committee on Taxonomy of Viruses (King, A.M.Q., Adams, M.J., Carstens, E.B. and Lefkowitz, E.J., eds) Pp. 351-373.

Burgyán, J and Havelda, Z. 2011. "Viral suppressors of RNA silencing". *Trends in Plant Science*, 16(5). Pp. 265-272.

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". *Proceedings of the National Academy of Sciences of the United States of America*, 99 (16). Pp. 10310-10315.

Castel, S. E and Martienssen, R. A. 2013. "RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nature reviews". *Genetics*, 14(2). Pp. 100-112.

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. Pp. 381-394.

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., Sherman, A. *et al*. 2016. "Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology". *Molecular Plant Pathology*, 17. Pp. 1140-1153.

Chang, Z and Wang, A. 2016. "The *Potyvirus* Silencing Suppressor Protein VPg Mediates Degradation of SGS3 via Ubiquitination and Autophagy Pathways". *Journal of Virology*, 91 (1). Pp. 478-16.

Charvat, T. T., Lee, D. J., Robinson, W. E and Chamberlin, A. R. 2006. “Design, synthesis, and biological evaluation of chicoric acid analogs as inhibitors of HIV-1 integrase”. *Bioorganic and Medicinal Chemistry*, 14. Pp. 4552-4567.

Chatterji, A., Chatterji, U., Beachy, R. N and Fauquet, C. M. 2000. “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. Pp. 341–350.

Chavarriaga-Aguirre, P., Brand, A., Medina, A., Prías, M., Escobar, R., Martinez, J and Tohme, J. 2016. “The potential of using biotechnology to improve cassava: a review. In vitro cellular and developmental biology”. *Plant: journal of the Tissue Culture Association*, 52(5). Pp. 461-478.

Chen, G., F., T and Inouye, M. 1994. “Role of the AGG codons, the rarest codons in global gene expression in Escherichia coli”. *Genes and Development*, 8(21). Pp. 2641-2652.

Chowdhury, J., Chakraborti, S. T. P., Pramanik, K and Chakraborti, S. 2016. “Protective role of epigallocatechin-3-gallate in health and disease: a perspective,” *Biomedicine and Pharmacotherapy*, 78. Pp. 50-59.

Csorba, T., Pantaleo, V and Burgyan, J. 2009. “RNA Silencing: An Antiviral Mechanism”. *Advances in Virus Research*, 75. Pp. 35-71.

Daniell, H., Wurdack K. J., Kanagaraj, A., Lee, S. B., Saski, C and Jansen R. K. 2008. “The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and the evolution of atpF in Malpighiales: RNA editing and multiple losses of a group II intron”. *Theory Applied Genetics*, 116. Pp. 723-737.

De Groot, N. S and Ventura, S. 2006. “Effect of temperature on protein quality in bacterial inclusion bodies”. *FEBS Letters*, 580(27). Pp. 6471-6476.

Department of Agriculture. 1997. “Food security policy for South Africa”. http://faolex.fao.org/docs/pdf/saf149624.pdf.

Dong, O. X and Ronald, P. C. 2019. “Genetic Engineering for Disease Resistance in Plants: Recent Progress and Future Perspectives”. *Plant physiology*, 180(1). Pp. 26-38.

Douglas, J.T., Wu, X.Q., Augustin, G.P and Marang, M.G. 1988. “Evaluation of inexpensive blocking agents for ELISA in the detection of antibody in leprosy”. *Leprosy Review*, 59. Pp. 37-43.

Du, G. J., Zhang, Z., Wen, X. D., Yu, C., Calway, T., Yuan, C. S and Wang, C. Z. 2012. "Epigallocatechin Gallate (EGCG) is the most effective cancer chemopreventive polyphenol in green tea". *Nutrients*, 4(11). Pp. 1679-1691.

Duong-Ly, K. C and Gabelli, S. B. 2015. "Affinity Purification of a Recombinant Protein Expressed as a Fusion with the Maltose-Binding Protein (MBP) Tag". *Methods in Enzymology*, 559, Pp. 17-26.

EL-SHARKAWY, M. A. 2007. "Physiological characteristics of cassava tolerance to prolonged drought in the tropics: implications for breeding cultivars adapted to seasonally dry and semiarid environments. *Brazilian journal of plant physiology*, 19 (4). Pp. 257-286.

Eskandarian Z, Fliegauf M, Bulashevska A, Proietti M, Hague R, Smulski CR, Schubert D, Warnatz K and Grimbacher B. 2019. "Assessing the functional relevance of variants in the IKAROS family zinc finger protein 1 (IKZF1) in a cohort of patients with primary immunodeficiency". *Frontiers in Immunology*, 10(568). Pp. 1-18.

Fakruddin, M., Mohammad Mazumdar, R., Bin Mannan, K. S., Chowdhury, A and Hossain, M. N. 2013. "Critical Factors Affecting the Success of Cloning, Expression, and Mass Production of Enzymes by Recombinant *E. coli*". *ISRN Biotechnology*, Pp. 1-7.

