

https://dspace.org/

South African cassava mosaic virus movement and nuclear shuttle proteins: uncovering their structures

Nankoo, Nikita

2022

https://hdl.handle.net/10539/34574

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



South African cassava mosaic virus movement and nuclear shuttle proteins: uncovering their structures

by

Nikita Nankoo

(389088)

PhD Thesis

Submitted in fulfilment of the requirements for the degree

Doctor of Philosophy

in

Molecular and Cell Biology

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

Supervisor: Prof MEC Rey

Co-supervisor: Dr IA Achilonu

June 2022

Declaration

I, Nikita Nankoo (389088), am a student registered for the post graduate degree Doctor of Philosophy in science (PhD) in the academic year 2022.

I hereby declare the following:

• I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.

• I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise.

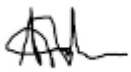
• I have not submitted this work before for any other degree or examination at any other
University.

• I have followed the required conventions in referencing the thoughts and ideas of others.

• I understand that the University of the Witwatersrand may take disciplinary action against me if there is 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.

==========

Nikita Nankoo (Miss)



1st day of June 2022

Acknowledgements

My deepest gratitude goes to my supervisor Prof Chrissie Rey for her valuable support and mentorship without whom this research would have not been possible.

I thank my co-supervisor Dr I. A. Achilonu and the University of Witwatersrand Protein Structure, Function and Research Unit for valuable input, support, and equipment usage.

I acknowledge the National Research Foundation and Prof Chrissie Rey for financial support during the course of this research.

Special thanks must go to Mr Ruan van Deventer, Mr Dylan Moodley, Miss Sasha Richardson, Mr Banele Ndlovu, Miss Alison Brigden, Mr Alexander Zwolinski, Dr Angela Botes and Mr Michael Tobin for their continued support and assistance during the course of this research.

I thank the University of Witwatersrand Plant Biotechnology Research members for continued support, guidance, and advice.

Lastly, I thank God and mostly importantly my loved ones for their support, understanding, and for the continued courage and strength to see this research to completion.

Contents

Declaration ii
Acknowledgements iii
List of tables vii
List of figures viii
List of abbreviations xii
Abstract 1
Chapter 1- Introduction 3
1.1 Overview of Cassava Mosaic Disease 3
1.2 Problem Statement 4
1.3 Rationale of the study 5
1.4 Novelty of the study 6
1.5 Aims and Objectives 6
1.6 Structure of the thesis 7
Chapter 2- Literature review 8
2.1 Cassava 8
2.1.1 Cassava production 9
2.1.2. Cassava products 12
2.1.3. Threats affecting cassava production 12
2.2 Cassava mosaic disease (CMD) 13
2.2.1. Symptoms of CMD 15
2.3. Geminiviruses 17
2.3.1. Structure of geminiviruses 19
2.3.2. Replication of geminiviruses 19
2.4 Cell-to-cell movement of viruses 22
2.4.1 *Geminivirus* movement proteins 23
2.4.2 Cell-to-cell MP and NSP- host interactions 27
2.5. Protein expression systems 28
2.6. Protein expression and purification 29
2.7. Structural protein characterization 31
Chapter 3 34
Computational analysis of the MP and NSP of SACMV 34
3.1. Background 34
3.2. Materials and Methods 37
3.2.1. Truncation of MP and NSP from full length proteins 37

3.2.2. Determining familial matches of the truncated MP and NSP proteins using InterProScan38
3.2.3. Secondary structure predictive analysis 38
3.2.4. Tertiary structure, binding sites and disordered region analysis..... 39
3.2.5. Building the 3-D models of truncated MP and NSP 39
3.3. Results 40
3.3.1. Truncation of MP and NSP from full length protein 40
3.3.2. Determining familial matches of the truncated MP and NSP proteins using InterProScan40
3.3.3. Secondary structure predictive analysis 41
3.3.4. Tertiary structure, binding sites and disordered region analysis..... 44
3.3.5. Building the 3-D models of truncated MP and NSP 47
3.4. Discussion 50
3.5. Conclusions 53
Chapter 4..... 54
Expression and Purification of truncated SACMV MP and NSP 54
4.1. Background 54
4.2. Materials and Methods 56
4.2.1. Cloning of the MP and NSP genes into the pCOLDI vector 56
4.2.2. Transformation of MP and NSP gene inserts into competent *E.coli* cells and gene insert validation 57
4.2.3. Protein overexpression 59
4.2.4. Protein extraction 61
4.2.5 Protein denaturation 61
4.2.6. Protein purification 61
4.2.7. Protein refolding 62
4.3. Results 62
4.3.1. Cloning of the MP and NSP genes into the pCOLDI vector 62
4.3.2. Transformation of MP and NSP gene inserts into competent *E.coli* competent cells and gene insert validation 63
4.3.3. Protein overexpression 65
4.3.4. Protein purification 67
4.3.5. Protein refolding 69
4.4. Discussion 72
4.5. Conclusions 76
Chapter 5..... 77
Secondary Structural characterization of truncated Movement Protein and Nuclear Shuttle Protein..... 77

5.1. Introduction 77
5.2. Materials and Methods 79
5.2.1. Protein quantification 79
5.2.2. Far-UV Circular Dichroism 79
5.2.3. Intrinsic tryptophan fluorescence 80
5.2.4. Extrinsic ANS fluorescence binding assay 81
5.2.5. Intrinsic Fluorescence MANT-ATP binding Assay 81
5.2.6. Phosphate assay for determination of functionality 82
5.2.7. Size exclusion high performance liquid chromatography (SE-HPLC) 82
5.3. Results 83
5.3.1. Protein quantification 84
5.3.2. Far-UV Circular Dichroism 85
5.3.3. Intrinsic tryptophan fluorescence 87
5.3.4. Extrinsic ANS fluorescence binding assay 89
5.3.5. Intrinsic Fluorescence MANT-ATP binding Assay 90
5.3.6. Phosphatase assay 92
5.3.7. Size exclusion high performance liquid chromatography (SE-HPLC) 94
5.4. Discussion 97
5.5. Conclusions 101
Chapter 6 102
Conclusions 102
Appendix A 104
Appendix B 109
References 110

List of tables

Table 2.1: Viruses affecting cassava in Latin America, Africa and South Asia……13

Table 2.2: Symptoms caused by CMGs which cause CMD…… 15

Table 2.3: Economically important viral diseases caused by geminiviruses……..17

Table 4.1: Table showing reagents used for gene validation of the MP truncated…59

Table 4.2: Table showing reagents used for gene validation of the NSP truncated…59

Table 5.1: Secondary structure content of the truncated MP as calculated using….86

Table 5.2: Secondary structure content of the truncated NSP as calculated using…87

List of figures

Figure 2.1: The most produced commodities in Africa…….8

Figure. 2.2: Cassava leaf and roots……8

Figure 2.3: The amount of cassava harvested and the production of cassava……...9

Figure 2.4: The amount of cassava produced (in tonnes) from the top 10 cassava producers worldwide……..9

Figure 2.5.: Comparison between cassava production and the area harvested for cassava in Africa……..10

Figure 2.6: Indication of the production share of cassava by region……..11

Figure 2.7: Symptoms of CMD on cassava plants…….15

Figure 2.8: The whitefly which transmits geminiviruses…….17

Figure 2.9: Diagram of a bipartite begomovirus……..20

Figure 2.10: The replication of begomoviruses as it begins in a plant cell…….20

Figure 2.11: Amino acid sequence of SACMV MP……24

Figure 2.12: The geminivirus MP and its predicted domains…….24

Figure 2.13: The geminivirus NSP and its predicted domains…….25

Figure 2.14: Amino acid sequence of SACMV NSP……..25

Figure 2.15: Schematic representation of recombinant protein expression ……..29

Figure 2.16: Schematic of an *E.coli* expression vector……..30

Figure 2.17. Chemical structure of ANS……...32

Figure 3.1: Amino acid sequence of full-length SACMV BC1……..36

Figure 3.2: Amino acid sequence of full-length SACMV NSP……..37

Figure 3.3: The truncated region of SACMV MP……39

Figure 3.4: The truncated region of SACMV NSP…….39

Figure 3.5: Annotation of the family homology of the truncated MP……...40

Figure 3.6: Annotation of the family homology of the truncated NSP…….40

Figure 3.7: The predicted secondary structure composition of the truncated MP of SACMV from the Predict Protein webserver……41

Figure 3.8: The predicted secondary structure composition of the truncated NSP of SACMV from the Predict Protein webserver……41

Figure 3.9: An annotation grid of the truncated MP of SACMV from the Psipred database……42

Figure 3.10: An annotation grid of the truncated NSP of SACMV from the Psipred database……42

Figure 3.11: The predicted solvent accessibility of the truncated MP of SACMV from the Predict Protein webserver……43

Figure 3.12: The predicted solvent accessibility of the truncated NSP of SACMV from the Predict Protein webserver……44

Figure 3.13: The predicted binding sites of the truncated MP (protein, DNA and RNA) of SACMV as predicted by the Predict Protein webserver……..44

Figure 3.14: The predicted binding sites of the truncated NSP (protein, DNA and RNA) of SACMV as predicted by the Predict Protein webserver……..45

Figure 3.15: The predicted disordered regions of the truncated MP of SACMV as predicted by the Predict Protein webserver……..45

Figure 3.16: The predicted disordered regions of the truncated NSP of SACMV as predicted by the Predict Protein webserver……..46

Figure 3.17: Partial predicted model of truncated MP of SACMV……..47

Figure 3.18: Partial predicted model of truncated NSP of SACMV……..47

Figure 3.19: Predicted model of the truncated MP of SACMV as predicted by SWISS-MODEL……...48

Figure 3.20: Predicted model of the truncated NSP of SACMV as predicted by SWISS-MODEL……49

Figure 4.1: The 528 bp gene sequence of the SACMV MP N-terminus truncation..56

Figure 4.2: The 267 bp gene sequence of the SACMV NSP C-terminus truncation 56

Figure 4.3: pColdI vector map showing……57

Figure 4.4: Schematic diagram showing the heat shock transformation protocol….58

Figure 4.4: LB agar plates, showing transformed MP-pCOLDI and BL21 DE3 pLysS clones……..63

Figure 4.5: LB agar plates, showing transformed NSP-pCOLDI and BL21 DE3 *E.coli* clones……..64

Figure 4.6: A 1% agarose gel showing the restriction digests, confirming the presence of the cloned MP and NSP genes……65

Figure 4.7: A 12% SDS-PAGE gel, showing the expression profile of the truncated MP from un induced till a 1 mM IPTG induction profile…….66

Figure 4.8: A 12% SDS-PAGE gel, showing the expression profile of the truncated NSP from un induced till a 1 mM IPTG induction profile…….67

Figure 4.9: A 12% SDS-PAGE gel, showing the elution profile of the purification of the truncated MP…….68

Figure 4.10: A 12% SDS-PAGE gel, showing the elution profile of the purification of the truncated NSP…….69

Figure 4.11: A 12% SDS-PAGE gel, showing the refolded truncated MP of SACMV ……..70

Figure 4.12: A 12% SDS-PAGE gel, showing the refolded truncated NSP of SACMV……...71

Figure 5.1: Quantification of the truncated MP……84

Figure 5.2: Quantification of the truncated NSP…….85

Figure 5.3: Far-UV circular dichroism spectrum for the truncated MP……86

Figure 5.4: Far-UV circular dichroism spectra for the truncated NSP…….87

Figure 5.5: Intrinsic tryptophan fluorescence spectra of truncated MP……..88

Figure 5.6: Intrinsic tryptophan fluorescence spectra of truncated NSP…….89

Figure 5.7: Extrinsic ANS fluorescence spectra of truncated MP……89

Figure 5.8: Extrinsic ANS fluorescence spectra of truncated NSP…….90

Figure 5.9: Intrinsic fluorescence spectra of truncated MP……..91

Figure 5.10: Intrinsic fluorescence spectra of truncated NSP…….91

Figure 5.11: 96 well plate showing the colour changes for the phosphate standards and the MP sample……..92

Figure 5.12: Quantification of released phosphate from hydrolysed ATP by the truncated MP…….93

Figure 5.13: 96 well plate showing the colour changes for the phosphate standards and the NSP sample…….93

Figure 5.14: Quantification of released phosphate from hydrolysed ATP the by the truncated NSP……...94

Figure 5.15: SE-HPLC spectra of the native truncated MP……..95

Figure 5.16: SE-HPLC spectra of the native truncated NSP……96

Figure 5.17: Standard curve used for weight determination of truncated MP and NSP…….96

List of abbreviations

Aa - Amino acid

ACMV - African cassava mosaic virus

ACMBFV - African cassava mosaic Burkina Faso virus

ANS - 1-anilino-naphthalene-8-sulfonate

ATP - Adenosine triphosphate

BDMV - Bean dwarf mosaic virus

C-ter - C-terminus

CaLCUV - Cabbage leaf curl virus

CBSV - Cassava brown streak virus

CBVs - Cassava begomoviruses

CD - Circular dichroism

CFSD - Cassava frogskin disease

CMD - Cassava mosaic disease

CMG - Cassava mosaic geminivirus

CMMV - Cassava mosaic Madagascar virus

CP - Coat protein

CR - Common region

dsDNA - Double stranded DNA

DTT - Dithiothreitol

E.coli - Escherichia coli

EACMV - East Africa cassava mosaic virus

EACMMV - East African cassava mosaic Malawi virus

EACMZV - East African cassava mosaic Zanzibar virus

EACMKV - East African cassava mosaic Kenya virus

EDTA - Ethylenediaminetetraacetic acid

EK - Eukaryotic

ER - Endoplasmic reticulum

IMAC - Immobilized metal ion affinity chromatography

IPTG - Isopropyl-beta-D-thiogalactopyranoside

IR - Intergenic region

MANT-ATP - 2’(3’)-N-methylanthraniloyl-ATP

MP - Movement protein

N-ter - N-terminus

NaCl - Sodium chloride

NCBI - National Centre for Biotechnology Information

Ni-NTA - Nickel-nitrilotriacetic acid

NSP - Nuclear shuttle protein

OD - Optimal density

ORF - Open reading frame

PD - Plasmodesmata

PDB - Protein Database

PK - Prokaryotic

PTGS - Post transcriptional gene silencing

REP- Replication-associated protein

SACMV - *South African cassava mosaic virus*

SDS-PAGE - Sodium dodecyl sulphate polyacrylamide gel electrophoresis

SE-HPLC - Size exclusion high performance liquid chromatography

siRNAs - Short interfering RNAs

SLCMV - Sri Lankan cassava mosaic virus

SqLCV - Squash leaf curl virus

ssDNA - Single stranded DNA

TMV - Tobacco mosaic virus

TrAP - Transcriptional activator protein

TuMV - Turnip mosaic virus

TYLCV - Tomato yellow leaf curl virus

UV - Ultraviolet

Abstract

South African cassava mosaic virus (SACMV) is a plant-infecting circular ssDNA bipartite begomovirus, whose genome comprises of DNA-A (encodes six genes) and DNA-B [encodes BC1 cell-to-cell movement (MP) and BV1 nuclear shuttle protein (NSP)]. Expression of these viral proteins *in vitro* and their purification have not been achieved to date and this study aimed to achieve this for a truncated region of the MP (N-ter) and NSP (C-ter), respectively. Furthermore, SACMV movement and nuclear shuttle proteins were truncated due to the difficulties in working with the full-length proteins; and the secondary structures of the truncated proteins were then investigated. Computational characterization and homology modelling were also undertaken. The truncated MP (amino acid 1-258) and NSP (amino acid 171-258) were cloned into a pCOLDI expression vector and transformed into BL21 (DE3) pLYSs and BL21 (DE3) *E. coli* respectively. Optimal conditions for the induced expression of the MP were found to be 0.50 mM IPTG at an OD_{600} of 0.45 expressed for 24 h. In contrast, the optimal conditions of the expression of the truncated NSP were found to be 0.25 mM IPTG at an OD_{600} of 0.40 expressed for 24 h. 8-Anilinonaphthalene-1-sulfonic acid (ANS) has been used as a folding/unfolding monitoring tool. This study showed that both proteins bind ANS with greater affinity in the denatured form, indicating that there are more hydrophobic regions for the polar fluorescent dye to bind. Both truncated proteins do not have the ability to bind ATP, and this was determined by using the MANT-ATP intrinsic fluorescence assay. In order to confirm that both proteins were refolded and did not bind ATP, intrinsic tryptophan fluorescence was assessed. This assay confirmed that both proteins were refolded due to the observed difference in spectra profiles. The assay also confirmed that ATP caused no conformational changes in both proteins in the native or denatured forms. Secondary structure analysis employing Circular Dichroism confirmed that the truncated NSP and MP contained predominantly β-sheet secondary structures. The estimated molecular weights by SDS-PAGE, were determined to be 23 kDa and 13kDa for MP and NSP, respectively. The employment of Mass Spectrometry in future studies could assist in confirming the exact molecular weight of the protein as Size Exclusion High Performance Liquid Chromatography did not yield desirable results as

the proteins eluted at the incorrect retention times possibly due to the proteins being expressed in the insoluble form.

Chapter 1- Introduction

1.1 Overview of Cassava Mosaic Disease

Cassava (*Manihot esculenta Crantz*) is a woody, tropical, perennial shrub. Cassava is considered to be one of the most important crops in a number of African and Asian countries due to its ability to provide food security and an income for disadvantaged farmers (Adekunle et al., 2016, Ariyo et al.,2006; Legg et al., 2015). One of the major threats impacting cassava crops is cassava mosaic disease (CMD), which is caused by cassava mosaic begomoviruses (CBVs) and is transmitted by the whitefly *Bemisia tabaci* (Fondong, 2017; Patil and Fauquet, 2009). Currently, CMGs are considered to be the most severe and widespread vector-borne plant viral pathogens affecting cassava crops. Cassava mosaic disease is a notable disease largely in Africa and Asia where cassava cultivation is of high economic importance (Alabi et al., 2011; Legg et al., 2006). CMD is endemic to sub-Saharan African countries and is caused by six distinct species of bipartite begomoviruses CBVs (Brown et al., 2015 Legg et al., 2015, Rey and Vanderschuren, 2017). These species are: *African cassava mosaic virus* (ACMV), *East African cassava mosaic virus* (EACMV), *East African cassava mosaic Kenya virus* (EACMKV), *East African cassava mosaic Malawi virus* (EACMMV), *East African cassava mosaic Zanzibar virus* (EACMZV) and *South African cassava mosaic virus* (SACMV) (Brown et al., 2015; Zerbini et al., 2017). There are a number of symptoms associated with CMD, and these vary in their type, severity and extent. CMD produces a range of foliar symptoms including green to yellow mosaic of the affected leaves, coupled with misshapen and twisted leaves. Other symptoms of CMD also include leaf distortion and mottling of leaves and can also lead to plants with reduced height. Cassava plants affected with CMD also produce no or fewer tubers depending on the severity of CMD infection (Alabi et al., 2011; Thresh and Cooter 2005). The effects of CMD on cassava crops greatly affects income and food security for many farmers in already poor regions of Africa and Asia. Tuber production of cassava plants is greatly affected due to CMD and this affects the various industrial processes and applications that are fast becoming dependant on cassava starch. These include the food and beverage industry, and animal feed and biofuel production industries (Alabi et al., 2011; Baguma et al., 2017; Morgan and Choct, 2016).

1.2 Problem Statement

Cassava is a popular crop amongst small-scale and subsistence farmers compared to similar crops such as wheat, rice and maize. This due to its ability to tolerate unfavourable conditions and abiotic stresses. Following the production of sugar cane and maize, cassava is the third most produced crop in Africa (FAOSTAT, 2021). When compared to other crops, cassava provides a higher carbohydrate content (Adekunle et al., 2016; Alabi et al., 2011). Recently the global production of cassava has vastly increased by an average of 100 million tonnes owing to the rising demands for cassava food products largely in Africa and for dried cassava and starch in Asia (Shakelford et al., 2018). It is estimated that approximately 500 million people depend on cassava as a source of carbohydrates, making cassava the third largest source of carbohydrates in the world for human food. Currently the world average yield of cassava is 12.8 tonnes per hectare. Cassava is fast becoming one of the most critical staple food sources for lower-income families in rural areas on the Sub-Saharan African continent. Cassava is also a primary crop for smallholder farmers. There is vast expansion of the cassava industry in Africa which is mainly driven by the constant increasing food demand for cassava and cassava products on the continent and globally (Morgan and Choct, 2016). More than 50% of the worlds' production of cassava occurs within Africa amongst cassava- growing regions of the world (Shakelford et al., 2018). Cassava has the ability to grow successfully under a wide range of agro-ecological zones where maize, rice and many other crops cannot thrive. Cassava is tolerant to diseases, drought, poor soils as well as other stresses (biotic and abiotic), making cassava an ideal crop for poor farmers to cultivate under marginal environments in Africa (Alabi et al., 2011; Morgan and Choct, 2016; Veiga et al., 2016). Further posing a threat to farmers and food security, causing great economic losses are plant viruses called Geminiviruses (Rizvi et al., 2014). Cassava mosaic geminiviruses (CMGs) are one of the largest impacting threats affecting cassava and are transmitted by the whitefly *Bemisia tabaci* (Fondong, 2017; Patil and Fauquet, 2009). Geminiviruses which belong to the genus, have circular single stranded DNA (ssDNA) and are monopartite (containing DNA-A only) or bipartite (containing two DNA components, DNA A and DNA B) genomes (Brown et al., 2015). South African cassava mosaic virus (SACMV), the virus of interest for this study contains a bipartite genome, where DNA A and B encode six and two proteins, respectively (Berrie et al., 2001). The DNA A genome encodes three proteins in the antisense direction while DNA B encodes for two

proteins, namely the cell-to-cell movement protein (MP) from the BC1 ORF on the complimentary strand, and the nuclear shuttle protein (NSP) from the BV1 ORF on the sense strand (Patil and Fauquet, 2015). There are two existing models proposed for the active roles of geminivirus MP and NSP in cell-to-cell movement (Hehnle et al., 2004). The 'couple-skating' model (Hehnle et al., 2004; Kleinow et al., 2009) indicates that the MP binds to a ss/dsDNA-NSP complex on the cytoplasmic side of the plasma membranes or microsomal vesicles and consequently transfers the nucleoprotein complex into the cell adjacent to the plasma membrane or through the ER. The 'relay-race' model infers that after the NSP-mediated nuclear export, the viral ss/dsDNA is taken by the MP. The MP then further transports the viral DNA into an adjacent cell, most likely with the assistance of other accessory proteins (Fontes et al., 2004; Krentz et al., 2012). The specific mechanisms which may be employed by the MP for successful cell-to-cell movement and infection by geminiviruses, and some of the host proteins associated with the DNA-protein movement complex, have not to date been well studied. In contrast, the functionality of geminivirus NSP has been explored to a greater degree (Fontes et al., 2004; Zhou et al., 2011).

1.3 Rationale of the study

Given that SACMV is one of the geminiviruses species which affects cassava plants in southern and eastern Africa, it is of utmost importance to understand the mechanisms by which the virus causes disease. BC1-encoded MP and BV1-encoded NSP are movement proteins that not only aid in the cell-to-cell movement of the virus but play a role in pathogenicity and infection. It is therefore vital to fully understand the functionality of the proteins in order to develop strategies to limit or eliminate SACMV movement once the virus has gained entry into the plant. Furthermore, the tertiary or molecular structures of the MP and NSP proteins have not yet been elucidated due to the insoluble nature of these proteins. It is for these reasons that this study was undertaken to express *in vitro*, and structurally characterize, these novel uncharacterized proteins.

1.4 Novelty of the study

South African cassava mosaic virus, the virus of interest for this study, is a bipartite begomovirus, and DNA B is of particular interest for this study. The DNA B BC1 ORF transcribes the cell-to-cell movement protein MP from the complimentary strand (MP) while the nuclear shuttle protein NSP is encoded from the sense strand by the BV1 ORF (Patil and Fauquet, 2015; Vanderschuren et al., 2007; Zhang et al., 2002). To date, the precise functions and mechanisms employed by geminiviral MP and NSP for successful cell-to-cell movement and infection are poorly studied. Additionally, some of the proteins associated with the DNA-protein movement complex of SACMV in cassava have not to date been elucidated. In order to further understand how SACMV and other geminivirus NSP and MP proteins interact with host proteins to potentiate infection, the secondary and tertiary structures of these proteins and their functional domains need to be revealed. Taking into consideration that the MP and NSP proteins of bipartite geminiviruses have not been successfully expressed *in vitro* in order to unravel tertiary structure, this study on SACMV is novel in this regard.

1.5 Aims and Objectives

This study aimed to conduct expression, purification, and structural characterization on truncated domains of the movement and nuclear shuttle proteins of the *South African cassava mosaic virus* while also assessing the computational parameters of these proteins that may relate to functions.