FAOSTAT.2016.FAOSTATDatabase.Availableat:http://faostat3.fao.org/faostatgateway/go/to/download/Q/QC/E.

Fazaeli, A., Golestani, A., Lakzaei, M., Rasi Varaei, S. S and Aminian, M. 2019. "Expression optimization, purification, and functional characterization of cholesterol oxidase from Chromobacterium sp". *PLOS ONE*, 14(2). Pp. 212-217.

Florentino, L. H., Santos, A. A., Fontenelle, M. R., Pinheiro, G. L., Zerbini, F. M., Baracat-Pereira, M. C. and Fontes, E. P. B. (2006). "A PERK-Like Receptor Kinase Interacts with the *Geminivirus* Nuclear Shuttle Protein and Potentiates Viral Infection". *Journal of Virology*, 80(13). Pp. 6648-6656.

Fondong, V. N. 2013. "*Geminivirus* protein structure and function". *Molecular Plant Pathology*, 14 (6). Pp. 635-649.

Fondong, V. N. 2017. "The Search for Resistance to Cassava Mosaic Geminiviruses: How Much We Have Accomplished, and What Lies Ahead". *Frontiers in Plant Science*, 8(408). Pp. 1-19.

Futatsumori-Sugai, M., Abe, R., Watanabe, M., Kudou, M., Yamamoto, T., Ejima, D and Tsumoto, K. 2009. "Utilization of Arg-elution method for FLAG-tag based chromatography". *Protein Expression and Purification*, 67(2). Pp. 148-155.

Garner, M. M and Revzin, A. 1981. "A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system". *Nucleic Acids Research*, 9(13). Pp. 3047-3060.

Glick, B. R and Whitney, G. K. 1987. "Factors affecting the expression of foreign proteins in *Escherichia coli*," *Journal of Industrial Microbiology*, 1(5). Pp. 277-282.

Glynou, K., Ioannou, P.C and Christopoulos, T.K. 2003. "One-step purification and refolding of recombinant photoprotein aequorin by immobilised metal-ion affinity chromatography". *Protein Expression and Purification*, 27(2). Pp. 384-390.

Gomez, M. A., Lin, Z. D., Moll, T., Chauhan, R. D., Hayden, L., Renninger, K., Beyene, G., Taylor, N. J., Carrington, J. C., Staskawicz, B. J and Bart, R. S. 2019. "Simultaneous CRISPR/Cas9-mediated editing of cassava eIF4E isoforms nCBP-1 and nCBP-2 reduces cassava brown streak disease symptom severity and incidence". *Plant Biotechnology Journal*, 17. Pp. 421-434.

Han, M., Yagura, M and Itoh, T. 2007. "Specific Interaction between the Initiator Protein (Rep) and Origin of Plasmid ColE2-P9". *Journal of Bacteriology*, 189 (3). Pp. 1061-1071.

Han, S. O., Inui, M. and Yukawa, H. 2008. "Effect of carbon source availability and growth phase on the expression of *Corynebacterium glutamicum* genes involved in the tricarboxylic acid cycle and glyoxylate bypass". *Microbiology*, 154(10). Pp. 3073-3083.

Hanley-Bowdoin, L., Bejarano, E.R., Robertson, D and Mansoor, S. 2013. "Geminiviruses: Masters at redirecting and reprogramming plant processes". *Nature Reviews Microbiology*, 11. Pp. 777-778.

Harper, S and Speicher, D. W. 2011. "Purification of proteins fused to glutathione S-transferase". Methods in Molecular Biology (Clifton, N.J.), 681. Pp. 259-280.

Harrison, B. B., Swanson, M. M and Farfette, D. 2002. "*Begomovirus* coat protein: Serology, viariation and functions" *Physiological and Molecular Plant Pathology*, 60. Pp. 257-271.

Hartitz, M. D. Santa, G. and Bisaro D. M. 1999. "The *Tomato Golden Mosaic Virus* Transactivator (TrAP) is a single-stranded DNA and a Zinc-binding phosphoprotein with an Acidic Activation Domain". *Virology*, 263. Pp. 1-14.

Heyraud-Nitschke, F., Schumacher, S., Laufs, J., Schaefer, S., Schell, J and Gronenborn, B. 1995. "Determination of the origin cleavage and joining domain of *Geminivirus* Rep proteins". *Nucleic acids research*, 23(6). Pp. 910–916.

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). Pp. 189-198.

Hsieh, Y. W., Alqadah and Chuang, C. F. 2016. "An Optimized Protocol for Electrophoretic Mobility Shift Assay Using Infrared Fluorescent Dye-labeled Oligonucleotides". *Journal of Visualized Experiments*, 117(54863). Pp 1-6.

Huang, W. Y., Cai, Y. Z and Zhang, Y. 2009. "Natural Phenolic Compounds from Medicinal Herbs and Dietary Plants: Potential Use for Cancer Prevention". *Nutrition and Cancer*, 62(1) Pp. 1-20.