In order for the aim to be achieved the following objectives were carried out:

- Overexpression of the histadine tagged truncated movement and nuclear shuttle proteins using the pCOLDI expression vector.
- Purification of the truncated MP and NSP using nickel charged gravity flow affinity chromatography columns.
- Determination of the truncated MP and NSP secondary structure analysis (Far-UV CD) in the native and denatured states.
- Determination of the truncated MP and NSP tertiary structure using intrinsic and extrinsic fluorescence methods in the presence and absence of ATP (in the native and denatured states).

- Analysis of the truncated MP and NSP binding events using mant-ATP fluorescence in the native and denatured states.
- Confirmation of the molecular weight of the truncated MP and NSP using SE-HPLC.
- Derivation of homology models of the truncated MP and NSP using online modelling websites.
- Creation of molecular docking simulation between the truncated MP and NSP and other ligands/ proteins for validation and comparison

1.6 Structure of the thesis

This thesis commences with Chapter 1 which presents a background, rationale and novelty of the study. Chapter 2 is the Literature Review which encompasses information regarding cassava, the diseases affecting the crop and the putative mechanisms of how the movement and nuclear shuttle proteins of South African cassava mosaic virus could be affecting cassava crops. Chapter 3 will cover the computational characterization and homology modelling of the truncated MP and NSP protein. In chapter 4 the *in vitro* expression and purification of truncated SACMV MP and NSP proteins is described. In Chapter 5, the structural and functional analyses and characterization of truncated SACMV MP and NSP proteins will be described. Lastly, in chapter 6 (Concluding remarks) the importance and impact of the study findings are evaluated in context of the field, and limitations of the study and future work are highlighted and discussed.

Chapter 2- Literature review

2.1 Cassava

Cassava (*Manihot esculenta* Crantz), from the family *Euphorbiaceae* is a woody, perennial shrub. Cassava is an important food source in a number of developing tropical Asian and African countries, due to its ability to offer food security and an income to economically disadvantaged farmers in these regions (Patil and Fauquet, 2009). Cassava plants usually reach a height of between 1-4 metres upon physiological maturity. The cassava tubers are harvested between 6 months to 1 year after planting. These tubers contain a dry matter content of between 30-40% (Adekunle et al., 2016; Ariyo et al., 2005; Legg et al., 2015). Cassava is the third most produced crop in Africa, following the production of sugar cane and maize (***figure 2.1).*** Amongst other staple crops, cassava supplies higher carbohydrate content. The tubers (***figure 2.2B***) recovered from cassava plants are used for human and animal consumption as they provide an important dietary source, while the leaves (***figure 2.2A***) are consumed as a green vegetable, prevalently in East Africa, where these leaves provide a source of proteins, minerals and vitamins (Adekunle et al., 2016; Alabi et al., 2011; Berry and Rey, 2001; Blagbrough et al., 2010; El-Sharkawy, 2004; Legg et al., 2006). Cassava is considered an excellent crop when compared to similar crops such as wheat, rice and maize. This due to its ability to tolerate unfavourable conditions and abiotic stresses. The starch content extracted from cassava plants offers numerous industrial applications such as the production of bioethanol, thickeners in the food industry and processing for the paper industry (Alabi et al., 2011; Chen et al., 2015; Thresh and Cooter, 2005).

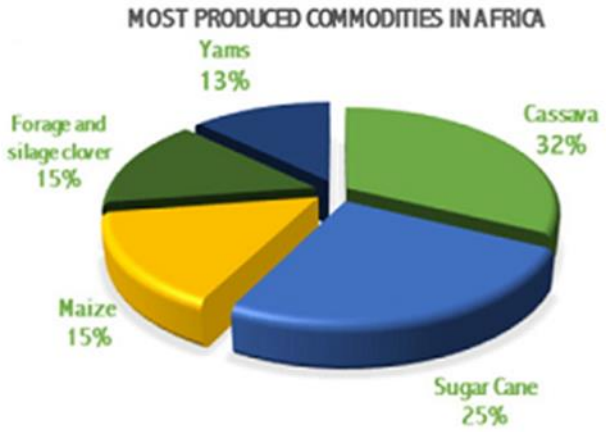


Figure 2.1: The most produced commodities in Africa (by percentage) (Source: FAOSTAT, 2016).

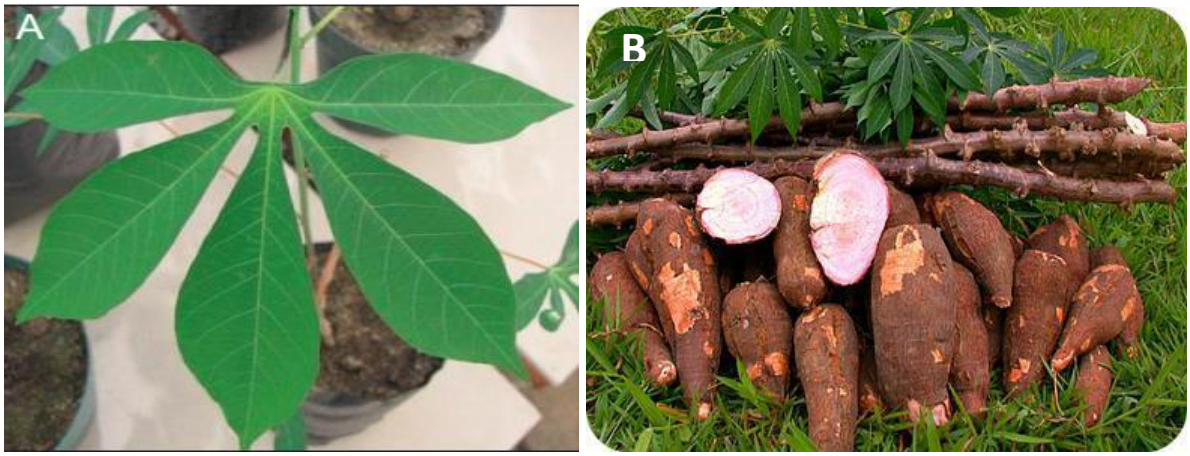


Figure. 2.2: A. -A cassava leaf (source: Legg et al., 2015) and **B.**- cassava roots (source: google images, 2018 (adapted from Malawi News24)).

2.1.1 Cassava production

The world average production of cassava has increased by roughly 100 million tonnes since the year 2000 (1 tonne = 1.000kg). This is driven by rising demands for cassava food products exclusively in Africa and for dried cassava and starch in Asia. In some developing countries cassava is believed to represent the future of food security. It is estimated that an average of 500 million people depend on cassava as a major carbohydrate. This makes cassava the third largest carbohydrate source in the world for human food. Presently the world average yield of cassava is 12.8 tonnes per hectare (world output of ~290 million tonnes) (***figure 2.3***), however there is the potential to produce an approximate of 23.2 tonnes of cassava per hectare (equivalent of ~500 million tonnes per year) (Fauquet, 1990; Garcia and Dale, 1999; Montagnac

et al., 2009; Morgan and Choct, 2016). Seventy percent of the world's cassava is produced in the countries: Nigeria (54 million tonnes per annum), Brazil (23.5 million tons), Thailand (22.5 million tons), Indonesia (24 million tons) and the Democratic Republic of Congo (***figure 2.4***) (Veiga et al., 2016). Of the estimated 13 million hectares of cultivated area in Africa and Asia, 70% has cassava growing on it (El-Sharkawy, 2004).

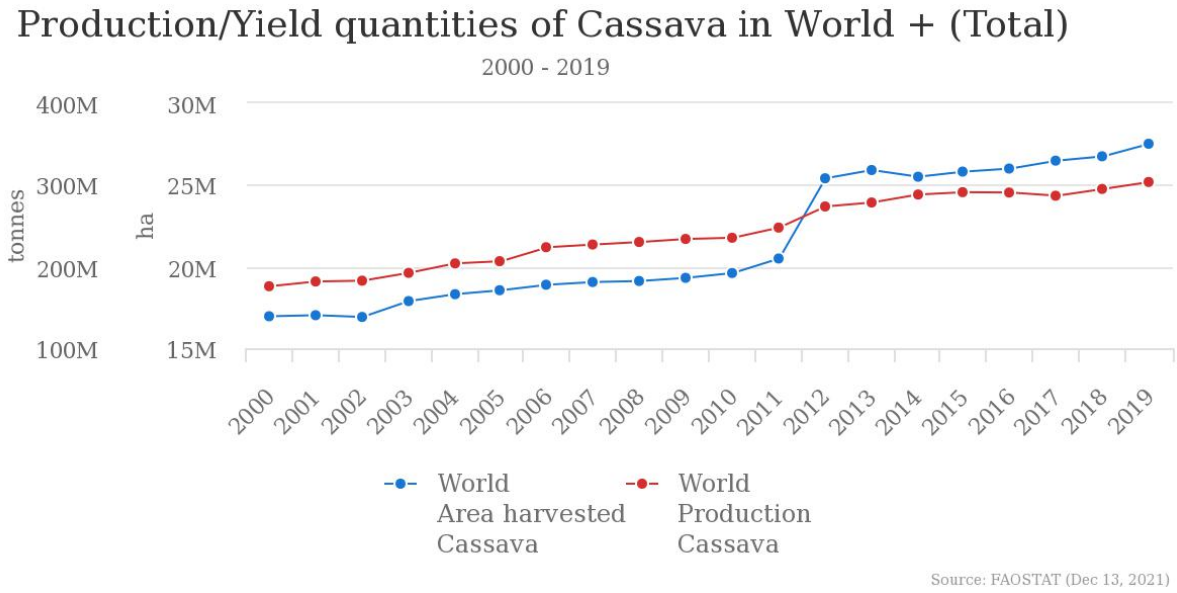


Figure 2.3: The amount of cassava harvested and the production of cassava worldwide between years 2000-2016 (Source: FAOSTAT, 2021).

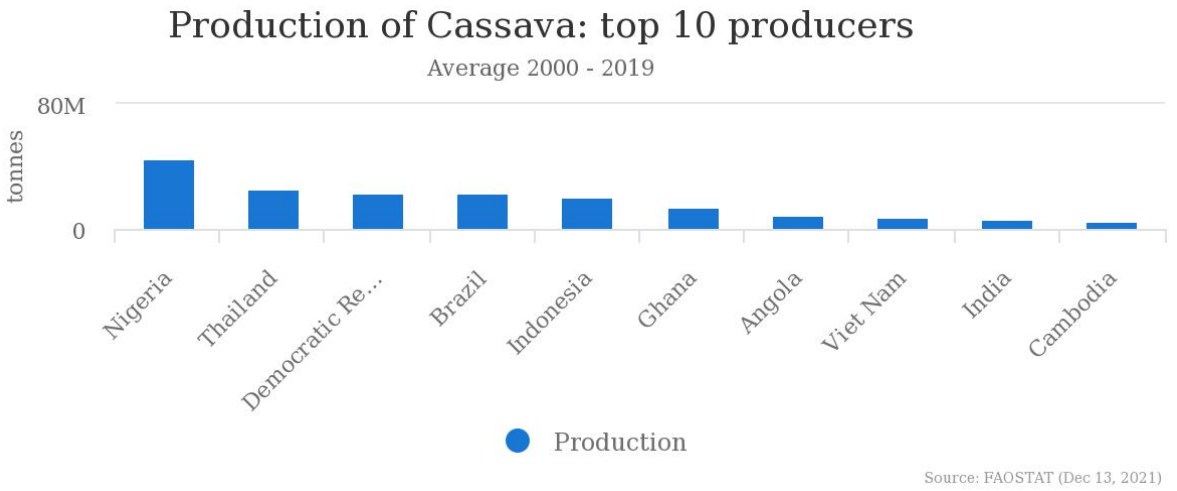


Figure 2.4: The amount of cassava produced (in tonnes) from the top 10 cassava producers worldwide (Source: FAOSTAT, 2021).

Cassava is one of the most critical staple food sources for low-income families in rural areas on the Sub-African continent. Cassava is also a primary crop for many smallholder farmers. The constant expansion of the cassava industry in Africa is driven mainly by the increasing food demand for the crop on the continent (***figure 2.5***) (Morgan and Choct, 2016). Over 50% of global production of cassava occurs within Africa amongst cassava- growing regions of the world (***figure 2.6***) (Alabi et al., 2011). Cassava has the ability to grow successfully under a wide range of agro-ecological zones where maize, rice and other crops cannot thrive. Cassava is tolerant to diseases, drought, poor soils as well as other stresses (biotic and abiotic). This makes cassava a suitable crop for poor farmers to cultivate under marginal environments in Africa (Alabi et al., 2011; Morgan and Choct, 2016; Veiga et al., 2016).

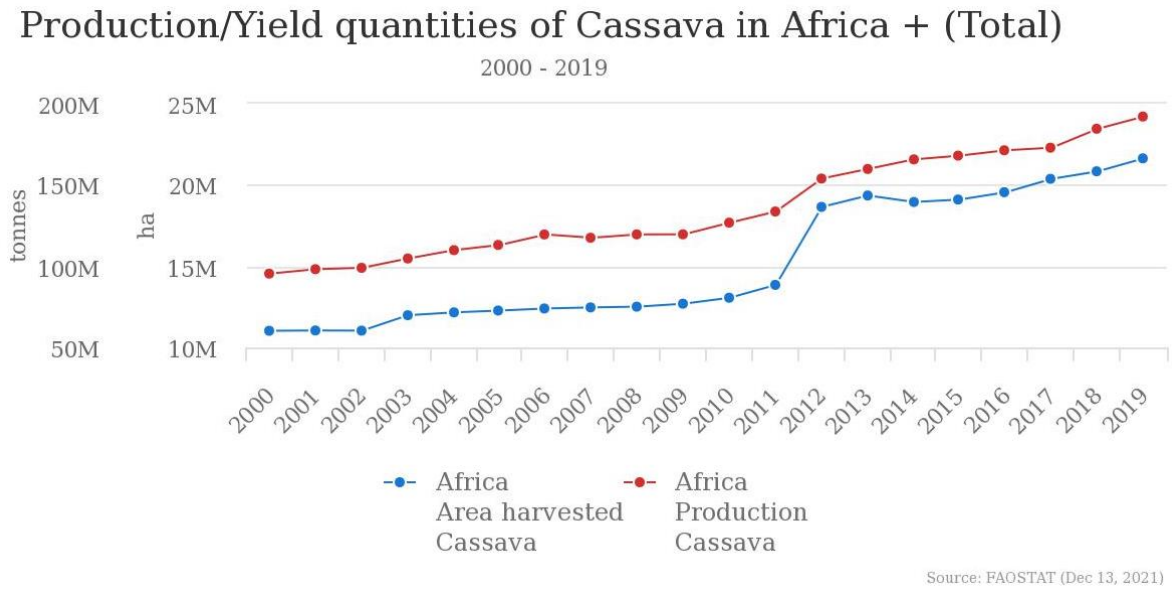


Figure 2.5: Comparison between cassava production and the area harvested for cassava in Africa over a period from 2000- 2016 (Source: FAOSTAT, 2021).

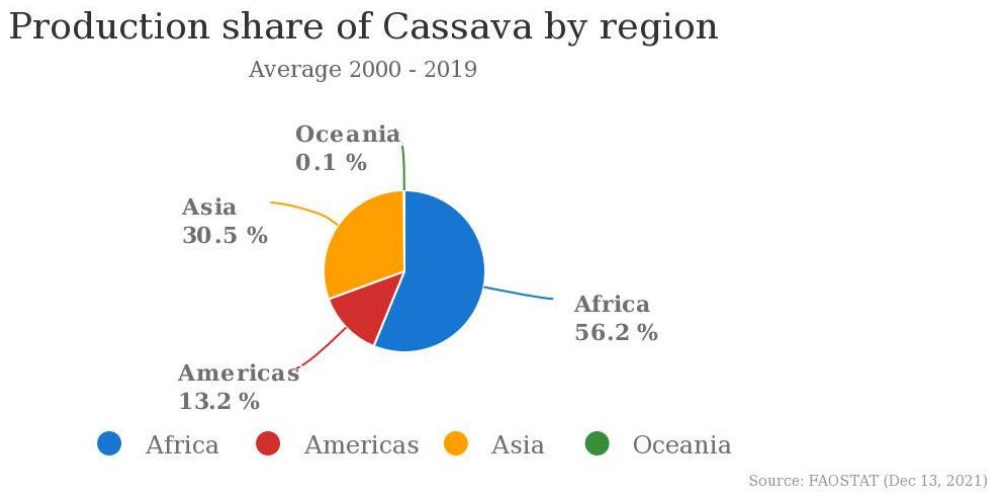


Figure 2.6: Indication of the production share of cassava by region (percentage) for the time period 2000- 2016 (Source: FAOSTAT, 2021).

2.1.2. Cassava products

Cassava plants are grown as tuberous root crops. The roots are a significant source of dietary starch. The roots are consumed either fresh or as a number of forms of processed foods. Cassava roots comprise mostly of carbohydrates as well as 1-3% of crude protein. Dried cassava roots are often produced into toasted cassava flour. Recently in the food industry dried cassava roots are also being used to produce cassava chips. Apart from human consumption cassava roots are also used as animal feed. The production of cassava roots for animal feed has increased by 76 million tonnes since 2007 (Alabi et al., 2011; Morgan and Choct, 2016). Cassava is also increasingly becoming highly popular as a feedstock for ethanol fuel production predominantly within China due to its low cost and high productivity. A single tonne of cassava roots is able to produce at least 188 litres of ethanol (Sánchez et al., 2017). Cassava leaves are highly nutritious and contain high levels of protein (between 16.6-39.9%). Cassava leaves also have high mineral levels and are a valuable source of vitamins B1, B2 and C as well as carotenes. (Alabi et al., 2011; Morgan and Choct, 2016; Veiga et al., 2016).

2.1.3. Threats affecting cassava production

A number of threats affect cassava production, the most important of these being pests and diseases (viral, bacterial and fungal). There are many virus species which affect cassava in Latin America, with the most economically important being the species

which gives rise to cassava frogskin disease (CFSD). Cassava mosaic geminiviruses (CMGs) which cause cassava mosaic disease (CMD) in Africa and Asia, and Cassava brown streak virus (CBSV) and Ugandan cassava brown streak virus (UCBSV) which cause cassava brown streak disease (CBSD) in Africa, have large economic impacts and are considered to be the paramount global threat to cassava production (Legg et al., 2015). There are approximately 15 virus species and several associated strains infecting cassava predominantly in Sub-Saharan Africa and the Asian sub-continent and their associated offshore islands (***table 2.1***). Of the 15 virus species listed, 11 of them are accountable for CMD and CBSD. Cassava green mottle virus, Cassava virus C and Cassava Kumi viruses A and B are also known to infect cassava, however, they are not well characterized and their significance is still not fully known (Legg et al., 2015).

2.2 Cassava mosaic disease (CMD)

As depicted in table 2.1, it is apparent that cassava is susceptible to a wide range of diseases caused by viruses. CMD is one of the most severe and widespread diseases that limits the production of cassava in sub-Saharan Africa. CMD is endemic to sub-Saharan African countries and is caused by six distinct species of bipartite cassava mosaic begomoviruses (CBVs) (Brown et al., 2015 Legg et al., 2015, Rey and Vanderschuren, 2017). These species are: *African cassava mosaic virus* (ACMV), *East African cassava mosaic virus* (EACMV), *East African cassava mosaic Kenya virus* (EACMKV), *East African cassava mosaic Malawi virus* (EACMMV), *East African cassava mosaic Zanzibar virus* (EACMZV) and *South African cassava mosaic virus* (SACMV) (Brown et al., 2015). CMD is a notable disease largely in Africa and Asia where cassava cultivation is of high economic importance (Alabi et al., 2011; Legg et al., 2006). In southern Africa, including South Africa, ACMV, SACMV and EACMV occur (Berrie et al., 2001; Berry et al., 2001).

Table 2.1: Viruses affecting cassava in Latin America, Africa and South Asia (Adapted from Legg et al., 2015).

VIRUS	GENUS/FAMILY	AREA FOUND IN	DISTRIBUTION BY AREA
Cassava common mosaic virus (CsCMV)	*Alphaflexiviridae/ Potexvirus*	Latin America	Colombia, Brazil (isolated casas from Africa, Asia)
Cassava vein mosaic disease (CsVMV)	*Caulimoviridae/ Cavemovirus*	Latin America	Brazil
Cassava virus X (CsVX)	*Alphaflexiviridae/ Potexvirus*	Latin America	Colombia
Cassava new alphaflexivirus (CsNAV)	*Alphaflexiviridae/ Potexvirus*	Latin America	Colombia
Cassava frogskin-associated virus (CsFSAV)	*Reoviridae/ Oryzavirus*	Latin America	Colombia, Brazil, Costa Rica, Argentina
Cassava polero –like virus (CsPLV)	*Luteoviridae/ Polerovirus*	Latin America	Colombia, Costa Rica
Cassava torrado- like virus (CsTLV)	*Secoviridae/ Torradovirus*	Latin America	Colombia, Argentina
Cassava symptomless virus (CsSLV)	*Rhabdoviridae/ Nucleorhabdovirus*	Latin America	Brazil
Cassava Caribbean mosaic virus (CsCAMV)	*Alphaflexiviridae/ Potexvirus*	Latin America	Colombia
Cassava Colombian symptomless virus (CsCSLV)	*Alphaflexiviridae/ Potexvirus*	Latin America	Colombia
Cassava American latent virus (CsALV)	*Secoviridae/ Nepovirus*	Latin America	Brazil, Guyana
African cassava mosaic disease (ACMV)	*Begomovirus/ Geminiviridae*	Africa	sub-Saharan Africa
African cassava mosaic Burkina Faso virus (ACMBFV)	*Begomovirus/ Geminiviridae*	Africa	Burkina Faso
Cassava mosaic Madagascar virus (CMMGV)	*Begomovirus/ Geminiviridae*	Africa	Madagascar
East African cassava mosaic Cameroon virus (EACMCV)	*Begomovirus/ Geminiviridae*	Africa	Sub-Saharan Africa and Comoros
East African cassava mosaic Kenya virus (EACMKV)	*Begomovirus/ Geminiviridae*	Africa	East Africa, Madagascar,

			Seychelles, Comoros
East African cassava mosaic virus – Ugandan Variant (EACMV-UG)	*Begomovirus/ Geminiviridae*	Africa	Sub-Saharan Africa
East African cassava mosaic Malawi virus (EACMMV)	*Begomovirus/ Geminiviridae*	Africa	Malawi
East African cassava mosaic virus (EACMV)	*Begomovirus/ Geminiviridae*	Africa	Sub-Saharan Africa
East African cassava mosaic Zanzibar virus (EACMZV)	*Begomovirus/ Geminiviridae*	Africa	Zanzibar, Madagascar
South African cassava mosaic virus (SACMV)	*Begomovirus/ Geminiviridae*	Africa	South Africa, Madagascar, Zimbabwe
Indian cassava mosaic virus (ICMV)	*Begomovirus/ Geminiviridae*	Asia	Southern Asia, Sri-Lanka
Sri-Lankan cassava mosaic disease (SLCMV)	*Begomovirus/ Geminiviridae*	Asia	Southern Asia, Sri-Lanka
Cassava virus C (CsVC)	*Ourmiavirus/* Unassigned	Africa	Ivory Coast
Cassava green mottle virus (CsGMV)	*Nepovirus/ Comoviridae*		Australia and Asia Pacific Islands, Solomin Islands
Cassava Ivorian bacilliform virus (CIBV)	*Amulavirus/ Bromoviridae*	Africa	Ivory Coast
Cassava Kumi viruses A and B	Uncharacterized	Africa	Kumi region of Uganda

2.2.1. Symptoms of CMD

Symptoms of CMD (***figure 2.7***, ***table 2.2***) vary in type, extent and severity. CMD produces a diversity of foliar symptoms including green to yellow mosaic of the affected leaves. This is often coupled with misshapen and twisted leaves as well as leaf distortion and mottling of leaves. The disease also results in an overall reduction in the size of cassava leaves and plants. The affected plants produce few or no tubers depending on the severity of the disease (Alabi et al., 2011; Thresh and Cooter 2005). Cassava leaves affected by 'green mosaic' display leaves showing light and dark green tissue. In contrast, plants affected by 'yellow mosaic' are more noticeable as the cassava leaves display contrasting green and yellow tissue. Severe chlorosis is

coupled with premature leaf abscission, S-shaped curvature of the petioles of the remaining leaves as well as a noticeable decrease in vegetative growth and yield of the cassava tubers.Severely affected plants are distinguished by extremely stunted growth and produce almost no yield of roots or stems (Thresh and Cooter, 2005).



Figure 2.7: Symptoms of CMD on cassava plants. An infected plant showing severe stunting and distortion of leaves (A) as compared to a healthy plant (B). Leaves of misshapen and twisted leaflets with mottling and mosaic symptoms (C and D) (Source: Alabi et al., 2011).

Table 2.2: CMD symptoms caused by CMGs, the genus and family they belong to the vector used to transmit the virus, and the areas it is predominantly distributed (Adapted from: Alabi et al., 2011).

Virus	Symptoms	Distribution
African cassava mosaic virus (ACMV)	Mosaic, leaf distortion and stunting	Africa and India
East African cassava mosaic Cameroon virus (EACMCV)	Mosaic, leaf distortion and stunting	West Africa, Tanzania
East African cassava mosaic Kenya virus (EACMKV)	Mosaic, leaf distortion and stunting	East Africa
East African cassava mosaic Malawi virus (EACMMV)	Mosaic, leaf distortion and stunting	Malawi
East African cassava mosaic virus (EACMV)	Mosaic, leaf distortion and stunting	East Africa

East African cassava mosaic virus-Uganda (EACMV-UG)	Mosaic, leaf distortion and stunting	Sub- Saharan Africa
East African mosaic Zanzibar virus (EACMZV)	Mosaic, leaf distortion and stunting	Zanzibar, Madagascar
Indian cassava mosaic virus (ICMV)	Mosaic, leaf distortion and stunting	India, Sri Lanka, Togo
South African cassava mosaic virus (SACMV)	Mosaic, leaf distortion and stunting	South Africa, Malawi, Madagascar, Zimbabwe
Sri Lankan cassava mosaic virus (SLCMV)	Mosaic, leaf distortion and stunting	India and Sri Lanka

2.3. Geminiviruses

Geminiviruses belong to the family *Geminiviridae* and are increasingly emerging as plant pathogens in a number of tropical and subtropical countries (***table 2.3***), eliciting a disastrous effect on productivity (Moffat, 1999; Boulton, 2003; Mansoor et al., 2003; Vanderschuren et al., 2007). Geminiviruses infect both monocotyledonous and dicotyledonous plants and are distributed world-wide. Geminiviruses are responsible for disastrous yield losses in a number of economically essential crops such as tomato, maize, cotton, chickpea and cassava (Varsani et al., 2017). The spread of geminivirus-associated disease correlates with the emergence of new virus species and strains as well as the increased presence of the viruliferous whitefly *Bemisia tabaci* Genn. which is the vector of geminiviruses (***figure 2.8***). The increase in the presence of the whitefly is most likely a result of advances in insecticide resistance and global warming (Vanderschuren et al., 2007; Seal et al., 2006).

Based on genome organization, host range, and insect vectors of geminiviruses, nine different genera have been distinguished*: Mastrevirus, Curtovirus, Begomovirus, Becurtovirus*, *Eragrovirus, Turncurtovirus, Topovirus* and the recently added *Capulavirus* and *Grablovirus* (Brown et al., 2015; Varsani et al., 2017). A large number of geminviruses have bipartite genomes and the two segments are similar in size. Viruses belonging to the *Begomovirus* genus, such as SACMV, are transmitted by many biotypes/genetic variants of the whitefly vector *Bemisia tabaci* (Seal et al., 2006). In South Africa and other surrounding southern African countries, SACMV is transmitted by several native and introduced haplotypes (Berry and Rey, 2001; Esterhuizen et al., 2013

Table 2.3: Economically important viral diseases caused by geminiviruses (Adapted from:Vanderschuren et al., 2007).

Disease name	Virus genus	Virion structure	Host crop	Epidemic region	Yield loss
Maize streak disease	*Mastrevirus*	Monopartite	Maize	Sub-Saharan Africa	Average 20-100%
Cassava mosaic disease	*Begomovirus*	Bipartite	Cassava	Africa, India	Overall 12-24% upto 100%
Cotton leaf curl disease	*Begomovirus*	Monopartite (associated β-component)	Cotton	Pakistan	Average 30%, upto 100%
Bean golden mosaic disease/ bean golden yellow mosaic disease	*Begomovirus*	Bipartite	Bean	Florida, Central and South America	10- 100%
Yellow mosaic disease	*Begomovirus*	Bipartite	Grain legumes	India	10-90%
Tomato leaf curl disease/ tomato yellow leaf curl disease	*Begomovirus*	Monopartite	Tomato	Europe, Asia, Americas, Australia	20- 80% upto 100%



Figure 2.8: The whitefly which transmits geminiviruses (Source: Inoue-Nagata et al., 2016).

2.3.1. Structure of geminiviruses

Begomoviruses have circular single stranded DNA (ssDNA) and are monopartite (containing DNA-A only) or bipartite (containing two DNA components, DNA A and DNA B) genomes (Brown et al., 2015). All bipartite begomoviruses have two circular DNA molecules, DNA A (2800 nucleotides) and DNA B (2760 nucleotides) (***figure 2.9A***); both of these are required for the systemic infection of plants. SACMV is one of the bipartite begomoviruses whose DNA A encodes for six proteins while its DNA B encodes for two (***figure 2.9B***). Both of these molecules have a conserved intergenic common region (CR) which entails a stem-loop structure with an invariant nonanucleotide sequence which is TAATATTAC (Berrie et al., 2001).

The DNA A molecule encodes three proteins in the antisense direction namely: the DNA replication associated protein Rep (AC1) which initiates rolling circle amplification by means of cleaving at the origin of replication in the stem-loop and has extensive host gene regulation; the REn protein which is a replication enhancer; TrAP and AC4 ORFs which encode for a transcriptional activator and host RNA silencing supressor protein, respectively. The structural coat protein (CP) is encoded by the AV1 in the sense direction (Patil and Fauquet, 2015). The DNA B ORF encodes for the cell-to-cell movement proteins from the complimentary strand (the MP and NSP are encoded from the sense strand by the BV1 ORF) (Patil and Fauquet, 2015; Vanderschuren et al., 2007; Zhang et al., 2002).

2.3.2. Replication of geminiviruses

Geminiviruses require insect vectors for their transmission. Upon delivery into plant cells, the geminivirus enters a life cycle which entails DNA replication for which it relies on host nuclear replication machinery for replication of its genome, since it does not encode its own DNA polymerases, DNA accumulation and virus assembly, followed by virus spread within the host (Hanley-Bowdoin et al., 1999, 2004, 2013; Gutierrez, 2000; Gutierrez et al., 2002). Geminiviruses require a number of viral proteins in order to replicate successfully in plant cells. One such protein is the Rep protein which is encoded by the AC1 ORF in bipartite geminiviruses (figure 1.10). Transcriptional activator protein (TrAP) (***figure 2.10***) is encoded by the AC2 ORF and activates transcription of the genes which encode for the CP and MP. The AC3 ORF encodes for the replication enhancer protein (REn). REn is not essential for viral

replication, however, the protein enhances viral DNA accumulation and enhances symptom developments in plants (affected by begomoviruses).

The AC4 ORF is found in the AC1 ORF and has the ability to supress RNA silencing. The CP is encoded by the AV1 ORF and has been suggested to be involved in the systemic and cell-to-cell spread of the virus. The nuclear shuttle protein (NSP) is encoded by the BV1 ORF and is required or the movement of ssDNA between the nucleus and cytoplasm. The movement protein (MP) is encoded by the BC1 ORF and is required for the movement of the virus cell-to-cell and long distance (Fondong, 2013).

Infection starts in a plant cell (***figure 2.10***) when viral ssDNA is released from virions and copied in order to generate dsDNA, that assembles with nucleosomes and is transcribed by host RNA polymerase II, which allows for the production of Rep. Rep initiates rolling-circle amplification by means of introducing a nick into a viral dsDNa molecule in order to generate a free 3`-hydroxyl end which primes ssDNA synthesis, this leads to the displacement of the parental strand or inset. The ssDNA released is then converted into dsDNA in order to re-enter the replication cycle. Viral replication transitions to recombination-dependent- replication and this is initiated by homologous recombination between a partially replicated ssDNA and a closed, circular dsDNA in order to form a looped molecule which serves as a template for single stranded and double stranded DNA synthesis. In later stages of infection, Rep represses its own transcription which leads to the activation of the TrAP expression which in turn activates the CP and NSP expression. Circular ssDNA may then be encapsidated by CP into virions which are then available for whitefly acquisition. The NSP then binds to viral DNA and moves the viral DNA across the nuclear envelope. The MP then transports it across the plasmodesma. It is not known if viral DNA moves ssDNA or dsDNA or if the the DNA is linear versus if the DNA is circular (Hanley-Bowdoin et al., 2013).

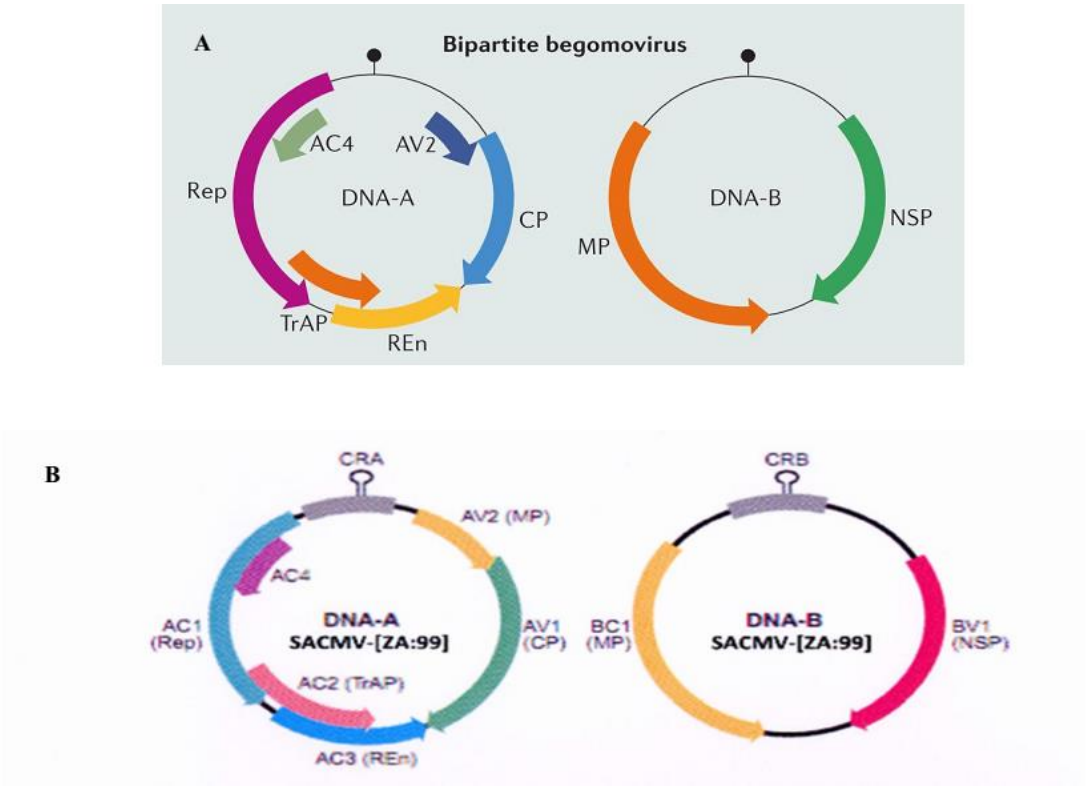


Figure 2.9: (A) A schematic diagram of a bipartite begomovirus, showing both DNA-A and DNA-B and the relative sizes and proteins which they encode for (Source: Hanley-Bowdoin et al., 2013) and **(B)** a detailed illustration of the SACMV bipartite genome, illustrating approximate position of all proteins present (Source: Berrie et al., 1998).

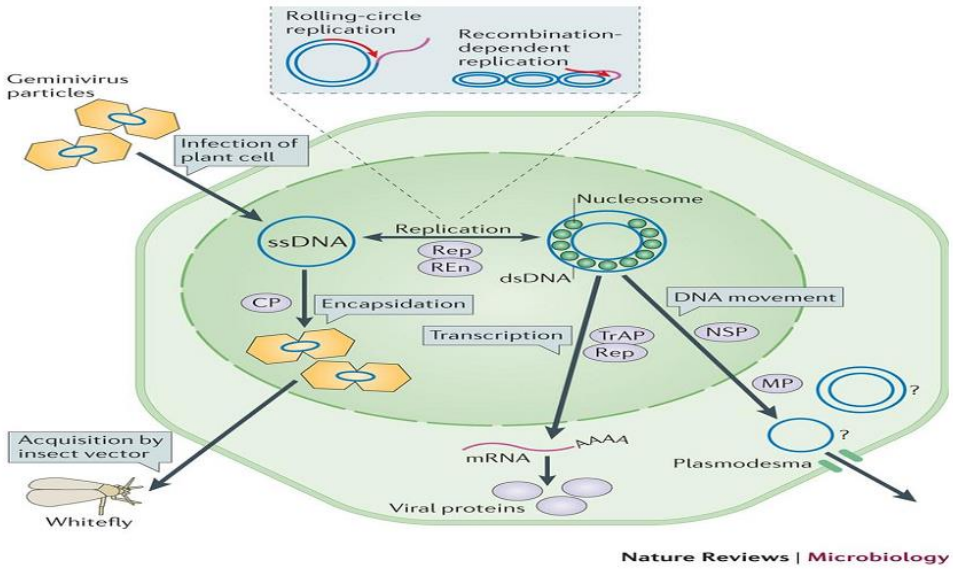


Figure 2.10: The replication of begomoviruses as it begins in a plant cell (host cell) (Source: Hanley-Bowdoin et al., 2013).

2.4 Cell-to-cell movement of viruses

In order to facilitate movement from cell-to-cell, viruses exploit the plasmodesmata, cytoplasmic bridges and cytoplasmic continuity between adjacent plant cells (Reichel et al., 1999). The plasmodesmata (PD) are channels contained within plant cell walls that enable communication amongst adjacent cells. The PD allows for the trafficking of macromolecules. The cell-to-cell movement of macromolecules may be specific or non-specific and is regulated during the course of plant development and differentiation (Haywood et al., 2002; Su et al., 2010). The movement of a virus can be divided into three distinct categories: (1) intracellular movement, (2) intercellular movement and (3) systemic movement. Intracellular movement infers the movement of the virus to the periphery of a cell and includes all metabolic activities which might be of essence to recycle the host components required to facilitate virus movement processes and host-interacting viral constituents required for the continued transport of the intracellular movement complex. Intercellular movement infers movement of the virus between cells. In order for successful plant virus infection, the initial plant virus infection must spread between cells and the virus is required to move through the modified PD which connects neighbouring cells. Upon establishment of intracellular and intercellular movement, the virus has the ability to invade the vascular cells of the plant and subsequently spread systematically throughout the plant. Movement through the PD entails modification of the pores of the PD of parenchyma cells as well as between the sugar- transporting phloem sieve elements and employing xylem vessels (Harries and Nelson, 2008).

In order for successful systemic infection, plant viruses employ host factors and these facilitate cell-to-cell as well as long distance movement (Min et al., 2010). To aid systemic infection in plants viruses, encode specialised movement protein (MPs) and these permit virus movement from cell-to-cell through the PD and other distant tissues through the phloem (Amari et al., 2012). In addition it has, been noted that interactions between virus encoded MPs and the host proteins are crucial in order for a possible systemic infection to occur (Zhou et al., 2011). Typically, a viral MP is described by its ability to increase the plasmodesmal size exclusion (SEL) and move cell-to-cell. But other viral proteins which do not themselves move may be of essence for the movement process (Boevink and Oparka, 2005).

2.4.1 *Geminivirus* movement proteins

As per descriptions from previous sections (1.3.1 and 1.3.2), DNA B, BV1 encodes the nuclear shuttle protein (NSP) and this controls trafficking of viral DNA in and out of the nucleus. In contrast, BC1 encodes the MP and this serves as an adaptor of the cell membrane and mediates cell-to-cell transfer of a ss/ds DNA complex through the PD as well as long-distance spread through the phloem (Krentz et al., 2012).

Existing models of viral cell-to-cell movement indicate that the translocation of plant viruses into adjacent uninfected cells occurs either by the PD or viral induced ER-derived tubules (Lazarowitz and Beachy, 1999; Fontes et al., 2004; Gafni and Epel, 2002). Based upon the biochemical activities of NSP and MP from Squash leaf curl virus (SLCV) and Bean dwarf mosaic virus (BDMV), it is suggested that geminiviruses use both mechanisms for cell-to-cell trafficking of viral DNA (Noueiry et al., 1994; Fontes et al., 2004; Sanderfoot and Lazarowitz, 1995).

The two existing models prominently suggested for the active role of geminivirus MP and NSP during cell-to-cell transport are the ‘couple-skating’ and ‘relay-race’ models. The ‘couple-skating’ model indicates that the MP binds to a ss/dsDNA-NSP complex on the cytoplasmic side of the plasma membranes or microsomal vesicles and consequently transfers the nucleoprotein complex into the cell adjacent to the plasma membrane or through the ER. The ‘Relay-race’ model infers that subsequent to the NSP-mediated nuclear export, the viral ss/dsDNA is taken by the MP. The MP then further transports the viral DNA into an adjacent cell, most likely with the assistance of other accessory proteins (Krentz et al., 2012; Fontes et al., 2004).

The precise mechanism which may be employed by the MP for successful viral infection in geminiviruses and some of the proteins associated with the DNA-protein movement complex has not to date been well studied. Little is known about the proteins (including those found in SACMV), the mechanisms of binding by the protein to DNA or other proteins as well as the associated functions.

Bean dwarf mosaic virus also a bipartite geminivirus has been investigated and may provide clues as to how SACMV may move within. Given that all geminiviruses, including SACMV, require functional MP and NSP in order to systematically infect its host, comparisons with other known geminivirus MPs may be advantageous. It has also been derived that as BDMV, SACMV also does not require the CP for cell

movement or nuclear import of the virus in the host plant (Levy and Tzfira, 2010; Sudarshana et al., 1998). Studies carried out on BDMV have suggested that the MP facilitates cell-to-cell movement of double stranded viral DNA in bipartite geminiviruses. In contrast NSP facilitates the export of single stranded and double stranded DNA from the nucleus. Such studies, supplemented by studies carried out on Cabbage leaf curl virus (CaLCV) and SqLCV, have led to the suggestion that MP and NSP of SACMV facilitate similar movements and functions (Sudarshana et al., 1998).

The BC1 gene of SACMV movement protein is 923 nucleotides and encodes for 307 amino acids (***figure 2.11***). The various MP domains as predicted by Fondong (2013) (fig.1.12) are: The pilot domain, required for the localization to the cell periphery; a central anchor domain, which targets the protein to the cell periphery and a C-terminal domain which facilitates oligomerization. There is also thought to be a NSP binding site located at the centre of the MP (***figure 2.12***). Their respective functions as well as the predicted secondary, tertiary and molecular structures have till date not been elucidated. Certain parameters however, such as the theoretical molecular weight and net charge can be obtained using online tools such as ‘Encorbio’ and ‘Protparam’. Based on the lack of information on the structure of the SACMV MP, it is essential to conduct in silico and experimental research on the MP of SACMV in order to determine the predicted functions and number of domains and other structural features. This will aid in determining potential binding sites to host proteins which facilitate movement of the virus as well as determining the actual molecular weight and net charge of the protein structures of geminivirus MPs have not to date

been determined.

```
        10         20         30         40         50         60
MDAQFTVTDN NYINSKRTEY ALTNDAAPIN LQFPGSFEQA TMRLKGRCMK IDHIIIEYRN

        70         80         90        100        110        120
QVPFNATGSV IVEIRDNRVS LDDAAQAAFT FPIACNVDLH YFSSTYFSIS EPSPWRIMYR

       130        140        150        160        170        180
VEDSNVIEGV KFASIKAKLR LSSAKHSTDI RFKPPTINIL SKGYTKDCID FWSVEKGETR

       190        200        210        220        230        240
RRLLNPTPTA RSQQPITHRP ITIHPGETWA TRSQIGLPSS LGPAQLEHFR SQSMRMDPST

       250        260        270        280        290        300
TPTDLDNDST EYPYQKLHRL HTPDLDPGDS ISQAQSDSVS RKDLETLLES TINKCLYKIK

SDAPRQL
```

Figure 2.11: Amino acid sequence of SACMV MP (Source: Genbank: AAF34900.2; Berrie et al., 2001).

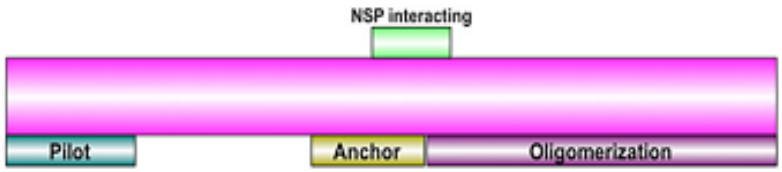


Figure 2.12: The geminivirus MP and its predicted domains (source Fondong, 2013).

The NSP protein has been shown to bind to DNA in a sequence non-specific manner as elucidated in studies for SqLCV (Pascal et al.,1994), BDMV (Rojas et al., 1998) and AbMV (Hehnle et al., 2004). In order to achieve cell-to-cell and long distance movement, the NSP-viral DNA complex is trapped by the MP in the cytoplasm and redirected to the cell wall, where it moves across into adjacent cells where the NSP directs the viral genome to the nucleus for new cycles of replication (Lazarowitz and Beachy, 1999; Noueiry et al., 1994; Sanderfoot and Lazarowitz, 1995; Sanderfoot et al., 1996; Ward et al., 1997, Fondong 2013). The NSP facilitates export of single stranded and double stranded DNA from the nucleoplasm to the cytoplasm through the nuclear pore complex. The NSP increases the export of viral DNA from the nucleus through the nuclear envelope towards the cytoplasm (Gafni and Epel, 2002; Rojas et al., 1998).

Thus far, the function of geminivirus NSP has been well explored. However, the complete and partial structures of this protein has not yet been elucidated. The predicted domains of NSP (Fondong, 2013) and the amino acid sequences are the only structural characteristics for geminivirus NSP thus far available. The BV1 gene encoding the SACMV nuclear shuttle protein is 777 nucleotides and encodes for 258 amino acids (***figure 2.14).*** There are no predicted domains for the NSP protein, however, Fondong, 2013, (***figure 2.13***) has suggested that there are: two nuclear localization signals (NLS-A and NLS-B); a DNA binding site which induces a hypersensitive response on the N-terminus; a nuclear export signal (NES), a MP interacting binding site and an Arabidopsis nuclear shuttle protein interactor (AtNSI) on the C-terminus. The predicted secondary, tertiary, quaternary and molecular structures of geminivirus NSPs have also till date not been elucidated. Certain parameters however, such as the theoretical molecular weight and net charge may be obtained using online tools such as 'Encorbio' and 'Protparam'. Based on the lack of information on the structure of the SACMV NSP, it is essential to conduct in silico and experimental research on the NSP of SACMV in order to determine the predicted functions and number of domains and other structural features.

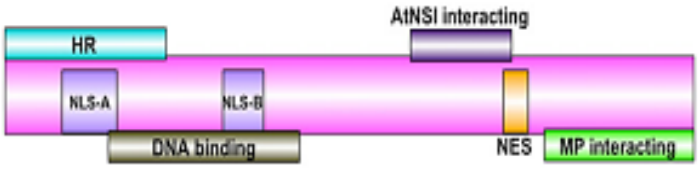


Figure 2.13: The geminivirus NSP and its predicted domains (Fondong, 2013).

```
        10         20         30         40         50         60         70
MYSVYRRGYK TPYRSPYGAR VTPYVYRKTS GKQTSKSRVP RKLVYESPK GLYTRRSLED IHNGASLKLS

        80         90        100        110        120        130        140
QQGDYTSYVS LPCRGIEGNG GRSVDHIKLL NLRVSGTVNV SQVGGDDNMG ERTTMRGIFF MACLVDKKPF

       150        160        170        180        190        200        210
VPEGVSILPT FNELFGEYES VYGMPRLKEN VRHRYRVIGT SKLYITTDED HIQKPFSLRRR LSGGKYPIWS

       220        230        240        250        258
SFKDVDNSST GGNYKNINKN AILVSYVWVS LCRTTCDVYS QFVLNYVG
```

Figure 2.14: Amino acid sequence of SACMV NSP (Source: Genbank: NP_620667.1; Berrie et al, 2001).

2.4.2 Cell-to-cell MP and NSP- host interactions

The ability of viruses to infect plants depends on the capacity of the virus to utilize cellular factors. Geminivirus proteins have been found to interact with various proteins and other factors within their respective hosts in order to establish infection. Geminiviruses can cause disease either by direct or indirect interactions of their proteins with host proteins thereby causing direct (by host protein interactions) or indirect (a consequence of infection) transcriptome changes. Functional genomics allows for the profiling of host plant genes, transcripts and proteins which may interact directly or indirectly with virus proteins, to facilitate replication and movement. The response elicited by the host plant depends on the host and virus protein structures and interactions which subsequently leads to signaling and biological processes downstream (Guitierrez, 2002; Jeske, 2004; Whitham and Wang, 2004;). In addition, geminivirus proteins have been implicated in various host-responsive processes such as transcriptional regulation, DNA replication, controlling the cell cycle, cell proliferation and differentiation and macromolecular trafficking in whole plants (Ascencio- Ibáñez et al., 2008; Mariano et al., 2004; Fontes et al., 2004). Several proteins have been shown to interact with geminivirus proteins (Hanley-Bowdoin et al., 2013). However, despite several transcriptome and geminivirus-host interaction studies, very few host proteins have been identified that interact directly with virus proteins, including the BC1-encoded MP, whose structure-function relationship is not yet elucidated. For determination of the complete function of the MP in SACMV, it is imperative that the structure of the protein is elucidated, and that host candidate protein interactors be determined. These host protein interactors should be tested through protein-protein interactions with the MP.