Hull R. 2014. "Replication of Plant Viruses". *Plant Virology*. Pp. 341-421.

International Committee on Taxonomy of Viruses (ICTV). Available online: https://talk.ictvonline.org/taxonomy/

Islam, M. A. 2012. "Cardiovascular effects of green tea catechins: Progress and promise," *Recent Patents on Cardiovascular Drug Discovery*, 7(2). Pp. 88-99.

Jaganathan, D., Ramasamy, K., Sellamuthu, G., Jayabalan, S and Venkataraman, G. 2018. "CRISPR for Crop Improvement: An Update Review". *Frontiers in Plant Science*, 9(985). Pp. 1-17.

Jia, B. and Jeon, C. O. 2016. "High-throughput recombinant protein expression in *Escherichia coli*: Current status and future perspectives". *Open Biology*, 6(8). Pp. 160-196.

Karamyshev, A L., Karamysheva, Z.N., Ito, K., Matsufuji S and Nakamura Y. 1999. "Overexpression and purification of recombinant eRF1 proteins of rabbit and Tetrahymena thermophile". *Biochemistry* (Mosc), 64(12). Pp. 1391-1400.

Kim, D., Song, S., Lee, M., Go, H., Shin, E., Yeom, J. H and Kim, Y. H. 2014. “Modulation of RNase E Activity by Alternative RNA Binding Sites”. *PLOS ONE,* 9(3). Pp. 1-11.

Kittelmann, K., Rau, P., Gronenborn, B and Jeske, H. 2009. “Plant *Geminivirus* Rep protein induces rereplication in fission yeast”. *Journal of Virology,* 83(13), Pp.6769-6778.

Kozlov, A. G., Galletto, R and Lohman, T. M. 2012. “SSB-DNA binding monitored by fluorescence intensity and anisotropy”. *Methods in molecular biology* (Clifton, N.J.), 922. Pp. 55-83.

Kram, K. E and Finkel, S. E. 2015. “Rich Medium Composition Affects *Escherichia coli* Survival, Glycation, and Mutation Frequency during Long-Term Batch Culture”. *Applied and Environmental Microbiology,* 81(13). Pp. 4442-4450.

Krämer, H., Amouyal, M., Nordheim, A and Müller-Hill, B. 1988. “DNA supercoiling changes the spacing requirement of two lac operators for DNA loop formation with lac repressor”. *EMBO Journal,* 7. Pp. 547-556.

Kuban-Jankowska, A., Sahu, K. K., Gorska, M., Tuszynski, J. A and Wozniak, M. 2016. “Chicoric acid binds to two sites and decreases the activity of the YopH bacterial virulence factor”. *Oncotarget,* 7(3). Pp. 2229-2238.

Kulbat, K. 2016. “The role of phenolic compounds in plant resistance”. *Biotechnology and Food Sciences,* 80 (2). Pp. 97-108.

Laemmli, U. K. 1970. “Cleavage of structural proteins during the assembly of the head of bacteriophage T4”. *Nature,* 227. Pp. 680-685.

Lakowicz, J. R. 1983. “Principles of Fluorescence Spectroscopy”. Plenum Press, New York.

Laufs, J. Desbiez, C., David, C., Mettouchi A., and Gronenborn, B .1995. “Rep protein of *tomato yellow leaf curl Geminivirus* has an ATPase activity required for viral DNA replication*”. Proceedings of the National Academy of Sciences,* 92. Pp. 5640-5644.

Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S. G., Schell, J and Gronenborn, B. 1995. “In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of *tomato yellow leaf curl virus*”. *Proceedings of the National Academy of Sciences of the United States of America,* 92(9). Pp. 3879-3883.

Lee C. H. 2017. A Simple Outline of Methods for Protein Isolation and Purification. *Endocrinology and metabolism* (Seoul, Korea), 32(1). Pp. 18-22.

Lee, P. Y., Costumbrado, J., Hsu, C. Y and Kim, Y. H. 2012. "Agarose gel electrophoresis for the separation of DNA fragments". *Journal of Visualized Experiments*: JoVE, 62(3923). Pp. 1-5.

Lentz, E. M., Kuon, J. E., Alder, A., Mangel, N., Zainuddin, I. M., McCallum, E. J and Vanderschuren, H. 2018. "Cassava *Geminivirus* agroclones for virus-induced gene silencing in cassava leaves and roots". *Plant Methods,* 14(1). Pp. 1-9.

Li W, Teng F, Li T and Zhou Q. 2013. "Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems". *Nature Biotechnology*, 31. Pp. 684-686.

Li, J. F., Norville, J. E., Aach, J., McCormack, M., Zhang, D., Bush, J and Sheen, J. 2013. "Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9". *Nature Biotechnology*, 31(8). Pp. 688-691.

Li, X., Tang, S., Wang, Q. Q., Leung, E. L. H., Jin, H., Huang, Y and Ding, J. 2017. "Identification of Epigallocatechin- Gallate as an Inhibitor of Phosphoglycerate Mutase 1". *Frontiers in Pharmacology*, 8(325). Pp. 1-9.