A host protein interactor of interest is Histone H3, a host nucleoprotein identified in *Arabidopsis* which has been found to interact with both the NSP and MP of BDMV (Zhou et al., 2011). The interaction was confirmed using co-immunoprecipitation assays *in vitro* and *in vivo*. Based on this research Histone H3 is a host protein of interest for the study for virus-host protein interactions such as that between the histone in the cassava host and the MP of SACMV. Host proteins such as cellular chaperones are essential for virus replication and cell-to-cell movement in RNA viruses (Verchot, 2012). Plant host proteins like heat shock proteins (HSPs) 40,70 and 90 have previously been found to interact with viral movement proteins in many DNA

and RNA viruses (Chen et al., 2008; Mayer, 2010). HSP70 particularly has been thought to be involved in the cell-to-cell transport of movement proteins of various RNA viruses such as closteroviruses (Lu et al., 2009; Shimizu et al., 2009; Soellick et al., 2000). Based on the suggestion that HSPs interact with MPs in many viruses (DNA and RNA) and that HSP70 perhaps is involved with the MPs of RNA viruses HSP70 becomes a host protein of interest in order to test for virus-host protein interactions between the cassava host and the MP of SACMV.

2.5. Protein expression systems

In order to successfully express and purify a recombinant protein a suitable expression system is required. There are a number of host expression systems which include bacteria, yeast, plants, insects, filamentous fungi and unicellular algae. A popular expression system employed for recombinant protein expression is prokaryotic expression using an *Escherichia coli* (*E.coli*) host (Yokoyama, 2003). *E.coli* has been a common choice for recombinant protein expression due to the many advantages it presents, such as its ability to grow on inexpensive media, its short doubling time, rapid growth and ease of selection of mutants (Kaur et al., 2018). Furthermore, *E.coli* cells possess great efficiency to incorporate foreign DNA and express recombinant proteins at a very high rate. An estimated 80% of proteins from the Protein Data Bank (PDB) whose three-dimensional structures were solved by the year 2003, were expressed in an *E.coli* host (Kaur et al., 2018). While E.coli as a recombinant host expression system is the most popular, other host expression systems also have their own benefits. Insect cells as a host expression system have gained popularity due to their ability to provide high-throughput expression of sequences for functional screening. Yeast systems using *Saccharomyces cerevisiae* and *Pichia pastoris* have become increasingly common in recombinant protein expression due to their ability to express proteins in very high densities for large scale productions, in addition expression in yeast systems provides stable expression of recombinant proteins with post-translational modifications (Krummen and Anderson, 2002).

2.6. Protein expression and purification

Recombinant protein expression is accomplished by the completion of multiple steps, an outline of these steps is shown in figure 2.15. There are two main components needed for recombinant protein expression, these are the expression vector and the expression host (Kaur et al., 2018).

The design of an expression vector ensures that the gene cloned into it will be transcribed and/or translated. The expression vector comprises of all components needed for the expression of foreign genes. The crucial components of an expression vector are illustrated in figure 2.16 (Kaur et al., 2018).

Once the desired gene has been successfully cloned into the selected host expression vector, an appropriate *E.coli* host strain must be selected in order for successful expression of the recombinant protein. The *E.coli* host strain BL21 is preferred for recombinant protein expression as it produces the RNA polymerase for T7 promoters. BL21 is also a preferred *E.coli* host strain due to its ability to grow vigorously in minimal media (Kaur et al., 2018).

Once the appropriate host strain has been selected, it is important that the induction conditions are selected, in order to produce the recombinant protein. These induction conditions are often optimized in order to increase the yield of the protein product. Induction conditions which are often optimized include, varying the concentration of the inducer such as Isopropyl-β- d-1- thiogalactopyranoside (IPTG), increasing the incubation time and varying the induction temperature (Kaur et al., 2018).

In order to carry out single step protein purification of the recombinant protein from the total protein pool of host cells, affinity tags are used. The host cell system also produces large amounts of metabolic proteins apart from the recombinant protein. Therefore, in order to recover the recombinant protein of interest from the crude host extract, it is vital to express an affinity tagged protein. There are a range of popular affinity tags including poly (His), chitin binding protein (CBP), maltose binding protein (MBP) and glutathione-S-transferase (GST). Affinity tags allow for the purification of a variety of proteins relatively easily. The most widely used affinity tag is the poly (His), as it binds to metal matrices such as nickel, cobalt, copper and zinc.

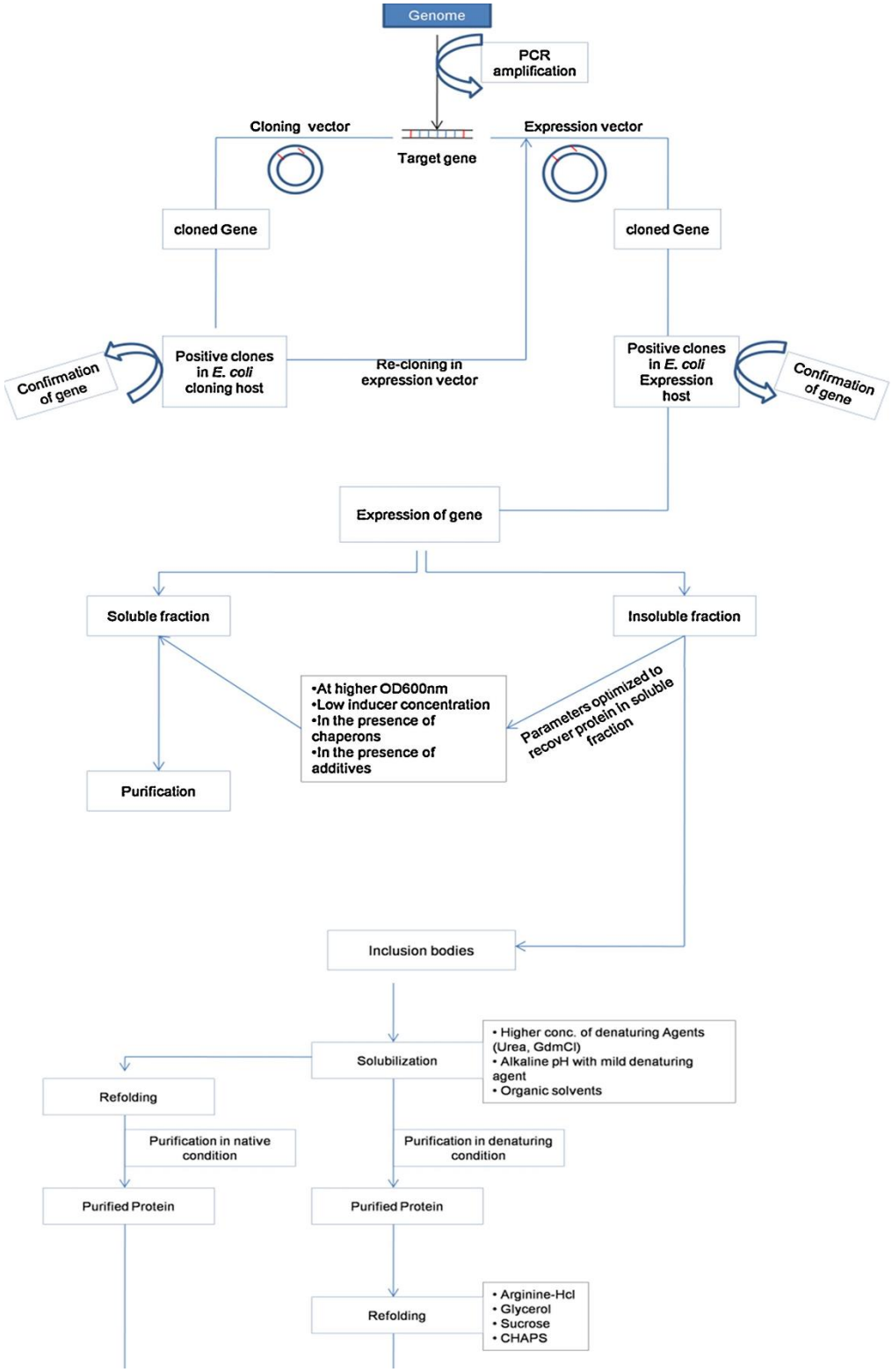


Figure 2.15: Schematic representation of recombinant protein expression in *E.coli* (Kaur et al., 2018).

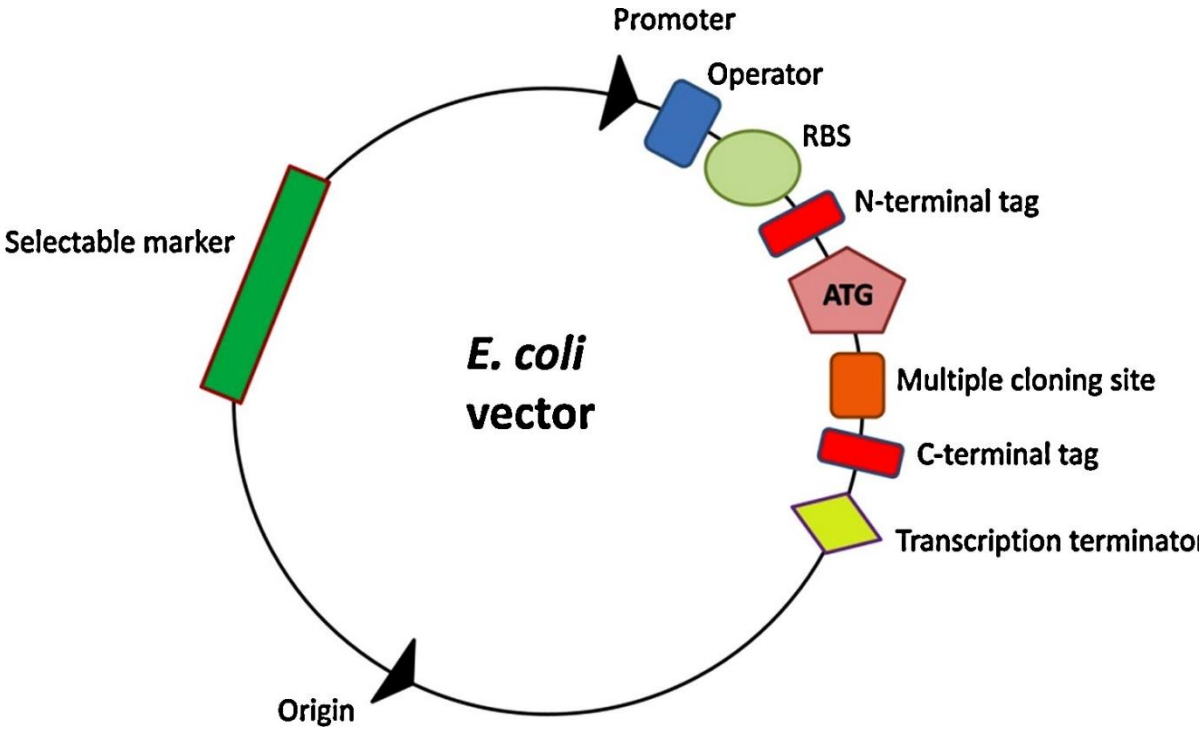


Figure 2.16: Schematic of an *E.coli* expression vector (Kaur et al., 2018).

2.7. Structural protein characterization

When investigating a protein, it is essential to identify the structural properties of the protein in order to determine its functionality. Proteins fold into three-dimensional confirmations, determined by their primary amino acid sequence. A complete protein structure is described by four levels of protein complexity: primary, secondary, tertiary and quaternary structure. The primary structure of a protein is described as the linear amino acid sequence of a proteins polypeptide chain. The secondary structure of a protein is the local spatial conformation of the polypeptide backbone, common secondary structure conformations include α-helices, β-sheets and turns. The tertiary structure of a protein refers to the three-dimensional arrangement of all the atoms within the protein, the tertiary structure can constitute of arrangements of domains, motifs and folds. The quaternary structure of a protein (mainly those >100kDa in mass) are oligomers which consist of more than one polypeptide chain. A number of protein structures (tertiary) and their functions have been solved using x-ray crystallography and nuclear magnetic resonance (NMR) (Sun et al., 2004). Circular dichroism (CD) is a structural technique commonly used to assess the secondary structure of a protein. CD signals which are detected by the spectropolarimeters only arise when absorption of radiation occurs. Therefore the spectral bands are easily assigned to individual

structural features of a molecule. The technique of CD provides an advantage in the study of proteins, as complementary structural information can be obtained from a number of spectral regions. In proteins, the chromophores which are of interest include the peptide bond which absorbs below 240 nm, the aromatic amino acid side chains, which absorb in the range of 260 – 320 nm and disulphide bonds which absorb around 260 nm. The technique of CD reveals much information about proteins including; the secondary structure composition which can include α-helices, β-sheets and turns from the peptide bond region (below 240 nm), the tertiary structure fingerprint by assessment of the aromatic amino acid side chains (between 260 - 320 nm), integrity of co-factor binding sites, conclusion about the overall structure features of proteins (CD measurements in the far UV give quantitative estimates of secondary structures which can be compared to results attained from x-ray crystallography and NMR) and conformational changes in proteins (by assessing the structural changes in proteins upon binding of ligands) (Kelly et al., 2005).

Fluorescence spectroscopy has been commonly employed as a tool for structural protein analysis. Intrinsic protein fluorescence employing a natural protein fluorophore, tryptophan provides information on the conformational changes of a protein. Extrinsic fluorescent dyes are employed to offer information about other structural properties of proteins. Extrinsic dyes can covalently attach to the protein. One such extrinsic dye used for protein characterization is 1-anilinonaphthalene-8-sulfonate (ANS) **(*figure* 2.17)**, which binds tightly to a protein in a polar environment and can therefore apart from tertiary structure determination also be used to assess a proteins folding/unfolding status (Hawe et al., 2008).

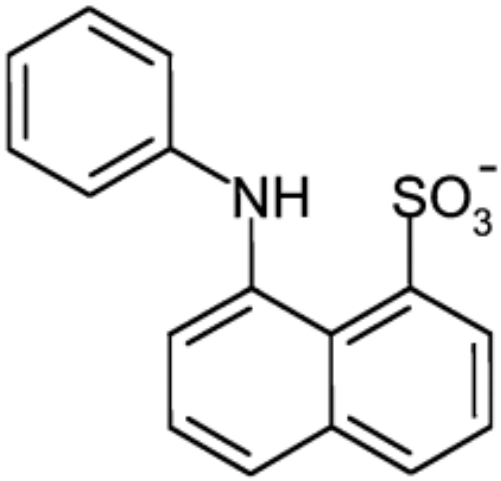


Figure 2.17. Chemical structure of ANS (Hawe et al., 2008).

The function of a protein depends on the ligands or molecules which they interact with. Adenosine triophosphate (ATP) is a common ligand which binds to proteins and is often used as a residue to determine protein-ligand binding (Chauhan et al., 2009). The binding of methylanthraniloyl (mant) derivatives such as mant-ATP is often used to study the mechanism of nucleotide binding. Upon the binding of the mant-ATP nucleotides to the protein of interest there is a large increase in fluorescence energy transfer at 440 nm allowing for the direct measurement of nucleotide affinity (Ni et al., 2000).

Chapter 3
Computational analysis of the MP and NSP of SACMV

3.1. Background

In order to fully understand any protein's function, and the mechanisms employed by the protein to serve its function it is important to determine the biological properties of the protein. Biological properties include a protein's secondary structure, hydrophobicity, the number of domains, the ability to bind to other molecules and the possible intrinsically unfolded regions of the protein (Radivojac et al., 2007).

South African cassava mosaic virus (SACMV) belongs to the *Begomovirus* genus and comprises of a circular single stranded (ss) bipartite genome. The genome consists of a DNA A molecule and a DNA B molecule. The ORF of the DNA A molecule encodes for three proteins in the antisense direction namely, the Rep-associated (AC1), replication enhancer Ren (AC2) and transcriptional activator TrAP (AC3) proteins. An overlapping AC4 ORF encoding for a putative host RNA silencing suppressor is also located within the DNA A molecule (Patil and Fauquet, 2015). The DNA B molecule encodes for two ORFs; the BC1 ORF encodes for the movement protein (MP) from the complementary strand of the while the BV1 ORF encodes for the nuclear shuttle protein (NSP) from the sense strand (Patil and Fauquet, 2015; Vanderschuren et al., 2007). A putative AC5 ORF has been proposed in bipartite geminiviruses, including SACMV (Rey and Vanderschuren, 2017).

The structural and functional characterization of proteins, specifically those which are novel and have no homologs in any databases, may pose many uncertainties in experimental approaches (Pierre and Porcelli, 2010). The determination of protein structures is experimentally and timeously onerous. To date only a fraction of viral protein structures have been fully determined (Dorn et al., 2014). However, the structural and functional characterization of proteins can be successfully predicted using computational approaches, driving successful follow-upexperimental assays (Pierre and Porcelli, 2010). Recently the growing availability of computational genomic-level sequence information databases for thousands of viruses and other species has brought along high-throughput computational experimental data. The availability of protein structure and function databases offer new opportunities for

computational protein function prediction. A number of methods have been recommended to generate computational data. These include function prediction from amino acid sequences, inferred evolutionary relationships, protein-protein interaction networks, protein structure data, microarrays or a combination of data types (Radivojac et al., 2013).

Computational approaches are employed to find predictive secondary structures, generate models and to provide virtual screening of biological properties of proteins and other compounds (Estrada and Peña, 2000). Computational methods are becoming increasingly popular and are often favoured in protein structure predictions and modelling as it is often the only way to obtain structural information about novel proteins especially if no experimental data is available (Pierre and Porcelli, 2010; Prajapat et al., 2010). Protein sequence data bases such as 'Swiss-Prot'; 'TrEMBL' (http://www.uniprot.org) and the 'Protein Data Bank' (PDB) (http://www.wwpdb.org/) are databases which contain amino acid sequences which are translated by nucleic acid sequences (Lesk, 2010). These databases are employed by web programmes, services and software such as 'PsiPred' (http://bioinf.cs.ucl.ac.uk/psipred/) and 'Predict Protein' (http://www.predictprotein.org/), which create alignments to other homologous proteins and predicts various aspects of structure and function based on the annotations listed within the databases (Roche et al., 2011, Rost and Liu, 2003). Homology modelling programmes such as 'Swiss-model', 'Phyre2' and ' PyMOL' also employ protein databases in order to generate 3D models of proteins for molecular visualization (Prajapat et al., 2011). Computational studies have been carried out on other geminivirus proteins such as the Replication associated protein (Rep) of Tomato yellow leaf curl virus (TYLCV), where the study focussed on homology modelling and the docking of Rep with other proteins using '3D JIGSAW' and UCSF Chimera software (Prajapat et al., 2011).

While some structural research (*in silico* and *in vitro*) has been carried out on some of the proteins encoded by DNA A such as Rep, very little has been done on the proteins encoded by DNA B. Until recent studies undertaken on SACMV in our laboratory, very little was known about the structural properties of the MP of geminiviruses. While some functional studies have been carried out on the NSP of geminiviruses, there has been no structural research done to date. This chapter specifically focusses on the computational analysis of truncated regions of SACMV MP (N-ter) and NSP (C-ter)

encoded by DNA B, a bipartite begomovirus from the family *Geminiviridae*. Computational analyses of these proteins are crucial to attain as both of these truncated proteins are novel with mostly little to no comparable information available structurally.

3.2. Materials and Methods

3.2.1. Truncation of MP and NSP from full length proteins

As per previous studies undertaken on the full length (308 amino acids) MP (Nankoo, 2018), it was determined that the full-length protein contains multiple intrinsically disordered regions contributing to the proteins' insolubility. In an effort to attain soluble proteins, it was decided that to avoid working with full- length proteins with possibly long intrinsically disordered regions it would be easier to work with specific shorter domains and truncate both proteins. Since the domain with less intrinsically disordered regions was the N-ter for the MP (Fondong, 2013; Nankoo, 2018) this domain was selected (amino acid 1- 176). Since the C-ter for NSP contains a MP interacting site with not many intrinsically disordered regions this domain (amino acid 171-258) was chosen for expression and characterization (Fondong, 2013).

The full-length templates (***figures* 3.1 and 3.2**) from NCBI were used to create the truncations of both proteins.

```
        10         20         30         40         50         60
MDAQFTVTDN NYINSKRTEY ALTNDAAPIN LQFPGSFEQA TMRLKGRCMK IDHIIIEYRN

        70         80         90        100        110        120
QVPFNATGSV IVEIRDNRVS LDDAAQAAFT FPIACNVDLH YFSSTYFSIS EPSPWRIMYR

       130        140        150        160        170        180
VEDSNVIEGV KFASIKAKLR LSSAKHSTDI RFKPPTINIL SKGYTKDCID FWSVEKGETR

       190        200        210        220        230        240
RRLLNPTPTA RSQQPITHRP ITIHPGETWA TRSQIGLPSS LGPAQLEHFR SQSMRMDPST

       250        260        270        280        290        300
TPTDLDNDST EYPYQKLHRL HTPDLDPGDS ISQAQSDSVS RKDLETLLES TINKCLYKIK

SDAPRQL
```

Figure 3.1: Amino acid sequence of full-length SACMV BC1 (Source: Genbank: AAF34900.2; Berrie et al., 2001).

.

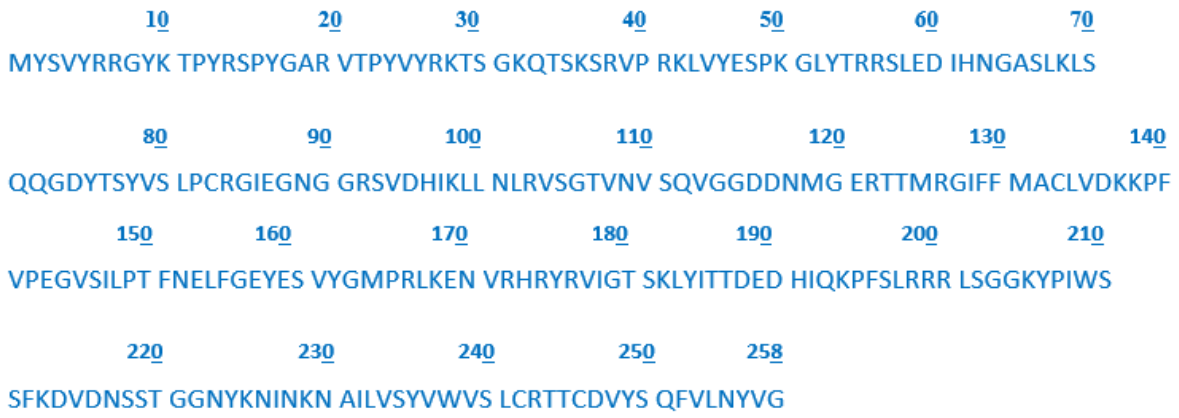


Figure 3.2: Amino acid sequence of full-length SACMV NSP (Source: Genbank: NP_620667.1; Berrie et al, 2001).

3.2.2. Determining familial matches of the truncated MP and NSP proteins using InterProScan

'Inter-ProScan' (https://www.ebi.ac.uk/interpro/) employs functional analysis of proteins by classifying them into families and further predicts functional domains. This database was chosen amongst others as it generates results after scanning a number of diverse databases and updates its scanning frequency every eight weeks (Finn et al., 2017). Using this biological database, amino acid sequence matches to other families of proteins were determined. Here detailed sequence matches as well as family membership were indicated for both truncated proteins.

3.2.3. Secondary structure predictive analysis

In order to elucidate the predictive secondary structure of the truncated MP and NSP, a web server called 'Predict Protein' (https://www.predictprotein.org/) was employed. This web server used the FASTA input format of the truncated proteins primary amino acid sequences and generated a predictive secondary structure analysis. In order to further explore the secondary structure predictions of the two truncated proteins, an additional web database was employed. The PsiPred database (http:bioinf.cs.ucl.ac.uk.psipred) used the FASTA input format of the truncated proteins (primary amino acid sequences) and generated an annotated grid showing detailed secondary structure.

3.2.4. Tertiary structure, binding sites and disordered region analysis

The Predict Protein webserver was further employed to determine the predicted solvent accessibility (tertiary structure in terms of hydrophobic and hydrophilic regions), protein disorder and protein and nucleotide binding regions. The FASTA input formats for both truncated proteins were put into the webserver and the output results were further analysed.

3.2.5. Building the 3-D models of truncated MP and NSP

In an attempt to create predictive 3-D models of the truncated MP and NSP, the Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) web portal was used. Phyre2 is a suite of tools available on a web portal in order to predict and analyse protein structure, function and mutations. The focus of Phyre2 is to provide biologists with a simple and intuitive interface to state-of-the-art protein bioinformatics tools (Kelley et al., 2009). The FASTA amino acid sequence of the proteins were entered into the database and the predicted 3-D models was modelled on the web portal using the nearest PDB family header. In a further attempt to create 3-D protein models for the truncated proteins, the FASTA amino acid sequences of the proteins were entered into the SWISS-MODEL (https://swissmodel.expasy.org) web server. The SWISS-MODEL webserver has become increasingly popular due to its user-friendly interface and output accessibility (Lesk, 2010).

3.3. Results

3.3.1. Truncation of MP and NSP from full length protein

The annotated images below show the truncated domains chosen for further computational analysis, expression, purification, and structural characterization. The N-ter pilot domain **(*figure* 3.3)** was selected for downstream application and was truncated between amino acids 1-176. The C-ter MP binding domain **(*figure* 3.4)** was selected for downstream application and was truncated between amino acids 171-258.

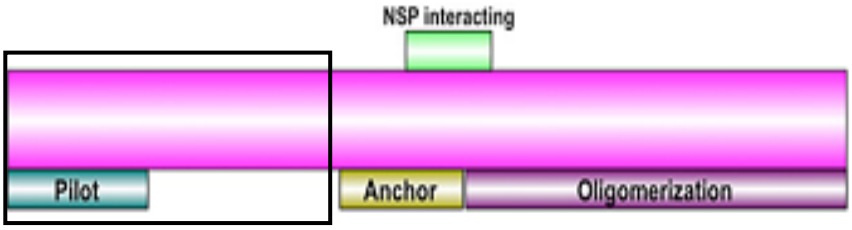


Figure 3.3: The truncated region of SACMV MP, showing the putative domain of interest (amino acids 1-176).

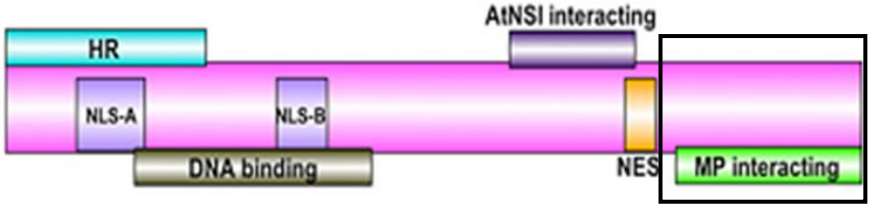


Figure 3.4: The truncated region of truncated NSP, showing the putative domain of interest (amino acids 171- 258).

3.3.2. Determining familial matches of the truncated MP and NSP proteins using InterProScan

The annotation of family homology of the truncated MP **(*figure* 3.5)** indicated that the SACMV truncated MP (pilot domain) shared no family homology match with any other non-geminiviral proteins. The only family homology match found was to that of the geminiviral MPs. This indicates that this domain is not common outside of the general *Geminiviridae* family. The annotation of family homology of the truncated NSP **(*figure* 3.6)** indicated that MP binding domain of the SACMV truncated NSP also shared no

family homology match with any other non-geminiviral proteins. The only familial homology matches were to geminiviral coat proteins. This also indicates that this domain is not found outside of the geminiviruses.

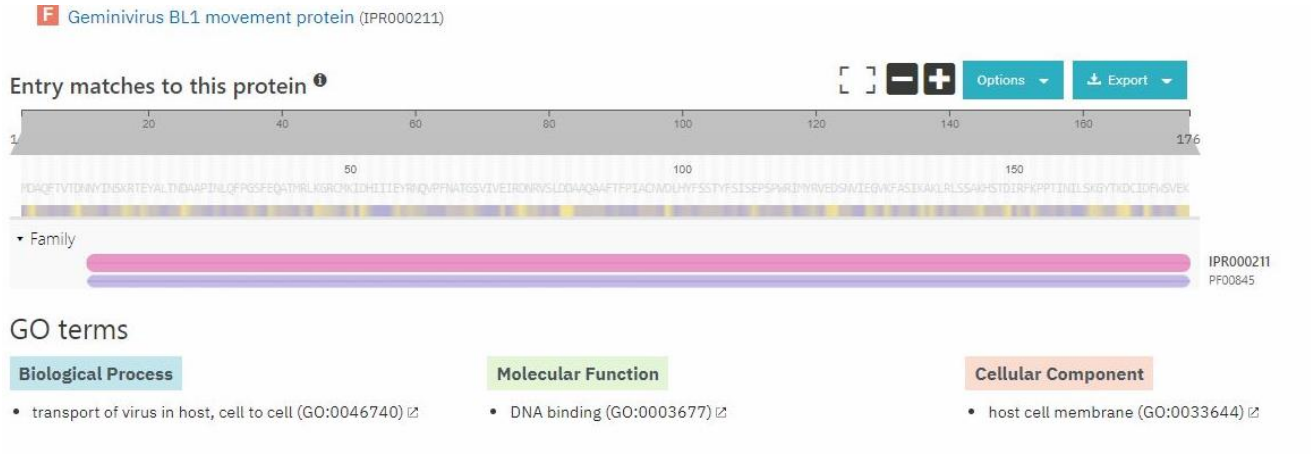


Figure 3.5: Annotation of the family homology of the truncated MP of SACMV from the InterProScan database, showing the family homology match of the truncated MP to the general geminivirus MP family.

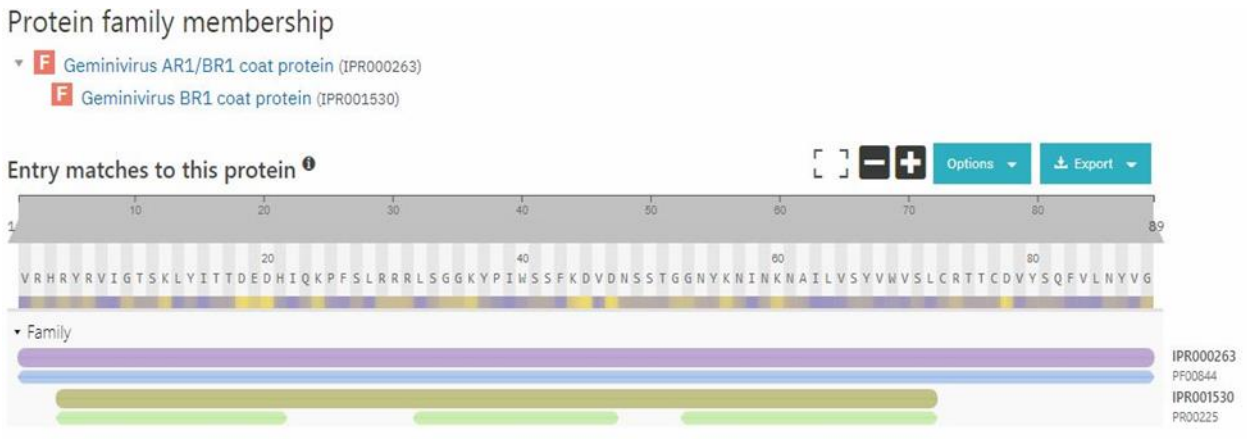


Figure 3.6: Annotation of the family homology of the truncated NSP of SACMV from the InterProScan database, showing the family homology match of the truncated NSP to the general geminivirus coat protein family.

3.3.3. Secondary structure predictive analysis

The secondary structure composition output generated by the 'Predict Protein' webserver **(*figure* 3.7)** indicated that the predictive predominant secondary structure

of the truncated MP is ‘other’ or random coils/loop. However, visually it can be seen that there is a high likelihood that the protein could also contain β-strands/ sheets. The ‘Predict Protein’ webserver also presents a possibility of the proteins secondary structure containing α-helices. The ‘Predict Protein’ webserver output **(*figure* 3.8)** indicated that the predictive predominant secondary structure for the truncated NSP would be evenly distributed between β-strands and ‘other’ or random coils/loop.

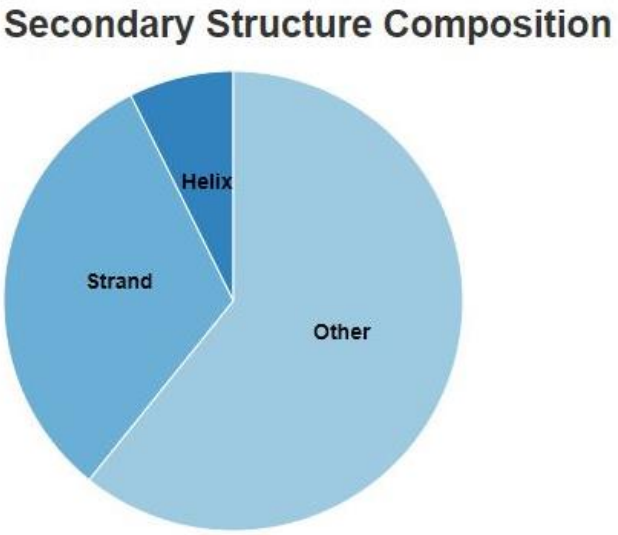


Figure 3.7: The predicted secondary structure composition of the truncated MP of SACMV from the Predict Protein webserver, showing a predominant random coil putative secondary structure.

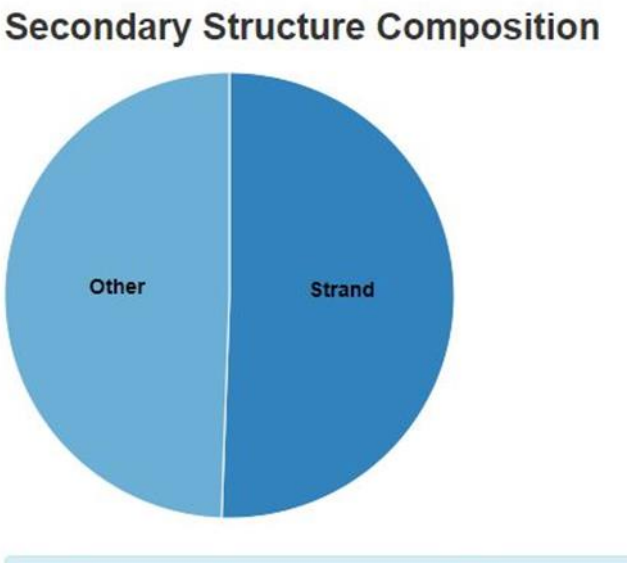


Figure 3.8: The predicted secondary structure composition of the truncated NSP of SACMV from the Predict Protein webserver, showing a predominant random coil and beta strand putative secondary structure.The annotation grid generated from the 'PsiPred' database for the truncated MP **(*figure* 3.9)** further confirms the prediction from 'Predict Protein' as the annotation indicates the predictive secondary composition is predominantly 'coil' followed by a high composition of β-strand. The annotation grid generated from the 'PsiPred' database for the truncated NSP **(*figure* 3.10)** also confirmed the results generated from the 'Predict Protein' webserver. The 'PsiPred' annotation grid confirms that the predictive secondary structure is evenly distributed between a coiled and β-sheet structure.

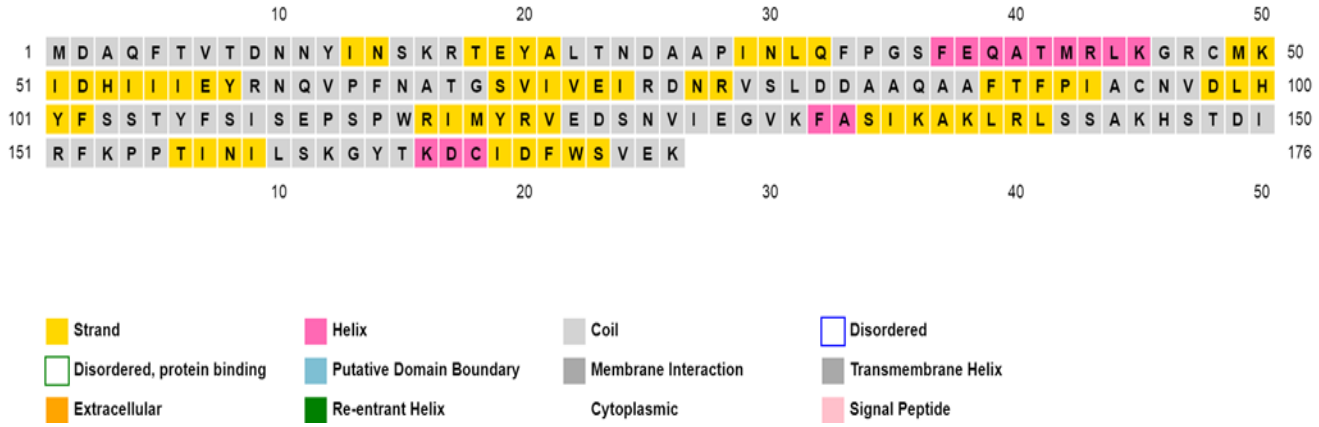


Figure 3.9: An annotation grid of the truncated MP of SACMV from the Psipred database.

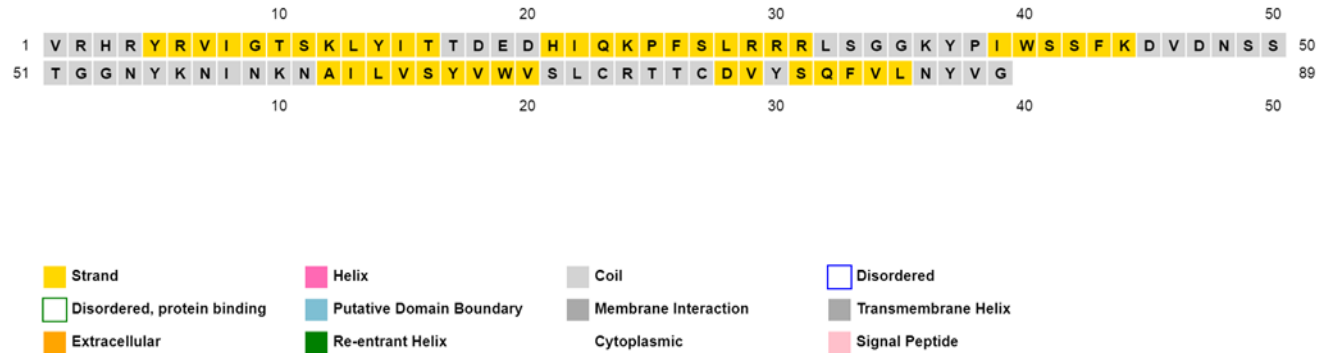


Figure 3.10: An annotation grid of the truncated NSP of SACMV from the Psipred database.

3.3.4. Tertiary structure, binding sites and disordered region analysis

The ‘Predict Protein’ webserver predicted a number of parameters pertaining to the overall computational characterization of the two truncated proteins. The webserver was able to create outputs for the solvent accessibility of the proteins (hydrophobicity), protein and nucleotide binding, and protein disorder. The solvent accessibility annotation of the truncated MP **(*figure* 3.11)** indicates that the predominant predicted tertiary structure of the protein in its native state is exposed or hydrophilic. However, a large surface area of the protein is predicted to be buried/ hydrophobic. While the solvent accessibility annotation of the truncated NSP **(*figure* 3.12)** appears to indicate a similar image, the tertiary structure appears to be more exposed (hydrophilic) than buried (hydrophobic).

The predictive protein binding sites for both truncated proteins are indicated by the purple squares in figures 3.11 and 3.12 below. While the truncated MP has a significantly high protein binding score level of 91 **(*figure* 3.13)**, the truncated NSP has a low protein binding score level of 14 **(*figure* 3.14)**. This indicates that the likelihood of proteins binding to the truncated MP in the highlighted regions are high, while the likelihood of proteins binding to the truncated NSP in the highlighted regions are less likely. Both annotations indicated that neither truncated NSP or MP proteins bind nucleic acids (DNA/RNA).

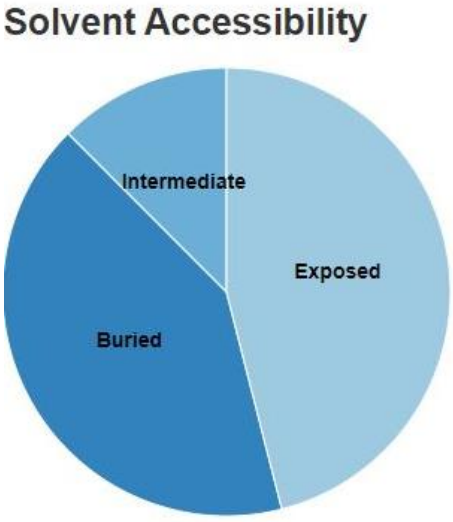


Figure 3.11: The predicted solvent accessibility of the truncated MP of SACMV from the Predict Protein webserver.

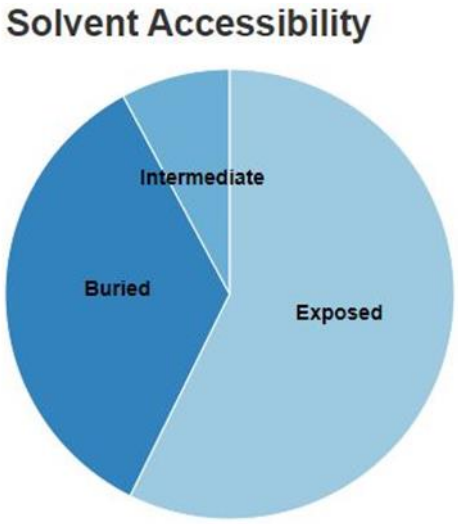


Figure 3.12: The predicted solvent accessibility of the truncated NSP of SACMV from the Predict Protein webserver.

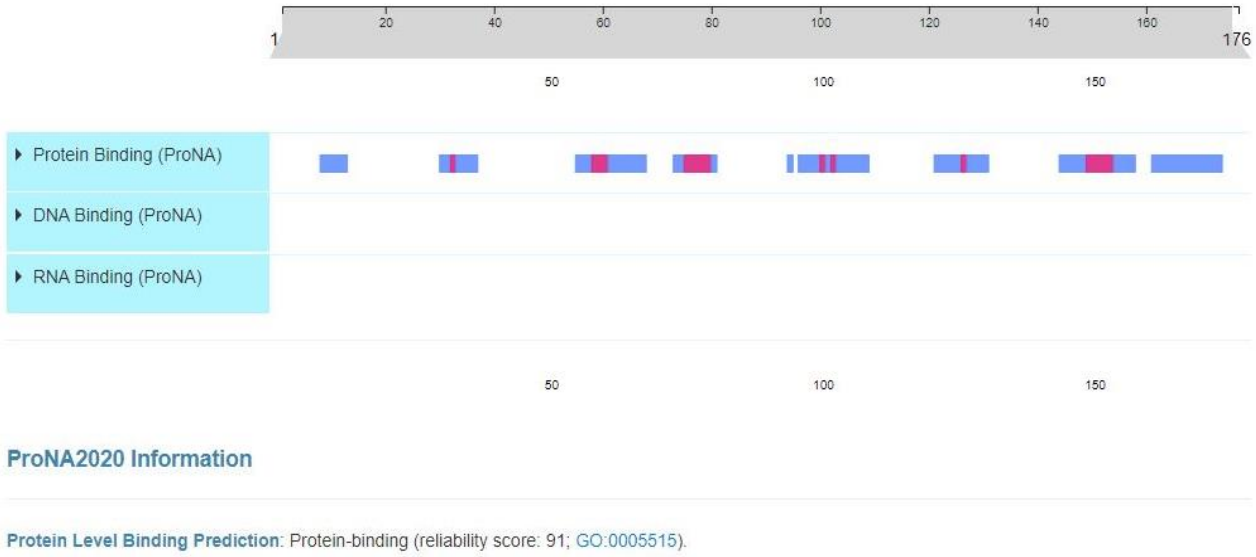


Figure 3.13: The predicted binding sites of the truncated MP (protein, DNA and RNA) of SACMV as predicted by the Predict Protein webserver.

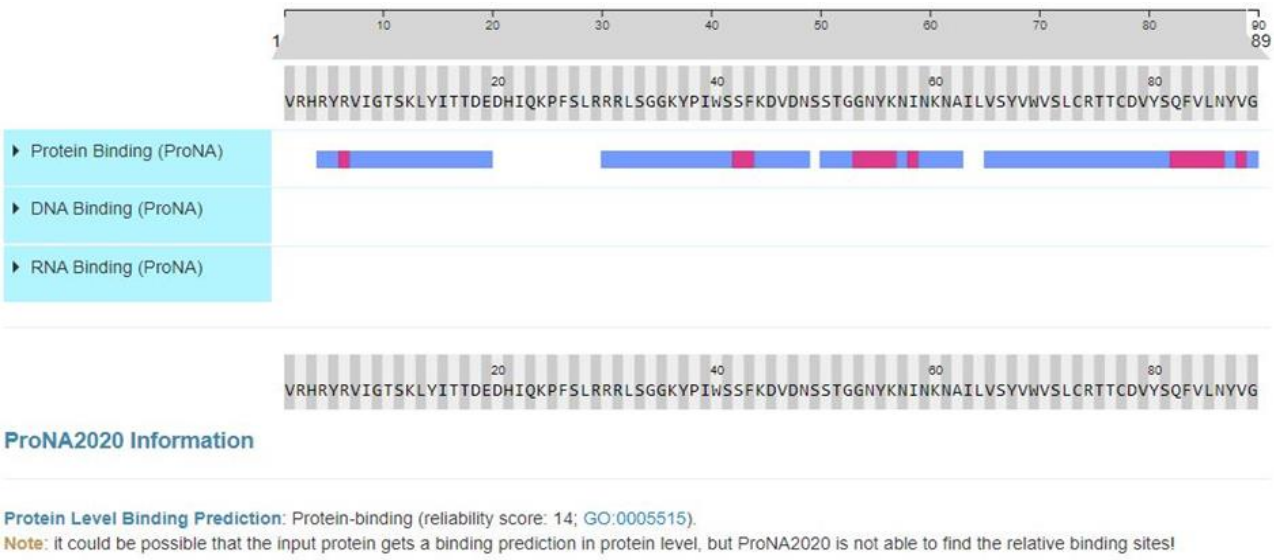


Figure 3.14: The predicted binding sites of the truncated NSP (protein, DNA and RNA) of SACMV as predicted by the Predict Protein webserver.

The predicted protein disorder annotations **(*figures* 3.15 and 3.16)** show high levels of predicted disorder for both truncated proteins. This can be correlated to the proteins' secondary structure analysis which predicts that both proteins contain a predominant 'other' or random coil secondary structure.

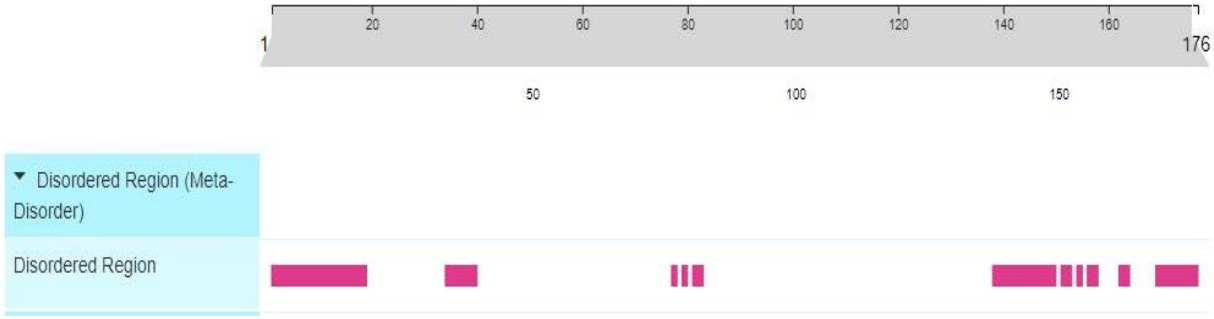


Figure 3.15: The predicted disordered regions of the truncated MP of SACMV as predicted by the Predict Protein webserver.