Long, G., Ji, Y., Pan, H., Sun, Z., Li, Y and Qin, G. 2015. "Characterization of Thermal Denaturation Structure and Morphology of Soy Glycinin by FTIR and SEM". *International Journal of Food Properties*, 18. Pp.763-774.

Lozano-Terol, G., Gallego-Jara, J., Sola Martínez, R. A., Cánovas Díaz, M and de Diego Puente, T. 2019. "Engineering protein production by rationally choosing a carbon and nitrogen source using *E. coli* BL21 acetate metabolism knockout strains". *Microbial Cell Factories*, 18(151). Pp. 1-19.

Luna, A.P., Rodríguez-Negrete, E.A., Morilla, G., Wang, L., Lozano-Duran, R., Castillo, A.G. and Bejarano, E.R. 2017. "V2 from a *Curtovirus* is a suppressor of post-transcriptional gene silencing", *Journal of General Virology*, 98. Pp. 2607-2614.

Martínez-Turiño, S and Hernández, C. 2009. “Inhibition of RNA silencing by the coat protein of *Pelargonium flower break virus*: distinctions from closely related suppressors”. *Journal of General Virology*, 90(2). Pp. 519-525.

Mauro, V. P and Chappell, S., A. 2014. “A critical analysis of codon optimization in human therapeutics”. *Trends in Molecular Medicine*, 20(11). Pp. 604-613.

Mehta, D., Stürchler, A., Anjanappa, R. B., Zaidi, S. S. A., Hirsch-Hoffmann, M., Gruissem, W and Vanderschuren, H. 2019. “Linking CRISPR-Cas9 interference in cassava to the evolution of editing-resistant Geminiviruses”. *Genome Biology*, 20(80). Pp. 1-10.

Miller, M., Bourassa, M. W and Smith, R.J. 2013. “FITR spectroscopic imaging of protein aggregation in living cells”. *Biochemica et Biophysica Act*, 1828. Pp. 2339-2346.

Mohammadian A., Kaghazian, H., Kavianpour, A and Jalalirad, R. 2018. “Solubilization of inclusion body proteins using low and very low concentrations of chemicals: implications of novel combined chemical treatment designs in enhancement of post-solubilization target protein purity and biological activity”. *Journal of Chemical Technology and Biotechnology*, 93(6). Pp. 1579-1587.

Morris, B., Richardson, K., Eddy, P., Zhan, X., Haley, A and Gardner, R. 1991. “Mutagenesis of the AC3 open reading frame of *African cassava mosaic virus* DNA A reduces DNA B replication and ameliorates disease symptoms”. *Journal of General Virology*, 72(6). Pp. 1205-1213.

Musa, S.D., Ifatimehin, O.O and Adeyemi, J.O. 2012. “Climate variability and Malaria incidence in Lokoja, Kogi State”. *Journal of Geography, Environment and Planning*, 8(2). Pp. 126-133.

Nielsen, M. M., Andersen, K. K., Westh, P and Otzen, D. E. 2007. “Unfolding of beta-sheet protein in SDS. *Biophysical Journal*, 92 (10). 3674-3685.

Nishimura H and Satoh A. 2006. “Antimicrobial and nematicidal substances from the root of chicory (*Cichorium intybus*). *Allelochemicals Biological Control of Plant Pathogen and Diseases*. 2. Pp. 177-180.

Njengele, Z., Kleynhans, R., Sayed, Y and Mosebi, S. 2016. “Expression, purification and characterization of a full-length recombinant HIV-1 Vpu from inclusion bodies”. *Protein Expression and Purification*, 128. Pp. 109-114.

Oliveira, A. F. C., Teixeira, R. R., Oliveira, A. S., Souza, A. P. M., Silva, M. L and Paula, S. O. 2017. "Potential Antivirals: Natural Products Targeting Replication Enzymes of *Dengue* and *Chikungunya* Viruses". *Molecules*, 22(3). Pp. 1-20.

Orozco P. 2018. "Argentina and Brazil Merge Law and Science to Regulate New Breeding Techniques". https://allianceforscience.cornell.edu/blog/2018/01/argentina-and-brazil-merge-law-and-science-to-regulate-new-breeding-techniques/

Orozco, B.M., Gladfelter, H.J., Settlage, S.B., Eagle, P.A., Gentry, R.N and Hanley-Bowdoin, L. 1998. "Multiple cis elements contribute to *Geminivirus* origin function". *Virology*, 242. Pp. 346-356.

Orozco, B.M., Kong, L.J., 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". *Journal of Biological Chemistry*, 275(9). Pp.6114-6122.

Orozco, B.M., Miller, A.B., Settlage, S.B. and Hanley-Bowdoin, L. 1997. "Functional Domains of a *Geminivirus* Replication Protein". *Journal of Biological Chemistry*, 272.

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 Science*, 4(11). Pp. 2411-2423.