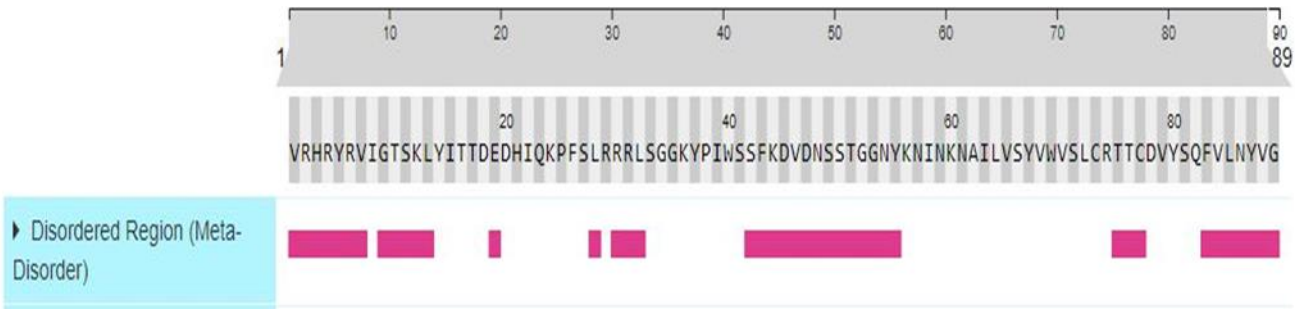


Figure 3.16: The predicted disordered regions of the truncated NSP of SACMV as predicted by the Predict Protein webserver.

3.3.5. Building the 3-D models of truncated MP and NSP

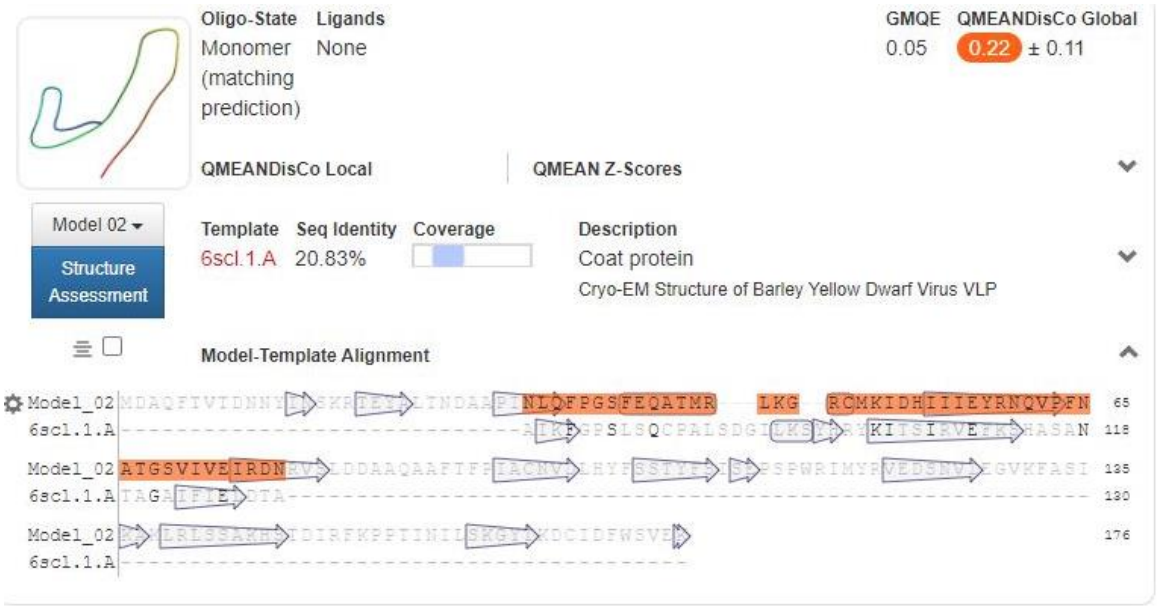
In an attempt to build a 3D model for the truncated MP or N-ter of the MP, two online tools were used, namely Phyre2 and SWISS-MODEL. The Phyre2 web tool, was only able to model the truncated MP **(*figure* 3.17)** using 11% of the query sequence. The 11% of this sequence was only modelled with 57.5% confidence and the Protein Data Bank linked the protein/ sequence to a transcription factor of a saga core molecule. In contrast, the SWISS-MODEL webserver built a 3D model based on 20.83% sequence identity to the cryo EM structure of *Barley Yellow Dwarf virus* **(*figure* 3.19)**.

The Phyre2 web tool and the SWISS-MODEL webserver both created a predicted 3D model based on the sequence identity (PDB file header) to the cryo EM structure of *Ageratum yellow vein virus* **(*figures* 3.18 and 3.20)**. the Phyre2 web tool was able to model 69% of the sequence with 92.2% confidence while the SWISS-MODEL webserver was able to match and model with a 19.35% sequence identity.

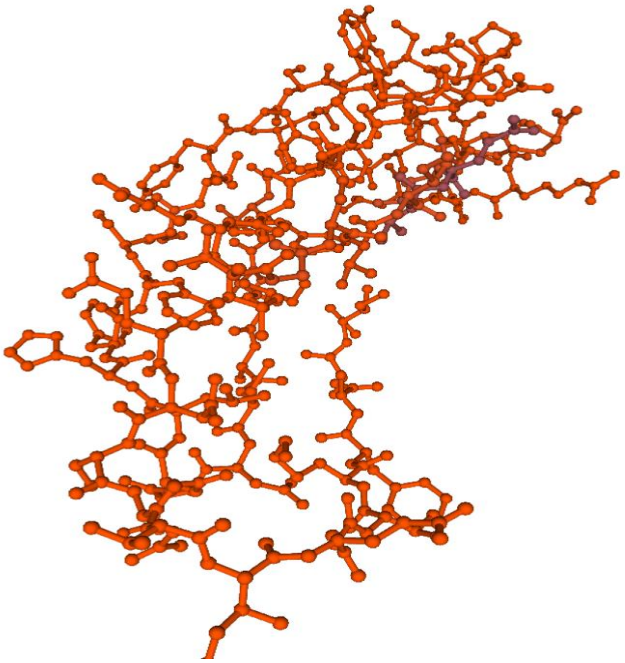
Figure 3.17: Partial predicted model of truncated MP of SACMV based on 11% coverage of the truncated protein, modelled with 57.5% confidence from the Phyre2 protein modelling site.

Figure 3.18: Partial predicted model of truncated NSP of SACMV based on 69% coverage of the truncated protein, modelled with 92.2% confidence from the Phyre2 protein modelling site.

A.



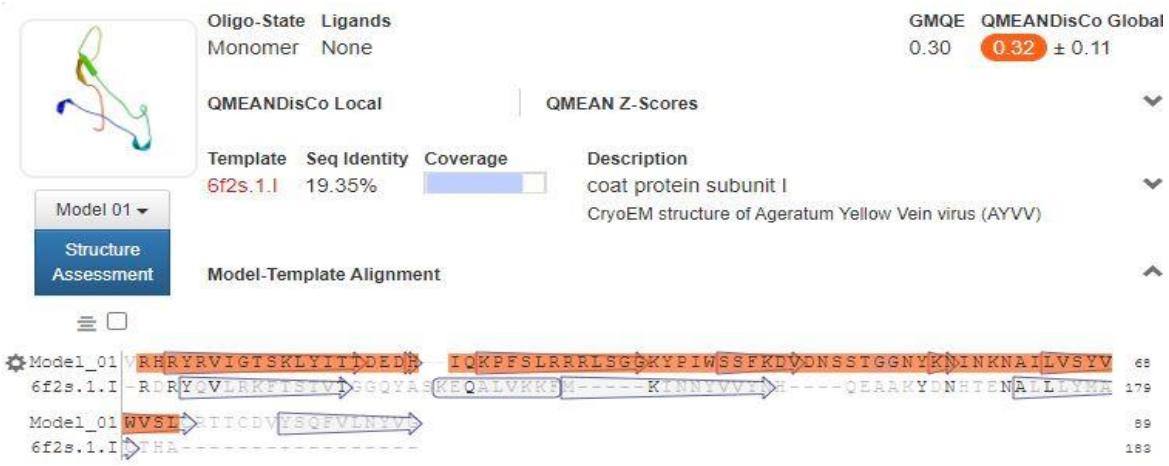
B.



.

Figure 3.19: (A) Model template alignment for the truncated MP of SACMV as aligned against the Cryo-EM structure of Barley Yellow Dwarf Virus VLP, (B) predicted model of the truncated MP of SACMV as predicted by SWISS-MODEL.

A.



B.

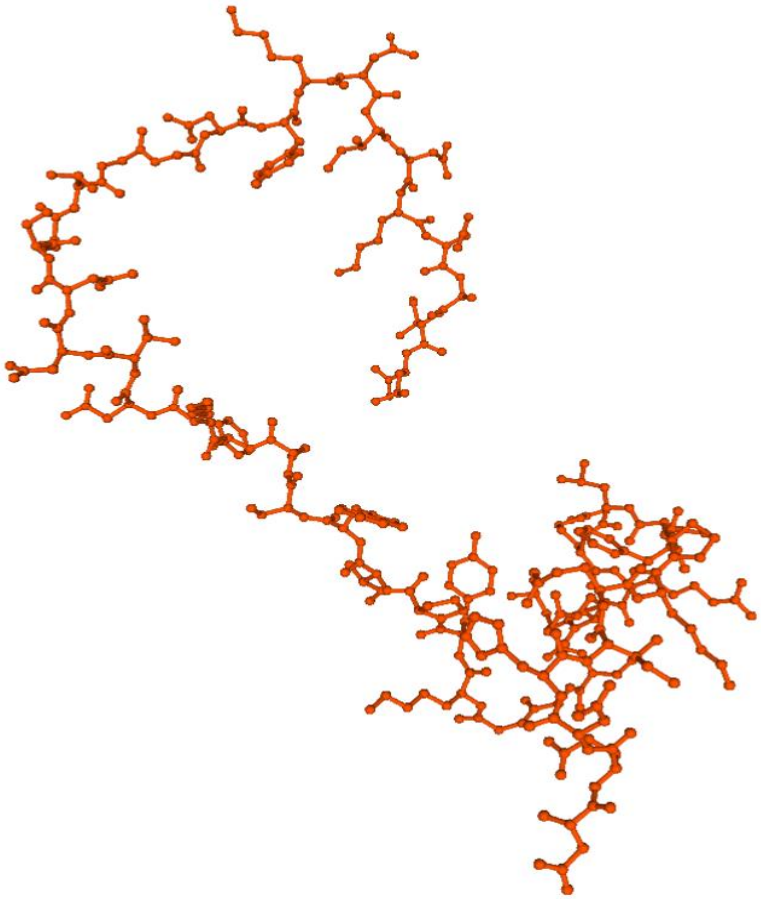


Figure 3.20: (A) Model template alignment for the truncated NSP of SACMV as aligned against the Cryo-EM structure of Ageratum Yellow Vein Virus (AYVV), (B) predicted model of the truncated NSP of SACMV as predicted by SWISS-MODEL.

3.4. Discussion

Predictive structural studies for the specific domains of the MP and NSP have never been studied before, and this makes computational studies useful in further understanding these proteins. While computational studies have been undertaken on the full length SACMV MP (Nankoo, 2018), none have been done on specific domains. No computational studies have been undertaken to date on the SACMV NSP or any of its domains. InterProScan was used to examine if the newly truncated proteins shared any similarities with any other protein families which could possibly share similar predicted functions such as the cell-to-cell and long-distance movement for the MP and NSP (Fondong, 2013). However, the results from this computational study did not reveal any family homology between the truncated MP (N-ter) and any other protein family except the geminivirus movement protein **(*figure* 3.5)**. This result is consistent with the results attained in the study of the full length SACMV MP where no family homology was found except with that of the geminivirus movement protein family (Nankoo, 2018). The results attained from InterProScan indicated that the truncated NSP (C-ter) shared no family homology with any other protein family besides the geminivirus coat protein family **(*figure* 3.6)**. The family homology matching between the SACMV truncated NSP and the geminivirus coat protein family could be explained since in monopartite geminiviruses the coat protein (CP), which is encoded by DNA A, is essential for virus movement (Unseld et al., 2001). This would support a common function, and possible similar host interacting proteins, for NSP and CP in bipartite and monopartitie geminiviruses, respectively.

Secondary structure analysis for the truncated MP indicated that the predominant predicated secondary structure for the protein would be 'other' **(*figure* 3.7)** which is defined by the PredictProtein webserver as random coils/loops. However, the prediction also indicates that the protein contains a high β-sheet predictive secondary structure. These results were further confirmed by the output results generated by the PsiPred database which produced the same predictive secondary structure **(*figure* 3.9)**. These results coincide with the predictive results attained for the full length MP where it was found that the predictive structure of the protein was predominantly random coils/loop and to a lesser extent, β-sheets (Nankoo, 2018).

The secondary structure analysis for the truncated NSP indicated that the predominant secondary structure for the protein would be random coils/loop and β-sheets

(*figure* 3.8). This prediction was further confirmed by the output prediction from the PsiPred database which confirmed that the truncated NSP protein possessed a predictive secondary structure of both random coils/loop and β-sheet. The predictive results attained are not uncommon as viral proteins often do not contain a well-defined secondary structure such as α-helix or β- strands but instead appear as random coils (Linding et al., 2003).

The hydrophobicity or solvent accessibility of a protein is one of the major forces in the stabilisation of protein folding. Prediction of the hydrophobicity of the specific novel protein allows for some determination of the tertiary structure of the protein. This since the 3D folding of the polypeptide chain of the protein folds in such a manner that in most cases the interior of the protein consists predominantly of hydrophobic, non-polar amino acids while polar, hydrophilic amino acids are found predominantly on the outside of the molecule (Wilson and Walker, 2012). The solvent accessibility of the truncated MP was predicted to be predominantly exposed/hydrophilic **(*figure* 3.11)** however, the buried/hydrophobic region was also predicted to be quite prevalent. This indicates that in the native form, it is still likely that ligands will bind to the buried/hydrophobic regions (Wilson and Walker, 2012). The solvent accessibility of the truncated NSP was predicted to be more exposed/ hydrophilic with less buried/ hydrophobic regions available **(*figure* 3.12)** for possible ligand binding.

In order for a proteins successful function, the protein must interact with other biological molecules such as nucleotides (DNA and/or RNA) and proteins. This allows for them to interact with these molecules upon binding to them. The predicted binding sites on both truncated proteins displayed by the 'Predict protein' database **(*figures* 3.13 and 3.14)** indicate a number of protein binding sites spanning across the domains of the two proteins. This supports the proposal that in regions where proteins are disordered, those regions that are more flexible are able to bind more readily to targets compared to ordered regions (Voet et al., 2013).

***Figures* 3.15 and 3.16** demonstrate that both the truncated MP and NSP are predicted to have regions of disorder. These regions of disorder seem to coincide with the regions of both proteins which are predicted to have random coils in their predicted secondary structure **(*figures* 3.9 and 3.10)**. The predicted disordered regions of a protein are indicative of the regions of the protein which hypothetically lack an ordered

or fixed 3D structure in the protein backbone. It is suggested that disordered regions of proteins give the disordered segments of the proteins more flexibility, allowing the disordered segment to perform relatively unhindered conformational searches when binding to their target molecules (Voet et al., 2013).

Phyre 2 is a protein homology/analogy recognition engine. The server uses a library of known protein structures from the SCOP (Structural Classification of Proteins) database and is supplemented with newer depositions in the Protein Database (PDB) (Kelley et al., 2009). The predictive 3D model created from the Phyre2 modelling site for the truncated MP **(*figure* 3.17)** was modelled with only 11% of the sequence which modelled with 57.5% confidence, which is extremely low for protein modelling. The sequence found similarity to transcription factor 20 within the PDB. In contrast the SWISS-MODEL site, modelled with 20.83% sequence identity to the coat protein of Barley Yellow Dwarf virus **(*figure* 3.19)**. SWISS-MODEL attained a more reliable model as it modelled against a similar geminivirus coat protein. The similarity of the SACMV MP to the Barley Yellow Dwarf virus is substantiated as the coat protein is essential for virus movement in monopartite geminiviruses (Unseld et al., 2001). The predictive 3D model created from Phyre2 for the truncated NSP **(*figure* 3.18)** was modelled with 69% coverage of the protein sequence with 92.2% confidence, which is a promising model, more so as the sequence similarity was to the capsid protein of ageratum yellow vein virus. The 3D model created by the SWISS-MODELL database produced similar results, with 19. 35% of the truncated NSPs sequence identity to the ageratum yellow vein virus **(*figure* 3.20)**. The 3D models generated for the NSP appear more consistent than those for the MP, showing high sequence similarity with the capsid protein of the geminivirus, Ageratum yellow vein virus.

3.5. Conclusions

Since little information is available on the structural aspects of the individual domains of the MP and NSP encoded by DNA B of geminiviruses, employing predictive computational approaches using a valuable resource of internet web servers, web portals and databases can be useful (Roche et al., 2011; Prajapat et al.,2010; Prajapat et al., 2011; Rost and Liu, 2003). These computational approaches may also be useful in accompanying *in vitro* experimental analysis (Prajapat et al., 2010).

Using computational approaches to predict protein structure or function may provide some insight as to what may be to expect when structurally characterizing the protein *in vitro*. The use of publicly available databases and webservers such as 'Predict Protein', 'PsiPred', Phyre2 and SWISS-MODEL have provided invaluable results for both the truncated MP and truncated NSP. While some results obtained during computational studies could sometimes be disputed, some results are validated *in vitro* and *in vivo* (Kelley et al., 2009; Rost and Lui., 2003).

Chapter 4

Expression and Purification of truncated SACMV MP and NSP

4.1. Background

The movement and nuclear shuttle proteins of *South African cassava mosaic virus* (SACMV) are novel proteins. Very little is known about the structure of both these proteins in geminiviruses. While some research has been done on the functionality of the nuclear shuttle protein (NSP) in geminiviruses, less focus on function in vivo has been done on the movement protein (MP).

In order to determine a proteins function, the structure of the protein has to be unravelled. However, before the structural characterization of a protein can ensue, the recombinant protein of interest must first undergo expression in a suitable host expression system before the protein is purified for downstream applications (Kaur et al., 2003; Yokoyama, 2003). In order for successful recombinant protein expression to occur, the gene of interest is cloned into a suitable expression vector which includes all the required components for the expression of foreign genes. The most crucial components of an expression vector include a promotor, operator, ribosome binding site, N-terminal tag, start codon, multiple cloning site, C-terminal tag, transcription terminal and a selectable marker (Kaur et al., 2003). Once the gene of interest is successfully cloned into the expression vector of choice, a suitable expression host is chosen. There are many expression host systems available for recombinant protein expression. However, the most popular host expression system used is the prokaryotic *Escherichia coli* (*E.coli*) (Yokoyama, 2003). Upon selection of an appropriate expression host strain and successful transformation of the expression vector, a panel of induction conditions are selected in order to produce the recombinant protein, often these conditions require optimization to improve yield and solubility of the recombinant protein (Kaur et al., 2003).

Proteins that are successfully expressed, undergo the process of purification before proceeding to downstream applications. Protein purification using immobilized-metal-ion exchange chromatography (IMAC) has become an increasingly popular method. This can be attributed to the increasingly popularity of engineering recombinant proteins which have poly-histidine tags on either the N- or C- terminals (Gutiérrez et

al., 2007). The purification of proteins using IMAC is a technique which is advantageous in the purification of his-tagged recombinant proteins, allowing for the separation of the his-tagged recombinant protein from the rest of the expression host proteins. The his-tagged recombinant protein binds to the metal charged matrix while all other expression host proteins elute through the column. Low concentrations of chemicals such as imidazole and EDTA are used as competitive agents to elute the his-tagged recombinant protein off the column. Once eluted the protein may be used for downstream applications such as structural characterization, depending on the purity of the protein attained off the IMAC column (Gutiérrez et al., 2007).

This chapter specifically focusses on the expression and purification of a truncated of the SACMV MP and NSP. Both these proteins can be found on the DNA B molecule of SACMV. Expression and purification is crucial for downstream structural characterization and downstream applications.

4.2. Materials and Methods

4.2.1. Cloning of the MP and NSP genes into the pCOLDI vector

The N-terminus (528 bp) and C-terminus (267 bp) truncated sequences of the SACMV MP **(*figure* 4.1)** and NSP **(*figure* 4.2)** genes, respectively, were sent to GenScript USA for synthesis (codon- optimized). These genes were cloned into the expression vector system pColdI **(*figure* 4.3)**. The restriction sites NdeI and BamHI were used to clone the truncated MP gene into the expression vector, while the restriction sites SacI and BamHI were used to clone the NSP gene into the expression vector.

ATGGACGCCCAATTTACCGTCACAGACAATAATTACATCAACAGCAAACGCACCGAG
TACGCATTAACAAACGATGCTGCACCAATCAATCTCCAATTCCCAGGCTCATTCGAGC
AGGCTACCATGCGACTCAAAGGCCGATGTATGAAAATCGACCACATTATAATTGAAT
ACCGAAACCAGGTCCCATTTAACGCAACTGGGTCGGTTATCGTAGAAATCAGAGACA
ATCGCGTCAGCCTTGACGACGCAGCTCAGGCAGCATTCACTTTCCCCATCGCTTGCAA
CGTGGACCTCCACTACTTCTCTTCTACATATTTTTCGATTTCTGAACCTTCCCCTTGGAG
AATCATGTACAGAGTCGAAGACTCAAACGTCATAGAAGGCGTGAAATTCGCATCCAT
CAAGGCCAAGCTCCGATTATCATCGGCCAAACATTCCACGGACATACGTTTCAAACCC
CCAACAATTAACATCTTATCCAAGGGATACACAAAAGACTGCATAGACTTCTGGTCCG
TGGAAAAA

Figure 4.1: The 528 bp gene sequence of the SACMV MP N-terminus truncation.

GTGAGGCACAGGTACAGGGTGATCGGCACCAGCAAGCTGTACATCACCACCGACGA
GGACCACATCCAGAAGCCCTTCAGCCTGAGGAGGAGGCTGAGCGGCGGCAAGTACC
CCATCTGGAGCAGCTTCAAGGACGTGGACAACAGCAGCACCGGCGGCAACTACAAG
AACATCAACAAGAACGCCATCCTGGTGAGCTACGTGTGGGTGAGCCTGTGCAGGAC
CACCTGCGACGTGTACAGCCAGTTCGTGCTGAACTACGTGGGC

Figure 4.2: The 267 bp gene sequence of the SACMV NSP C-terminus truncation.

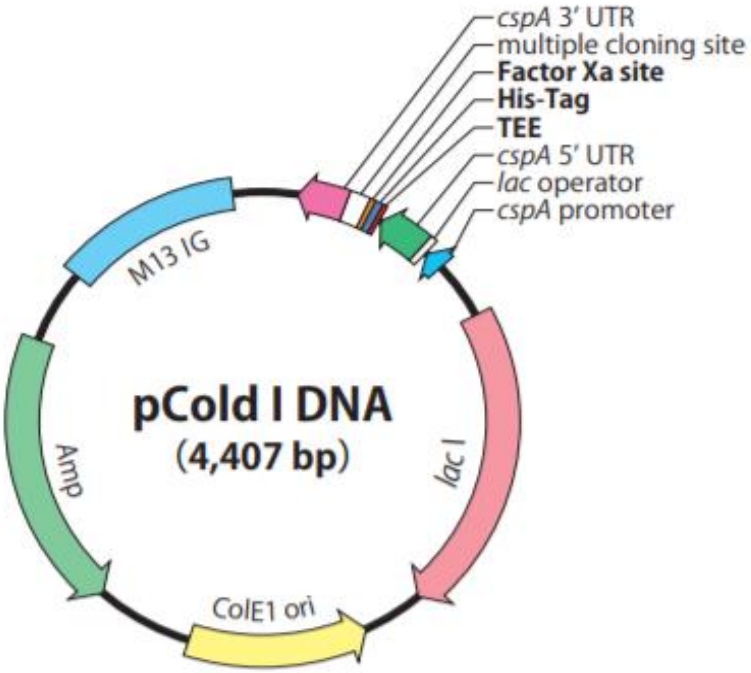


Figure 4.3: pColdI vector map showing annotations (Source:Takara Bio Inc).

4.2.2. Transformation of MP and NSP gene inserts into competent *E.coli* cells and gene insert validation

The lyophilized cells (4 µg) (GeneScript USA) were dissolved in sterile nuclease free water, making up a working concentration of 100 ng/ µl of cells. These cells were transformed via the GeneScript USA heat shock transformation protocol **(*figure* 4.4)**.