Pace, N., C and Tanford C. 1968. "Thermodynamics of the unfolding of beta-lactoglobulin A in aqueous urea solutions between 5 and 55 degrees". *Biochemistry*, 7. Pp. 198–208.

Pain, R. H. 2004. "Determining the Fluorescence Spectrum of a Protein". *Current Protocols in Protein Science*, 7(7). Pp. 1-7.

Pant, V. Gupta, D. Choudhury, N. R. Malathi, V.G. Varma, A and Mukherje, S.K. 2001. "Molecular characterisation of the Rep protein of the blackgram isolate of *Indian mungbean yellow mosaic virus*". *Journal of General Virology*, 82. Pp. 2559-2567.

Pantaleo, V., Szittya, G and Burgyan, J. 2007. "Molecular Bases of Viral RNA Targeting by Viral Small Interfering RNA-Programmed RISC". *Journal of Virology*, 81(8). Pp. 3797-3806.

Park, S. R., Lim, C. Y., Kim, D. S and Ko, K. 2015. "Optimization of Ammonium Sulfate Concentration for Purification of Colorectal Cancer Vaccine Candidate Recombinant Protein GA733-FcK Isolated from Plants". *Frontiers in Plant Science*, 6 (1040). Pp. 1-8.

Pasumarthy, K. K., Choudhury, N. R and Mukherjee, S. K. 2010. “Tomato *leaf curl Kerala virus* (ToLCKeV) AC3 protein forms a higher order oligomer and enhances ATPase activity of replication initiator protein (Rep/AC1)”. *Virology Journal*. 7(128). Pp. 1-8.

Patzke, H and Schieber, A. 2018. “Growth-inhibitory activity of phenolic compounds applied in an emulsifiable concentrate - ferulic acid as a natural pesticide against *Botrytis cinerea*”. *Food Research International,* 113. Pp. 18-23.

Pauthe, E., Pelta, J., Patel, S., Lairez, D and Goubard, F. 2002. “Temperature-induced h-aggregation of fibronectin in aqueous solution.” *Biochimica et Biophysica Acta*, 1597. Pp. 12 - 21.

Perumal, A., Vishwakarma, R. K., Hu, Y., Morichaud, Z and Brodolin, K. 2018. “RbpA relaxes promoter selectivity of M. *tuberculosis* RNA polymerase”. *Nucleic Acids Research*, 46(19). Pp. 10106-10118.

Peter, B., Ngubane, M. C. M., Fanucchi, S and Dirr, H. W. 2013. “Membrane Mimetics Induce Helix Formation and Oligomerization of the Chloride Intracellular Channel Protein 1 Transmembrane Domain”. *Biochemistry*, 52. Pp. 2739-2749.

Priyadarshini, C. G., Ambika, M. V., Tippeswamy, R and Savithri, H. S. 2011. “Functional Characterization of Coat Protein and V2 Involved in Cell to Cell Movement of *Cotton Leaf Curl Kokhran Virus-Dabawali*”. *PLOS ONE,* 6(11). Pp. 1-12.

Prokoph, N., Ormö, M., O’Mahony, G., Hogner, A., McPheat, J., Karlsson, U and Liu, J. 2016. “Development of an ELISA for High-Throughput Screening of Inhibitors of Cdk5-Mediated PPARγ Phosphorylation”. *ASSAY and Drug Development Technologies*, 14(4). Pp. 261-272.

Ramesh, S., Sahu, P., Prasad, M., Praveen, S and Pappu, H. 2017. “Geminiviruses and Plant Hosts: A Closer Examination of the Molecular Arms Race”. *Viruses*, 9(9). Pp. 1-21.

Rey, C and Vanderschuren, H. 2017. “Cassava Mosaic and Brown Streak Diseases: Current Perspectives and Beyond”. *Annual Review of Virology*, 4(1). Pp. 429-452.

Reyes, L. F and Cisneros-Zevallos, L. 2003. “Wounding stress increases the phenolic content and antioxidant capacity of purple-flesh potatoes (*Solanum tuberosum* L.)”. *Journal of Agricultural Food Chemistry,* 51. Pp. 5296-5300.

Riek, U., Ramirez, S., Wallimann, T and Schlattner, U. 2009. “A versatile multidimensional protein purification system with full Internet remote control based on a standard HPLC system”. *BioTechniques*, 46(6S). Pp. 4-7.

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”. *Archives of Virology*, 160. Pp. 375-387.

Sakamoto, S., Putalun, W., Vimolmangkang, S., Phoolcharoen, W., Shoyama, Y., Tanaka, H and Morimoto, S. 2018. “Enzyme-linked immunosorbent assay for the quantitative/qualitative analysis of plant secondary metabolites”. *Journal of Natural Medicines*, 72(1). Pp. 32-42.

Sanford, J.C and Johnston, S.A. 1985. “The concept of parasite-derived resistance-deriving resistance genes from the parasite’s own genome”. *Journal of Theory in Biology*, 113. Pp. 395-405.