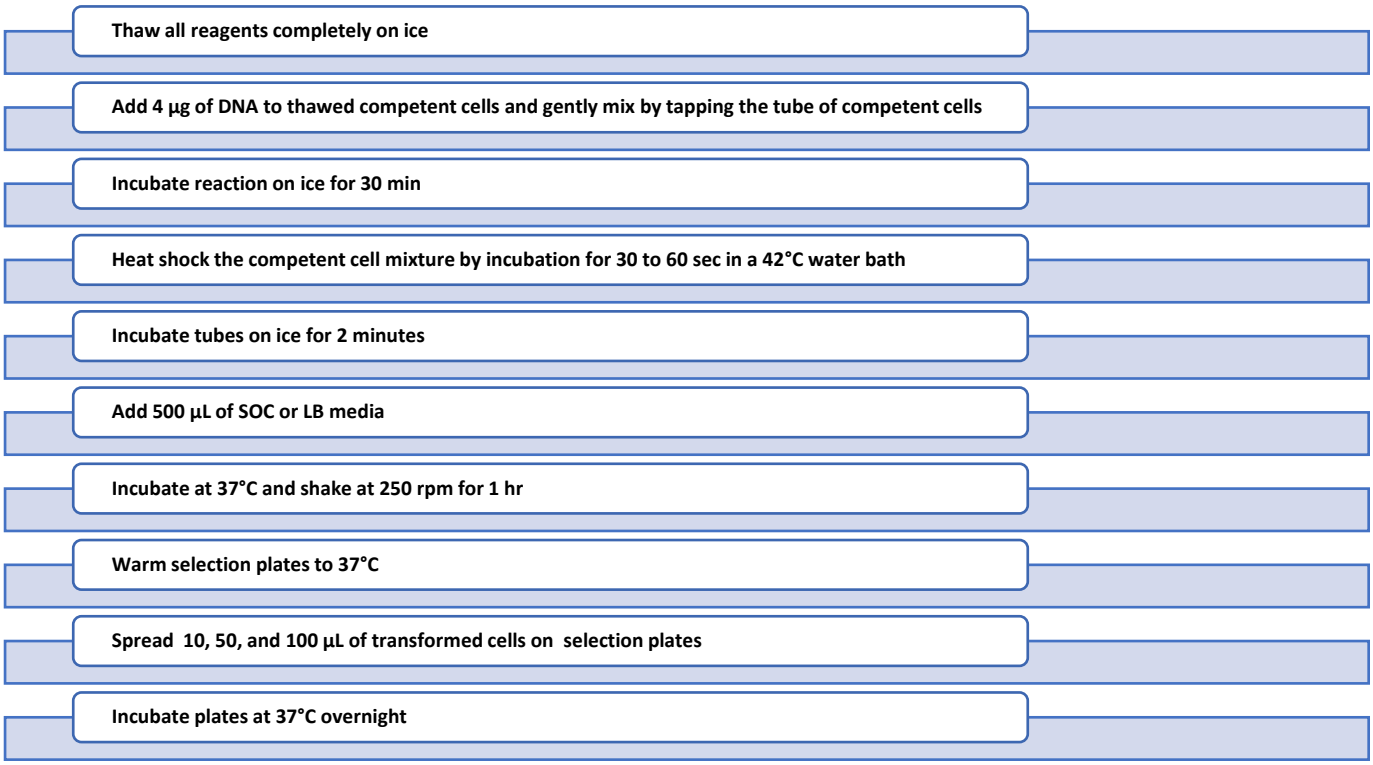


Figure 4.4: Schematic diagram showing the heat shock transformation protocol adapted from GenScript, USA.

The MP cells were transformed into BL21(DE3) pLysS *E.coli*, while the NSP cells were transformed into BL21(DE3) *E.coli*. The transformed colonies were selected and grown in liquid Luria Broth (LB) [10 % (w/v) tryptone (Sigma Aldrich, USA), 5 % (w/v) yeast extract (Sigma Aldrich, USA), 10 % (w/v) NaCl (Sigma Aldrich, USA) and 100 µg/mL ampicillin (Sigma Aldrich, USA)] overnight for further plasmid extraction and glycerol stock preparation. From the prepared glycerol stocks the gene inserts were validated by restriction digests using the respective restriction enzymes used for cloning **(*table* 4.1 and 4.2).**

Table 4.1: Table showing reagents used for gene validation of the MP truncated insert.

Reagent	Volume
Nuclease free distilled water (Sigma Aldrich ,USA)	40 µl
10 X Cut Smart buffer (New England Biolabs Inc)	5 µl
NdeI	1 µl
BamHI	1 µl
Plasmid DNA making up a concentration of 1 µg	3 µl
Total	**50 µl**

Table 4.2: Table showing reagents used for gene validation of the NSP truncated insert.

Reagent	Volume
Nuclease free distilled water (Sigma Aldrich ,USA)	38.5 µl
10 X Cut Smart buffer (New England Biolabs Inc)	5 µl
SacI	1 µl
BamHI	1 µl
Plasmid DNA making up a concentration of 1 µg	4.5 µl
Total	**50 µl**

4.2.3. Protein overexpression

4.2.3.1 Truncated MP overexpression

Overnight cultures of transformed *E.coli* BL21(DE3) plysS were prepared in 50 ml falcon tubes (Lasec, SA) in LB media [10 % (w/v) tryptone (Sigma Aldrich, USA), 5 % (w/v) yeast extract (Sigma Aldrich, USA), 10 % (w/v) NaCl (Sigma Aldrich, USA)] supplemented with 100 µg/ml^{-1} ampicillin (Sigma Aldrich, USA) and 50

µg/mLchloroamphenicol (Sigma Aldrich, USA). The cultures were placed in a 37 °C shaking incubator (United Scientific), the cultures were allowed to shake for no longer than 16 h at a shaking speed of 200 rpm. Following overnight incubation, the cultures were removed from the incubator and centrifuged at 15 000 xg for 20 min. The supernatants were discarded and the pellets were re-suspended in fresh LB medium supplemented with 100 µg/mL ampicillin and 50 µg/mL chloroamphenicol (total volume of 100 ml). The OD_{600} of the cultures were measured on a spectrophotometer (United Scientific) and were measured around ~ 0.1-0.2. The cultures were then placed in a 37 ° C shaking incubator and allowed to shake at 200 rpm for 2 h until the OD_{600} reached 0.45. The cultures were then induced with 0.50 mM of IPTG (Merck, Germany) and allowed to shake at 150 rpm for 24 h at 15**°C** ° in a shaking incubator in order to allow the protein to express. Following the 24 h expression period, the cultures were removed from the incubator and centrifuged at 25 000xg for 20 min. The supernatants were discarded and pellets were air dried and stored at -80**°C** ° for further use.

4.2.3.2 Truncated NSP overexpression

Overnight cultures of transformed *E.coli* BL21(DE3) were prepared in 50 ml falcon tubes (Lasec, SA) in LB media [10 % (w/v) tryptone (Sigma Aldrich, USA), 5 % (w/v) yeast extract (Sigma Aldrich, USA), 10 % (w/v) NaCl (Sigma Aldrich, USA)] supplemented with 100 µg/mL ampicillin (Sigma Aldrich, USA). The cultures were placed in a 37 ° C shaking incubator (United Scientific), the cultures were allowed to shake for no longer than 16 h at a shaking speed of 200 rpm. Following overnight incubation, the cultures were removed from the incubator and centrifuged at 15 000 xg for 20 min. The supernatants were discarded and the pellets were re-suspended in fresh LB medium supplemented with 100 µg/mL ampicillin (total volume of 100 ml). The OD_{600} of the cultures were measured on a spectrophotometer (United Scientific) and were measured around ~ 0.1-0.2. The cultures were then placed in a 37**°C**shaking incubator and allowed to shake at 200 rpm for 1.5 h until the OD_{600} reached 0.40. The cultures were then induced with 0.25 mM of IPTG (Merck, Germany) and allowed to shake at 150 rpm for 24 h at 15**°C** in a shaking incubator in order to allow the protein to express. Following the 24 h expression period, the cultures were removed from the incubator and centrifuged at 25 000xg for 20 min.

The supernatants were discarded and pellets were air dried and stored at -80**°C** for further use.

4.2.4. Protein extraction

To thawed protein pellets, 5 ml of lysis buffer [50 mM Tris HCl (Sigma Aldrich USA), 5% Glycerol (SRL chemicals, India), 250 mM NaCl (Sigma Aldrich, USA) and 5 mM $MgCl^2$ (Merck, Germany) pH 8.0] was added before sonicating (Sonicator- Microson Ultrasonic cell-disrupter, Lasec, SA) the cells [3 cycles, 30 sec on and 30 sec pause]. A protease inhibitor PMSF [1 mM] was added to the protein pellets before the cells were exposed to another 3 cycles of sonication. This was followed by the addition of lysozyme [1 mg/ml] (Sigma Aldrich, USA) to the cells which was followed by another 3 cycles of sonication. The cells were then left to incubate on ice for 20 min on a shaking platform [150 rpm] (United Scientific). This was followed by centrifugation at 25 000 xg for 20 min at 4˚C. The soluble fraction was kept for assessment on a 12 % SDS-PAGE gel for protein solubility (or stored at - 20˚ C for further use). While the insoluble fraction was washed in 0.3 parts lysis buffer and then washed with 0.3 parts of 1% SDS (Sigma Aldrich, USA), the washed insoluble fraction was then assessed on a 12 % SDS-PAGE gel to determine protein insolubility or stored for further use at -20˚C.

4.2.5 Protein denaturation

As will be indicated in section 4.3 (Results), both these truncated proteins are insoluble once expressed, therefore it was imperative to denature, purify and then subsequently refold these proteins. The insoluble protein pellets (4.2.4) were treated to 8 M urea solution [8 M urea (Sigma Aldrich, USA) and 50 mM Tris HCl (Sigma Aldrich, USA) pH 8.5] to a final volume of 1.25 % (v/v). The samples were then sonicated before being centrifuged at 25 000 xg for 20 min at 4**°C**. The supernatants were then dialysed in SnakeSkin dialysis tubing (10k MWCO, 16 mm) (ThermoFischer Scientific, USA) overnight at 4**°C** (separately for the two proteins (200 X the final supernatant volume)) in 4 M urea solution [4 M urea (Sigma Aldrich, USA), 20 mM sodium phosphate (Sigma Aldrich, USA), 0.5 M NaCl (Sigma Aldrich, USA) and 10 mM imidazole (SRL chemicals, India) pH 7.5].

4.2.6. Protein purification

Protein purification for both the truncated MP and NSP was carried out by gravity flow affinity chromatography. The resin used was a pre-charged nickel resin (Profinity™ IMAC resin, BIORAD, USA). In order to prepare the columns, 2 ml of the 50 % nickel resin slurry was added to each column. The storage buffer was removed from each of the columns by gravity flow elution. The resins from each column were washed with 5 column volumes of milli-q water (5 ml). The column resins were equilibrated with 3 column volumes each of dialysate (4 M urea dialysis buffer). Each of the proteins were loaded into the separate columns (5 ml in total per protein), the columns were then incubated at 4**°C** on a shaking platform for 30 min. Following the 30 min incubation, the flow through from the two separate columns were collected in 3, 2 ml fractions. The resins were washed with 5 column volumes (5 ml) of dialysate (25 mM imidazole in 4 M urea dialysis buffer pH 7.7) and collected in 2 ml fractions. The proteins were then eluted using the respective elution buffers (250 mM imidazole in 4 M urea dialysis buffer pH 7.5) and collected in 2 ml fractions.

4.2.7. Protein refolding

The eluted fractions for each of separate proteins (MP and NSP) were pooled together (based on protein purity) and diluted 15 X into 4 M urea solution. The proteins were then refolded using stepwise dialysis. The proteins were dialysed in the first overnight dialysis buffer (separately per protein) (50 mM Tris-HCl (Sigma Aldrich, USA) pH 10, 0.5 M L-arginine (SRL chemicals, India), 2% (v/v) Triton X-100 (SRL chemicals, Delhi, India), 10 mM DTT (SRL chemicals, Delhi, India)). The proteins were then dialysed overnight into the next dialysis buffer (50 mM Tris HCl pH 10, 0.5 M L-arginine). This overnight dialysis step was subsequently repeated.

Following protein refolding, the proteins were removed from the dialysis tubing and placed in protein concentration and ultrafiltration columns (Amicon® ultra centrifugal columns, Merck, Germany). The columns were centrifuged at 15 000 xg for 25 min at 4**°C**. The remaining protein that did not pass through the columns were stored at 4**°C** for downstream protein quantification and characterization.

4.3. Results

4.3.1. Cloning of the MP and NSP genes into the pCOLDI vector

The SACMV MP and NSP truncated codon-optimized genes were successfully cloned into the expression vector pCOLDI and met all quality controls (certificates provided by GenScript USA, provided in **Appendix A**.

4.3.2. Transformation of MP and NSP gene inserts into competent *E.coli* competent cells and gene insert validation

***Figure* 4.4** represents LB agar plates supplemented with ampicillin and chloroamphenicol antibiotics (plates A-C and D). The colonies present on the agar plates A-C (sample triplicate) are indicative of the positive clones from the transformation of the truncated MP gene cloned in the pCOLDI expression vector into the BL21 DE3 pLysS *E.coli*. LB agar plate (D) is a positive control containing the antibiotic chloroamphenicol only. The last agar plate (E) is a negative control plate containing no antibiotics and no truncated MP gene cloned into the pCOLDI expression vector. Transformation of the truncated MP-pCOLDI into the BL21 DE3 pLysS *E.coli* strain was successful.

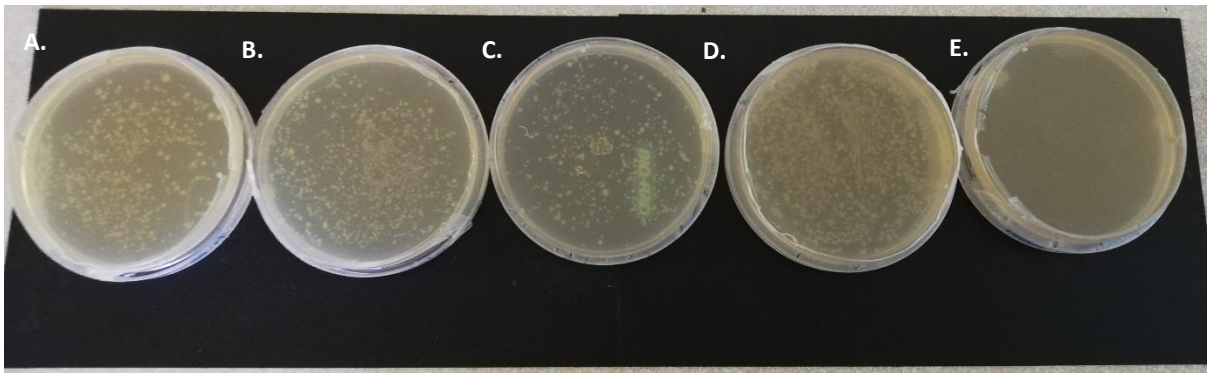


Figure 4.4: LB (ampicillin + chloroamphenicol) agar plates, showing transformed MP-pCOLDI and BL21 DE3 pLysS clones.

Transformation of the truncated NSP-pCOLDI into the BL21 DE3 *E.coli* strain was successful. ***Figure* 4.5** represents LB agar plates supplemented with ampicillinantibiotic (plates B, D and E). The colonies present on the agar plates B, D and E (sample triplicate) are indicative of the positive clones from the transformation of the truncated NSP gene cloned in the pCOLDI expression vector into the BL21 DE3 *E.coli*. LB agar plate (A) is a positive control containing no antibiotic. The middle agar

plate (C) is a negative control plate containing no antibiotics and no truncated NSP gene cloned into the pCOLDI expression vector.

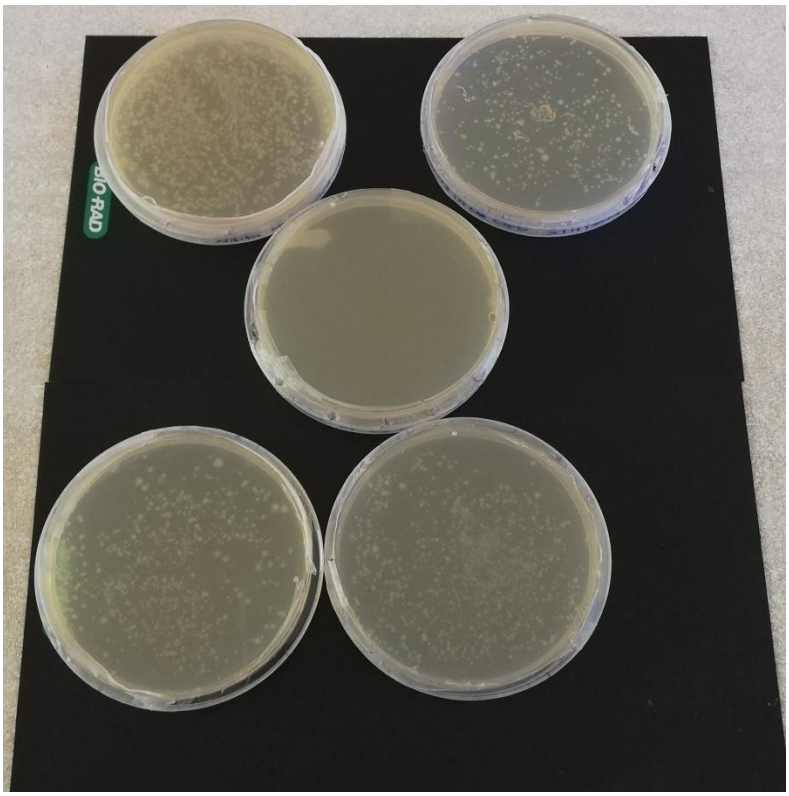


Figure 4.5: LB (ampicillin) agar plates, showing transformed NSP-pCOLDI and BL21 DE3 *E.coli* clones.

In order to confirm successful cloning, an agarose gel representing the restriction digests for both the MP and NSP plasmids subsequent to transforming the cloned pCOLDI expression vector into the respective BL21 DE3 pLysS and BL21 DE3 *E.coli* host strains was performed. ***Figure* 4.6** indicates that cloning was successful. The marker used was 1 kb plus (Thermo-scientific) an additional low range marker (Thermo-scientific) was used due to the small fragment size estimated to be retrieved from the truncated MP restriction digest. The resulting gel shows visible differences in the un-digested and digested plasmids for both the MP and NSP. The truncated MP plasmids which were digested with *NdeI* and *BamHI*, from the successful *E.coli* strain transformants (BL21 DE3 pLysS) are represented on the gel by the lanes labelled 'Trunc MP cut'. The truncated NSP plasmids which were digested with *SacI* and *BamHI*, from the successful *E.coli* strain transformants (BL21 DE3) are represented

on the gel by the lanes labelled ‘Trunc NSP cut’. The un-digested plasmids for both the truncated MP and NSP are labelled by ‘Trunc MP uncut’ and ‘Trunc NSP uncut’ respectively. The resulting restriction digestions show successful transformation of the expression vector pCOLDI containing the two separate genes truncated MP and truncated NSP into the *E.coli* strains BL21 DE3 pLysS and BL21 DE3 respectively.

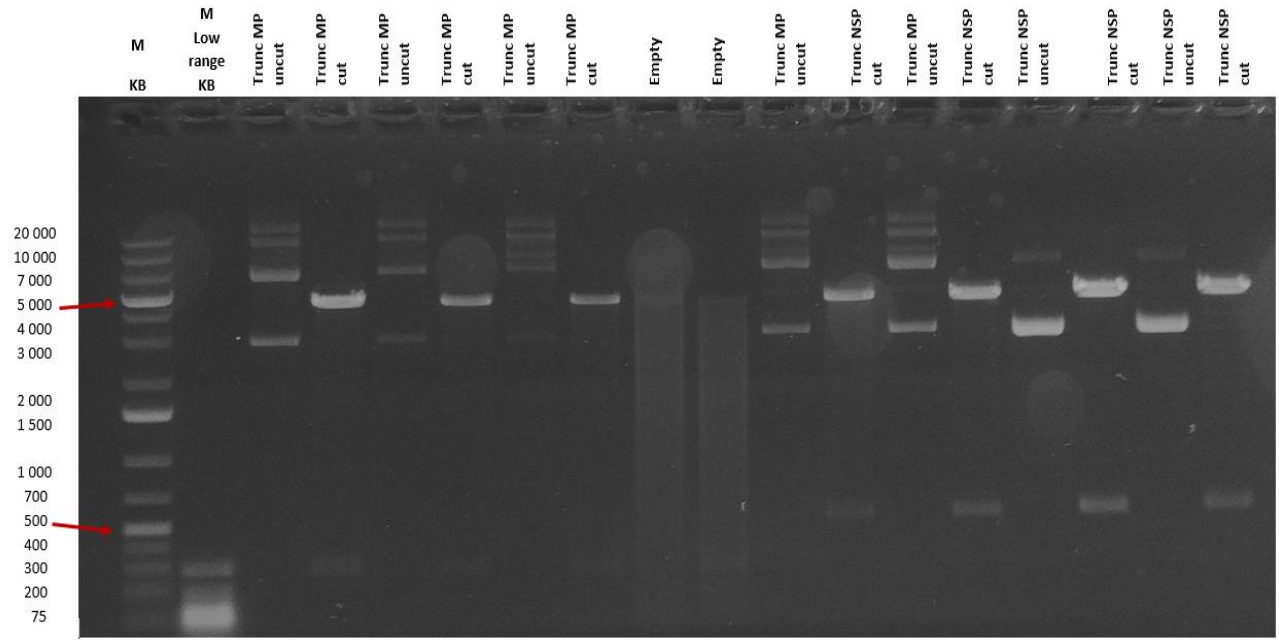


Figure 4.6: A 1% agarose gel showing the restriction digests, confirming the presence of the cloned MP and NSP genes in the pCOLDI expression vectors after transformation into their respective *E.coli* strains (BL21 DE3 pLysS and BL 21 DE3).

4.3.3. Protein overexpression

4.3.3.1. Truncated MP overexpression

***Figure* 4.7** below represents the expression trials of the truncated MP at 15 **°C** for 24 h using a pCOLDI expression vector. The supernatants and their pellets are shown with their corresponding IPTG inductions. The addition of IPTG occurs from 0.25 mM up to 1 mM. The increasing concentration of IPTG induction assists with the expression of the truncated MP, this can be seen by the thick bands visible at 0.50 mM and 1 mM IPTG induction. While successful expression is observed in the resulting gel, it is however in the insoluble fraction. A fair amount of expressed protein is visible in the insoluble fractions of 0.25 and 0.75 mM of IPTG while a

greater amount of expressed protein can be seen in the insoluble fractions attained at 0.50 and 1 mM of IPTG induction. While small fractions of soluble protein are visible at 0.50 and 0.75 mM IPTG inductions these fractions are insufficient for downstream applications.

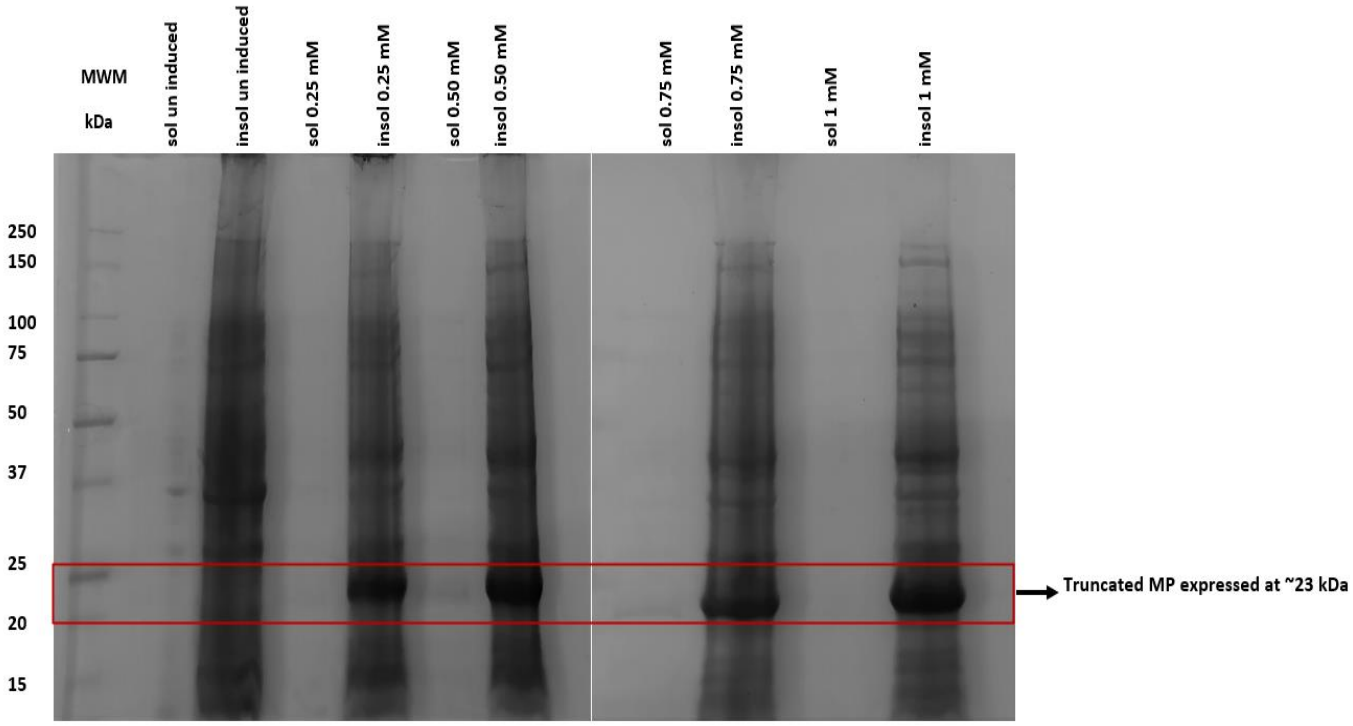


Figure 4.7: A 12% SDS-PAGE gel, showing the expression profile of the truncated MP from un induced till a 1 mM IPTG induction profile. MWM: Molecular weight marker, unstained (Bio-rad, USA); soluble fractions (sol); insoluble fractions (insol).