Saranraj, P., Behera, S. S and Ray, R. C. 2019. “Traditional Foods from Tropical Root and Tuber Crops”. *Innovations in Traditional Foods*, Pp. 159-191.

Saraswat, M., Musante, L., Ravidá, A., Shortt, B., Byrne, B and Holthofer, H. 2013. “Preparative Purification of Recombinant Proteins: Current Status and Future Trends”. *BioMed Research International*, Pp. 1-18.

Scarpati, M. L and Oriente, G. 1958. “Chicoric acid (dicaffeyltartic acid): its isolation from chicory (*Chicorium intybus*) and synthesis”. *Tetrahedron*, 4. Pp. 43-48.

Schein, C. H and Noteborn, M. H. M. 1988. “Formation of Soluble Recombinant Proteins in Escherichia Coli is favoured by Lower Growth Temperature”. *Nature Biotechnology*, 6(3). Pp. 291-294.

Schlager, B., Straessle, A and Hafen, E. 2012. “Use of anionic denaturing detergents to purify insoluble proteins after overexpression”. *BMC Biotechnology*, 12, (95). Pp. 1-7.

Silva, A. R. A., Morais, S. M., Marques, M. M. M., Lima, D. M., Santos, S. C. C., Almeida, R. R., Vieira, I. G. P and Guedes, M. I. F. 2011. “Antiviral activities of extracts and phenolic components of two Spondias species against *dengue* virus” *Journal of Venomous Animals and Toxins Including Tropical Diseases,* 17(4). Pp. 406-4013.

Sivashanmugam A, Murray V, Cui C, Zhang Y, Wang J and Li Q. 2009. “Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*”. *Protein Science*, 18(5). Pp. 936-48.

Snehi, S. K., Purvia, A. S., Parihar, S., Gupta, G., Singh and Raj, S. K. 2017. “Overview of *Begomovirus* genomic organization and its impact”, *International Journal of Current Research*, 9 (11). Pp. 61368-61380.

Steiner, S and Pfannschmidt, T. 2009. “Fluorescence-based Electrophoretic Mobility Shift Assay in the Analysis of DNA-binding Proteins”. *Methods in molecular biology* (Clifton, N.J.). Pp. 273-289.

Strange, R.N and Scott, P.R. 2005. “Plant disease: a threat to global food security”. *Annual Review of Phytopathology*, 43(1). Pp. 83-116.

Swartz, J. R. 2001. “Advances in Escherichia coli production of therapeutic proteins,” *Current Opinion in Biotechnology*, 12(2). Pp. 195-201.

Tabassum, B., Ahmad, I., Aslam, U and Husnai T, 2012. “How RNA interference combat” viruses in plants, in Functional Genomics, (MeroniGand Petrera F. InTech, Rijeka ed). Pp. 13-130.

Tajmir-Riahi, H. A., N’soukpoé-Kossi, C. N and Joly, D. 2009. “Structural analysis of protein–DNA and protein–RNA interactions by FTIR, UV-visible and CD spectroscopic methods”. *Spectroscopy*, 23(2). Pp. 81-101.

Tao, Q., Caoa, J., Zhua, L and Linb, H. 2019. “The complete mitochondrial genome of an important root crop cassava (*Manihot esculenta*)”. *Mitochondrial DNA Part B*, 4(1). Pp. 1081-1082.

Thomsen, M. O., Frette, X. C., Christensen, K. B., Christensen, L. P and Grevsen K. 2012. “Seasonal variations in the concentrations of lipophilic compounds and phenolic acids in the roots of *Echinacea purpurea* and *Echinacea pallida*”. *Journal of Agricultural and Food Chemistry*, 60. Pp. 12131-12141.

Tonukari, N.J. 2004. “Cassava and the future of starch”. *Electronic Journal of Biotechnology*, 7(1). Pp.5-8.

Torres-Herrera, S. L., Romero-Osorio, A., Moreno-Valenzuela, O., Pastor-Palacios., G., Carden as-Conejo, Y., Ramírez-Prado, J. H., Riego-Ruiz, L., Minero-García., Y., Ambriz-Granados., S and Argüello-Astorga, G. R. 2019. "A Lineage of Begomoviruses Encode Rep and AC4 Proteins of Enigmatic Ancestry: Hints on the Evolution of Geminiviruses in the New World Setting". *Viruses,* 11(7). Pp. 1-18.

Towbin, H., Staehelin, T and Gordon, J. 1979. "Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets. Procedure and some applications*". Proceedings of the National Academy of Sciences of the United States of America*, 76. Pp. 4350-4354.

Trinks, D., Rajeswaran, R., Shivaprasad, P.V., Akbergenov, R., Oakeley, E.J., Veluthambi, K., Hohn, T. and Pooggin, M. 2005. "Suppression of RNA silencing by a *Geminivirus* nuclear protein, AC2, correlates with transactivation of host genes". *Journal of Virology,* 79. Pp. 2517-2527.

United Nations. 2013. "The world population ageing. Department of Economic and Social Affairs,PopulationDivision" https://www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeing2013

United Nations. 2013. The millennium development goal. https://www.un.org/millenniumgoals/pdf/report-2013/mdg-report-2013-english.pdf

Upadhyay, S. K., Kumar, J., Alok, A and Tuli, R. 2013. "RNA-Guided Genome Editing for Target Gene Mutations in Wheat". *Genetics,* 3(12). Pp. 2233-2238.

Usoltsev D. A., Sitnikova V. E., Nosenko T. N., Olekhnovich R. O and Uspenskaya M.V. 2019. "Comparison of protein secondary structure calculation methods based on infrared spectra deconvolution". *Scientific and Technical Journal of Information Technologies, Mechanics and Optics,* 19(4). Pp. 586-593.

Usoltsev, Sitnikova, Kajava and Uspenskaya. 2019. "Systematic FTIR Spectroscopy Study of the Secondary Structure Changes in Human Serum Albumin under Various Denaturation Conditions". *Biomolecules,* 9(359). Pp. 1-17.

Van der Krol A. R., Mur, L. A., Beld, M., Mol, J. N and Stuitje, A. R. 1990. "Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression". *Plant Cell,* 2. Pp. 291-299.

Vance, V. 2001. “RNA Silencing in Plants-Defense and Counterdefense”. *Science*, 292(5525). Pp. 2277-2280.

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”. *Journal of Virology*, 78 (17). Pp. 9487-9498.

Wang, X. B., Jovel, J., Udomporn, P., Wang, Y., Wu, Q., Li, W.X., Gasciolli. V., Vaucheret. H and Ding. S., W. 2011. “The 21-nucleotide, but not 22-nucleotide, viral secondary small interfering RNAs direct potent antiviral defense by two cooperative argonautes in *Arabidopsis thaliana*”. *Plant Cell*, 23. Pp. 1625-1638.

Wegrzyn, K. Fuentes-Perez, M.E. and Bury, K. 2014. “Sequence-specific interactions of Rep proteins with ssDNA in the AT-rich region of the plasmid replication origin”. *Nucleic Acids Research*, 42(12). Pp. 7807-7818.

Widido, Y. 2018. “Cassava Productivity for Eradicating Hunger and Poverty in Rural Areas of Indonesia”. *Scindo*, 39(334). Pp. 32-40.

Wills, R. B. H and Stuart D. L. 1999. “Alkylamide and chicoric acid in *Echinacea purpurea* grown in Australia”. *Food Chemistry*, 67. Pp. 385-388

Xiao, Y and Isaacs, S. N. 2012. “Enzyme-linked immunosorbent assay (ELISA) and blocking with bovine serum albumin (BSA) not all BSAs are alike”. *Journal of immunological methods*, 384(1-2). Pp. 148-151.

Xing, H. L., Dong, L., Wang, Z. P., Zhang, H. Y., Han, C. Y., Liu, B and Chen, Q. J. 2014. “A CRISPR/Cas9 toolkit for multiplex genome editing in plants”. *BMC Plant Biology*, 14(1). Pp. 1-12.

Yakhnin, A. V., Yakhnin, H and Babitzke, P. 2012. “Gel mobility shift assays to detect protein-RNA interactions”. *Methods in Molecular Biology*, 905. Pp. 201-211.

Yakubov, B., Chen, L., Belkin, A. M., Zhang, S., Chelladurai, B., Zhang, Z. Y and Matei, D. 2014. “Small Molecule Inhibitors Target the Tissue Transglutaminase and Fibronectin Interaction”. *PLOS ONE*, 9(2). Pp. 1-8.

Yang, H., Yang, S., Kong, J., Dong, A. and Yu, S. 2015. “Obtaining information about protein secondary structures in aqueous solution using Fourier-transform IR spectroscopy”. *Nature Protocols*, 10(3). Pp. 382-396.

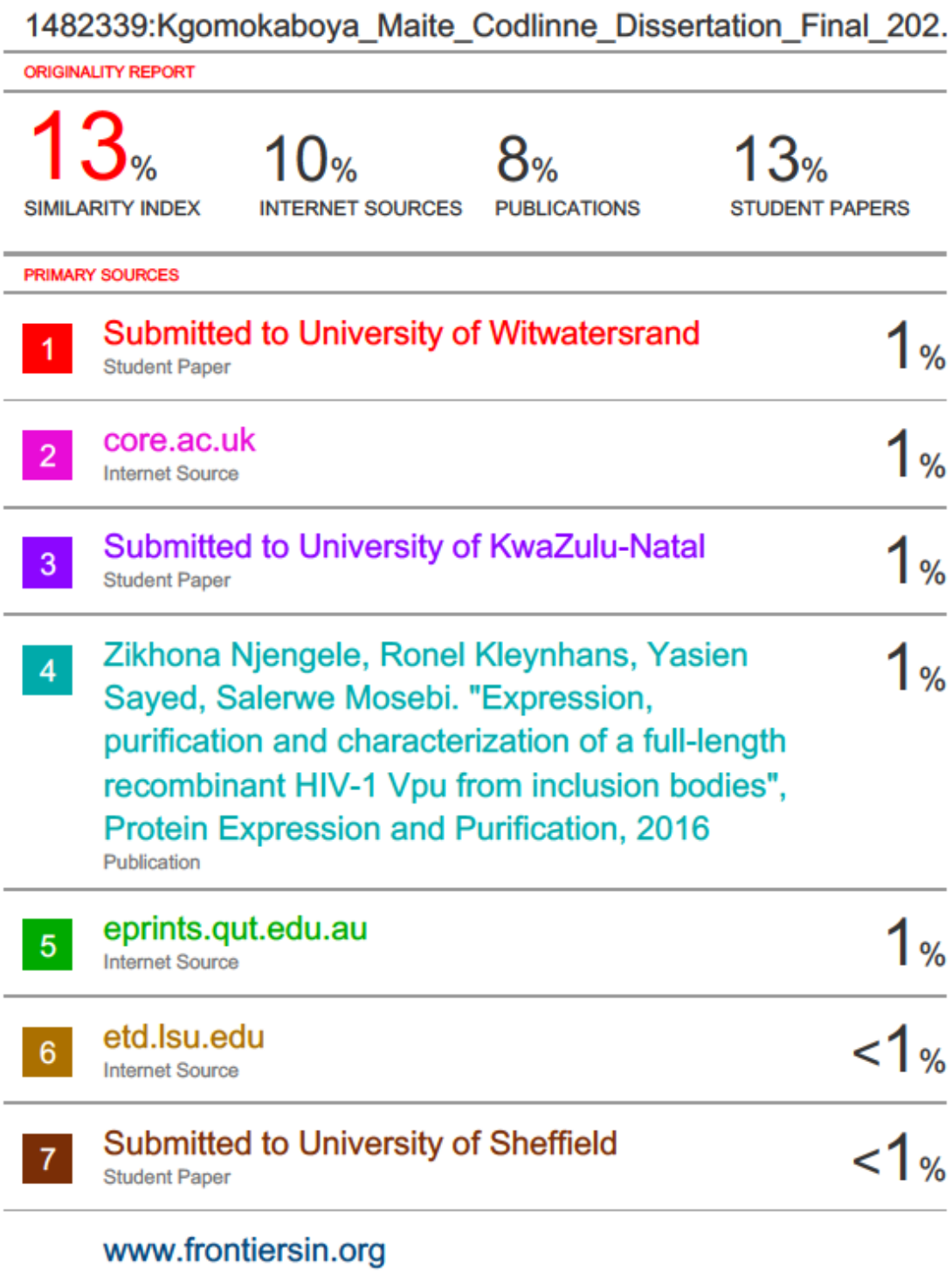
Ye, J., Qu, J., Mao, H. Z., Ma, Z. G., Rahman, N. E., Bai, C., Chen, W., Jiang, S. Y., Ramachandran, Sand Chua, N. H. 2014. "Engineering *Geminivirus* resistance in *Jatropha curcus*". *Biotechnology for Biofuels*, 7(1). Pp. 1-11.

Zerbini, F. M., Briddon, R.W., Idris, A., Martin, D. P., Moriones, E., Navas-Castillo, J., Rivera-Bustamante., R, Roumagnac, R., Varsani, A and ICTV Report Consortium. 2017. "ICTV Virus Taxonomy Profile: Geminiviridae". *Journal of General Virology*, 98. Pp. 131-133.

Zininga, T., Achilonu, I., Hoppe, H., Prinsloo, E., Dirr, H. W and Shonhai, A. 2015. "Overexpression, Purification and Characterisation of the *Plasmodium falciparum* Hsp70-z (PfHsp70-z) Protein". *PLOS ONE*, 10(6). Pp. 1-13.

Zotti, M., dos Santos, E. A., Cagliari, D., Christiaens, O., Taning, C. N. T. and Smagghe, G. 2018. "RNA interference technology in crop protection against arthropod pests, pathogens and nematodes". *Pest Management Science*, 74(6), Pp. 1239-1250.

**Appendix**



1482339:Kgomokaboya_Maite_Codlinne_Dissertation_Final_202.

ORIGINALITY REPORT

| 13% | 10% | 8% | 13% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | Submitted to University of Witwatersrand<br>Student Paper | 1% |
| 2 | core.ac.uk<br>Internet Source | 1% |
| 3 | Submitted to University of KwaZulu-Natal<br>Student Paper | 1% |
| 4 | Zikhona Njengele, Ronel Kleynhans, Yasien Sayed, Salerwe Mosebi. "Expression, purification and characterization of a full-length recombinant HIV-1 Vpu from inclusion bodies", Protein Expression and Purification, 2016<br>Publication | 1% |
| 5 | eprints.qut.edu.au<br>Internet Source | 1% |
| 6 | etd.lsu.edu<br>Internet Source | <1% |
| 7 | Submitted to University of Sheffield<br>Student Paper | <1% |
| | www.frontiersin.org | |