4.3.3.2. Truncated NSP expression

***Figure* 4.8** represents the expression trials of the truncated NSP at 15**°C** for 24 h using a pCOLDI expression vector. The supernatants and their pellets are shown with their corresponding IPTG inductions. The addition of IPTG occurs from 0.25 mM up to 1 mM. While increasing IPTG concentrations were introduced to determine the optimal IPTG concentration for induction of expression of the protein, it was found that the lower IPTG induction of 0.25 mM worked best for protein expression of NSP. While protein expression for the truncated NSP was successful, this was for insoluble protein fractions only. No soluble protein was found to be expressed.

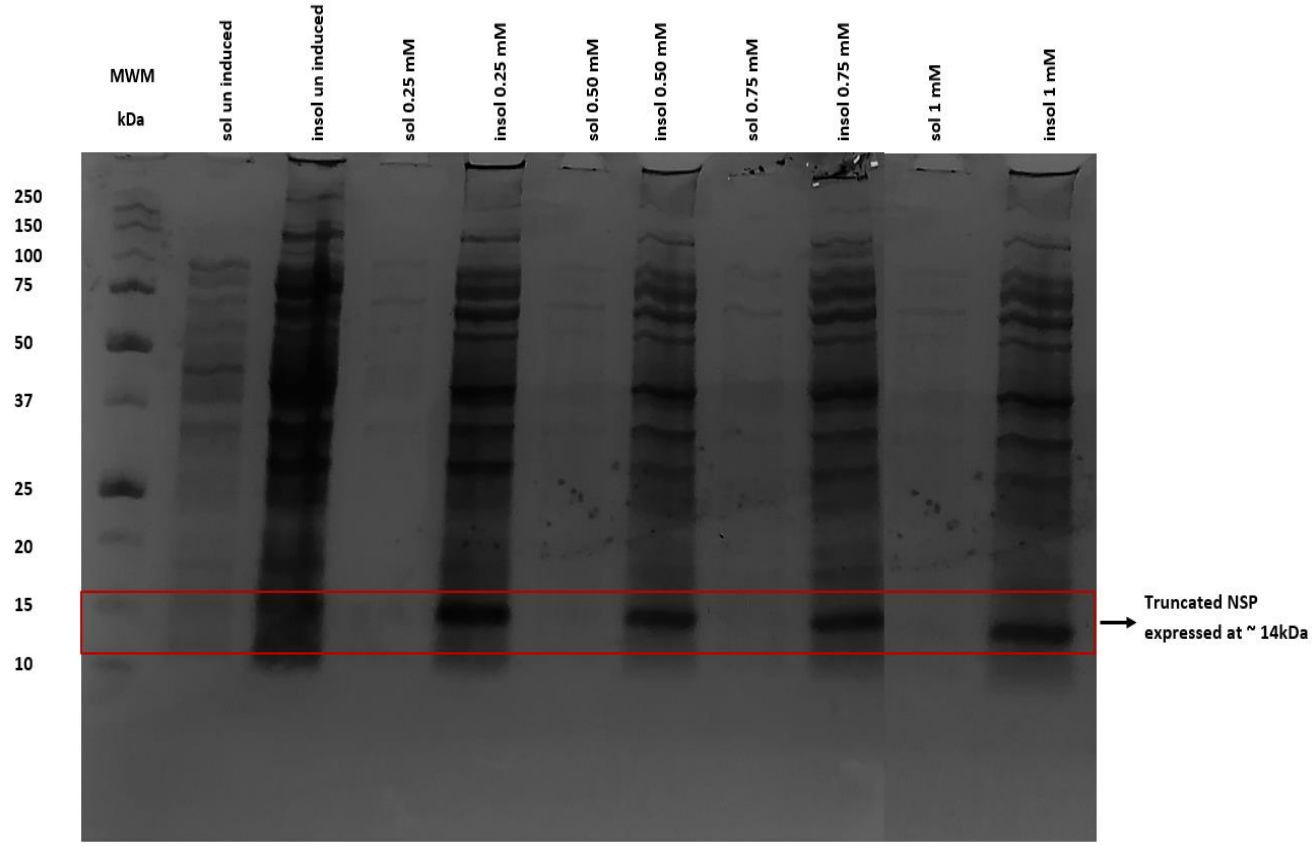


Figure 4.8: A 12% SDS-PAGE gel, showing the expression profile of the truncated NSP from un induced till a 1 mM IPTG induction profile. MWM: Molecular weight marker, unstained (Bio-rad, USA); soluble fractions (sol); insoluble fractions (insol).

4.3.4. Protein purification

4.3.4.1. Protein purification of the truncated MP

The purification profile of the SACMV truncated MP is viewed in ***Figure* 4.9** below. The purification was successfully carried out on a gravity flow-affinity chromatography column, employing a nickel charged resin. The solubilised protein represents the denatured form of the protein, subjected to overnight dialysis in 4 M urea solution. The flow through was collected subject to a 30 min incubation time allowing the protein to bind to the nickel resin. Column washes using the 4 M urea dialysate buffer with 25 mM imidazole indicated successful binding of the protein to the nickel resin as no significant protein can be visualised in the washes. Elution of the truncated MP with the addition of 250 mM imidazole to 4 M urea dialysate buffer successfully eluted the protein off the nickel resin. While elution fractions 1 till 3 show the truncated MP well

eluted with concentrated bands at ~ 23 kDa there are still impurities. However, the eluted fraction 4 while visibly less concentrated in its protein elution, contains no impurities and is suitable for downstream applications.

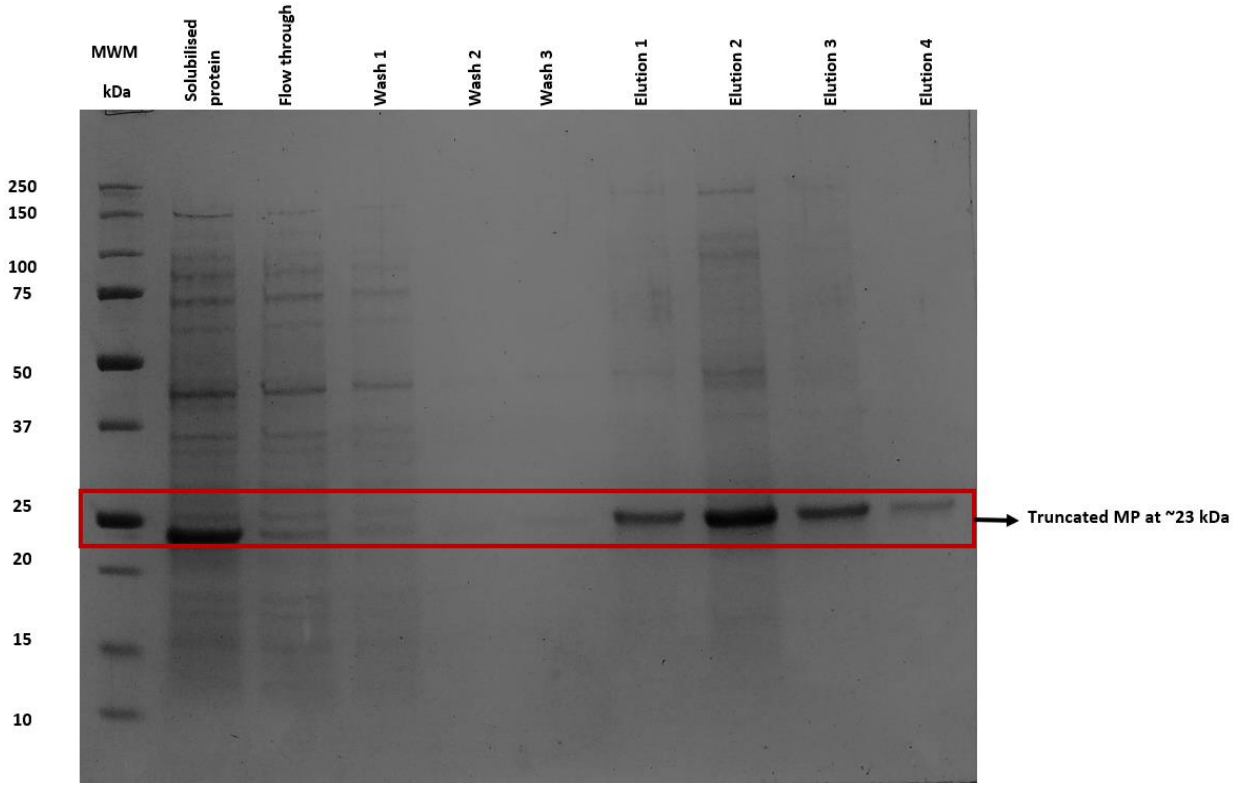


Figure 4.9: A 12% SDS-PAGE gel, showing the elution profile of the purification of the truncated MP of SACMV through a nickel charged resin using gravity flow-affinity chromatography.

4.3.4.2. Protein purification of the truncated NSP

The purification profile of the SACMV truncated NSP is seen in ***Figure* 4.10** below. The purification was carried out on a gravity flow-affinity chromatography column, employing a nickel charged resin. The solubilised protein represents the denatured form of the protein, subjected to overnight dialysis in 4 M urea solution. The flow through was collected subject to a 30 min incubation time allowing the protein to bind to the nickel resin. Column washes using the 4 M urea dialysate buffer with 25 mM imidazole indicated successful binding of the protein to the nickel resin as no significant protein can be visualised in the washes. Elution of the truncated MP with the addition of 250 mM imidazole to 4 M urea dialysate buffer successfully eluted the protein off the nickel resin. Despite numerous attempts the elution fractions 1 till 4

show the truncated NSP, eluted with concentrated bands at ~ 13 kDa, however, there are still impurities. The marginal impurities were less visible in elution fractions 3 and 4, which were used for downstream applications.

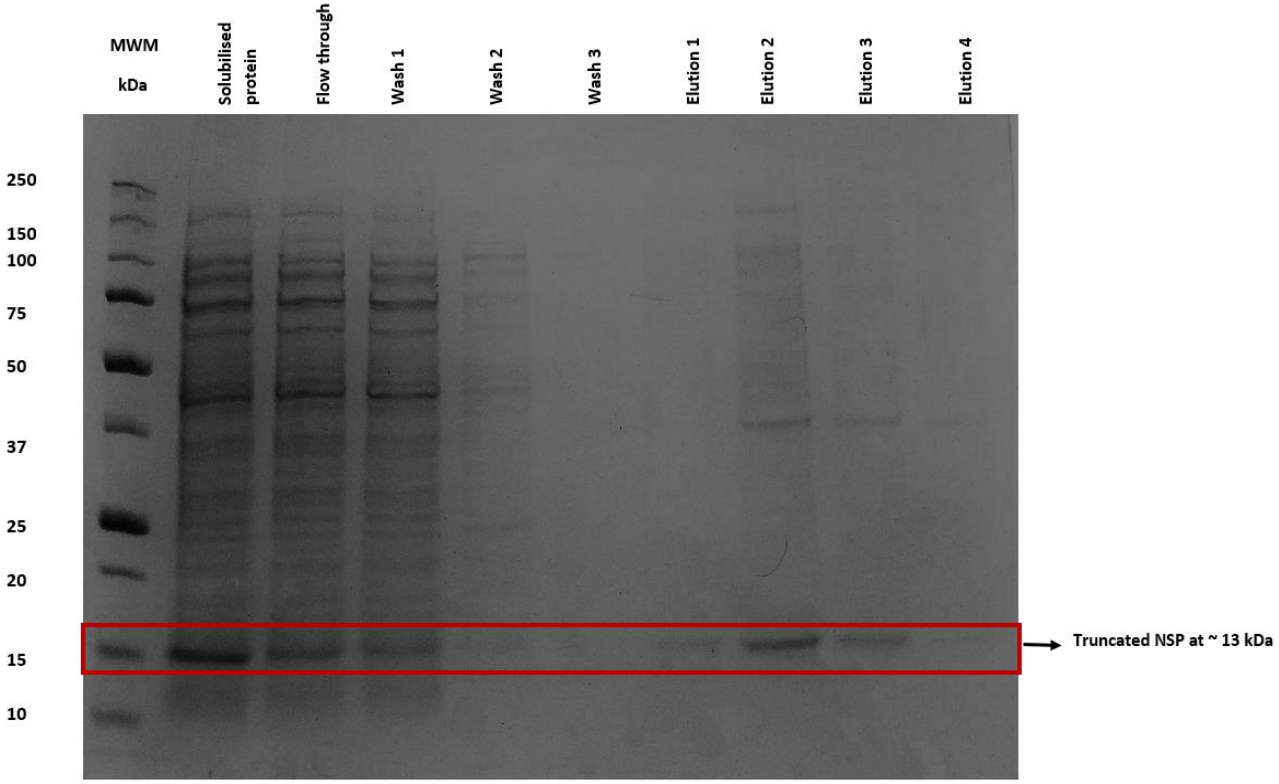


Figure 4.10: A 12% SDS-PAGE gel, showing the elution profile of the purification of the truncated NSP of SACMV through a nickel charged resin using gravity flow-affinity chromatography.

4.3.5. Protein refolding

The resulting gel **(*figure* 4.11)** represents the refolded truncated MP of SACMV. Lanes 1,2 and 3 all represent a sample triplicate. The refolded protein is clearly visible at ~ 23 kDa. Smears are visible in the sample lanes and this could be due to a refolded protein being run on a SDS gel.



Figure 4.11: A 12% SDS-PAGE gel, showing the refolded truncated MP of SACMV.

The resulting gel **(figure 4.12)** represents the refolded truncated NSP of SACMV. Lanes 1,2 and 3 all represent a sample triplicate. The refolded protein is present at ~ 13 kDa. Smears are visible in the sample lanes and this could be due to a refolded protein being run on a SDS gel.

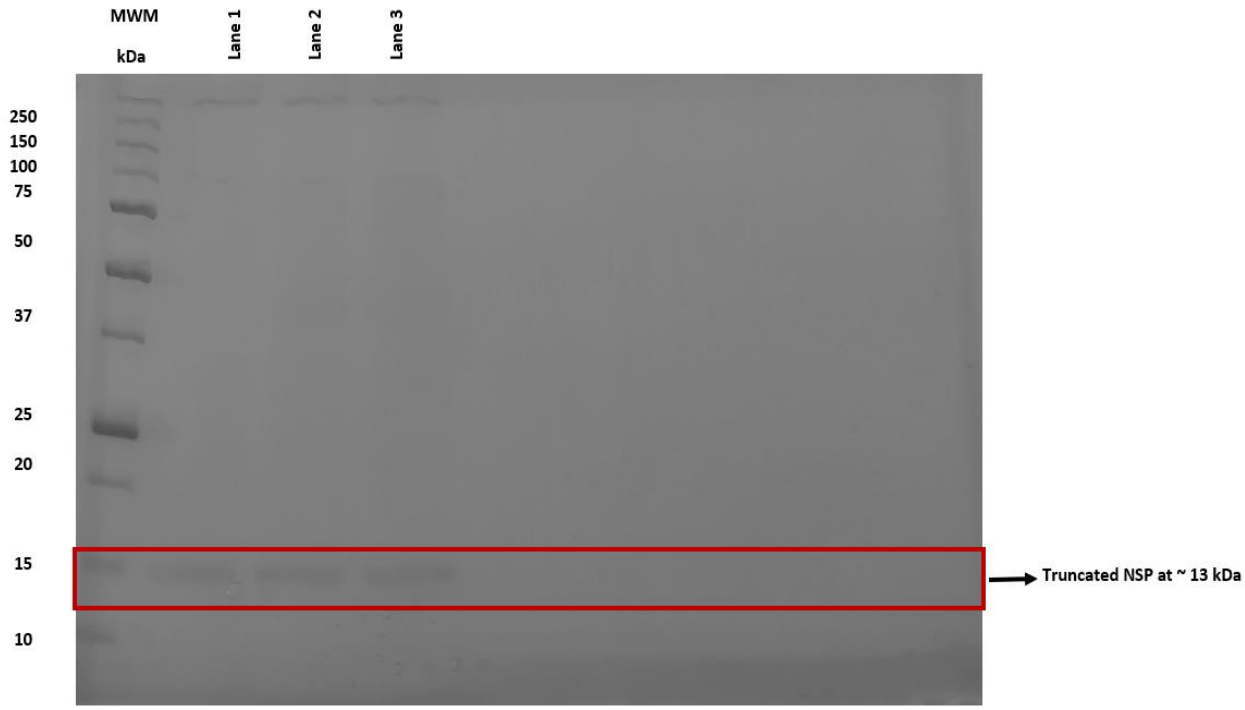


Figure 4.12: A 12% SDS-PAGE gel, showing the refolded truncated NSP of SACMV.

4.4. Discussion

In an effort to minimize possible insolubility of the proteins, (MP and NSP) truncations of these proteins were used for this study. The likelihood of insolubility of these full-length proteins were demonstrated by previous studies in our laboratory. Previous studies by Nankoo, 2018, indicated that the full-length MP of SACMV was insoluble in expression and the protein contained many intrinsically disordered regions contributing to the insolubility of the protein. The regions of truncation for both the proteins were selected based on the least intrinsically disordered domains as this was expected to give better solubility in terms of expression. Computational studies (further discussed in chapter 3 of this study) further support and highlight the motivation of using truncations of these proteins for this study.

The 528 bp region of the MP located predominantly on the N terminus was selected for cloning into the expression vector of choice (pCOLDI). While there is no current published model for the SACMV MP, Fondong, 2013 proposed a general model of geminivirus MPs functional domains. The model illustrates that geminivirus movement proteins contain a pilot domain on the N terminus which is essential for localization to the cell periphery and plays a critical role in the binding of DNA to the NSP-MP complex and its subsequent export from the nucleus, supporting the 'couple-skating' transport model (Fondong, 2013; Frischmuth, et al., 2007; Hehnle et al., 2004; Zhang.S.C et al., 2001). While the pilot domain of the MP was of most interest due to its potential ability to bind and export DNA out of the nucleus, computational data shows that the pilot domain of the SACMV MP is less likely to have disordered regions in its secondary structure composition and therefore better solubility was to be expected. In proteins where insolubility is a possibility, especially viral proteins, it is favourable to express each functional domain individually (Tokmakov et al., 2015).

In contrast the 267 bp region of the NSP predominantly located on the C terminus was selected for cloning into the expression vector of choice. Similarly, to the MP of SACMV there is no published model for the NSP of SACMV, however it is predicted that these proteins will be highly conserved due to functional similarities. The general geminivirus model of the NSP (Fondong, 2013) illustrates that the C terminus contains the nuclear export signal and MP interacting domain, which is critical in binding to the MP and forming the NSP-MP complex, further supporting the 'couple-skating'

transport model (Fondong, 2013; Frischmuth, et al., 2007; Hehnle et al., 2004; Zhang.S.C et al., 2001).

Both the truncated MP and NSP genes were codon optimized upon cloning into the chosen expression vector. Codon optimization is a popular cloning method as it involves synonymous codon changes which leads to enhanced protein expression in a heterologous system and as a result improves the overall expression and translation efficiency of the gene of interest. Studies support that there is increased protein expression and production especially when codon optimized genes are expressed in *E.coli* (Haridhasapavalan et al., 2020).

The truncated MP and NSP genes were respectively transformed into the BL21 DE3 pLysS and BL21 DE3 strains of *E.coli*. While both the strains of *E.coli* contain the DE lysogen which expresses the T7 polymerase upon IPTG induction, the pLysS strain of *E.coli* produces a T7 lysozyme, which reduces basal level expression of the gene of interest. This makes the pLysS strain of *E.coli* more suitable for expression of more toxic genes. While both proteins were initially transformed into a number of *E.coli* strains successful transformation was only found in the aforementioned strains. For the expression of recombinant proteins, a host system must be selected which is suited for the type of protein which is being expressed and allows for any post translational modification found in eukaryotes. *E.coli* continues to be one of the preferred bacterial organisms, for the expression of proteins, due to its genome which has been substantially studied and its largely understood mechanism regarding transcription, translation and protein refolding (Baneyx, 1999; Chen, 2012). While yeast is another well-studied eukaryotic system used for the expression of proteins (Wilson and Walker, 2012) it is more difficult to work with due to cell growth being slower and owing to the levels of expression not always being optimal, and post-translational modifications may also differ *in vitro* and in vivo. However, working with yeast is advantageous as the system offers a high level of intracellular expression and secretes few of its own proteins making expression products easier to purify for downstream work (Chen, 2012; Nasser et al., 2003).

Both expressed proteins (truncated MP and NSP) were found to be insoluble in their expression. While the truncated MP was insoluble, at 0.50 and 0.75 mM of IPTG induction of a small amount of soluble protein was observed. This protein however

was insufficient to continue with downstream applications as the yield was too low. In contrast, no soluble protein was detected in the expression of the truncated NSP. Expression trials usually entail optimization of incubation time, temperature and the concentration of ITPG. However due to the expression vector utilized for the study only the manipulation of the IPTG concentration was required (Clark, 2001; Papaneophytou and Kontopidis, 2014; Wilson and Walker,2012). The expression vector pCOLDI used in this study requires a cold shock expression system, the optimal expression for this expression vector especially when expressing proteins less likely to be soluble is 15 °C . The optimal incubation time is also a period of 24 h (Takara Bio Inc, 2018). However despite optimum conditions used to express both the truncated MP and NSP soluble protein, sufficient soluble protein was not expressed. For this reason proceeding studies continued with the insoluble fractions retrieved for MP and NSP. Since 1 mM IPTG induction gave the greatest yield of insoluble protein for the truncated MP, this was the expression parameter chosen for further expression of the truncated MP for downstream applications. An IPTG induction of 0.25 mM gave a greater yield insoluble protein for the truncated NSP, and this was the expression parameter chosen for further expression of the truncated NSP for downstream applications.

Immobilized metal ion affinity chromatography (IMAC) is a commonly used protein purification method. The purification method is based on histidine residues and charged transition metal ions which are immobilized on a matrix (in this study Ni^{2+} was employed) (Haridhasapavalan et al., 2020). The truncated MP produced a pure protein in elution 4 after the elution buffer containing 250 mM imidazole was passed through the nickel charged resin. This fraction was pooled together from multiple truncated MP purifications to proceed with downstream applications. The truncated NSP produced pure protein between fractions 1 till 4, with the elution buffer containing 250 mM imidazole is passed through the column. While all four elutions were not as pure as desired (some contaminating bands were observed), elutions 3 and 4 contained less contamination than elutions 1 and 2. An imidazole concentration of between 150-200 mM was also used for this elution, however the protein did not elute off the column. A concentration of 300 mM of imidazole in the elution buffer eluted all the protein out in the first elution with significant contamination and no protein eluted in proceeding fractions. Therefore, due to minimal contamination in elutions 3 and 4, these fractions

were chosen to be pooled from multiple truncated NSP purifications for downstream applications. It has been reported in some studies that it is difficult to attain highly pure recombinant protein when expressing in *E.coli*, such as in the purification of the recombinant human ETS protein (Haridhasapavalan et al., 2020).In particular, viral proteins are notoriously difficult (Linding et al., 2003), which may explain why purified geminiviral proteins have not been previously studied.

Following protein purification, both truncated proteins were refolded using stepwise dialysis. Stepwise dialysis brings the denatured protein to equilibrium with the aid of a high concentration of denaturant, this is followed by the concentration of the denaturant being decreased gradually. The addition of the reducing agent DTT, assists in promoting the formation of disulphide formation, which plays a critical role in the correct refolding of proteins eliminating protein aggregation (Clark, 1998; Yamaguchi and Miyazaki, 2014). While refolded proteins are considered native and ideally should be run on a native page gel, both these truncated proteins did not successfully run on the native page gels as the proteins did not migrate out of the wells. Therefore, both proteins were run on 12% SDS-PAGE gels (samples were not exposed to any β-mercaptoethanol or heating prior to being run on the gel), and truncated NSP and MP proteinswere detected despite some smearing, due to the native proteins being run on SDS-PAGE gels.

4.5. Conclusions

The successful expression and purification of two truncated proteins, NSP and MP, encoded by SACMV's DNA B is, to our knowledge, the first report. Expression *in vitro* and purification of the MP N terminus and the NSP C terminus have not been previously researched. It has now been established that both these proteins can only be expressed in the insoluble form. The truncated MP and NSP both bind successfully to an immobilized metal ion affinity chromatography column coated with a nickel resin, allowing pure protein to be easily attained. Due to the insolubility of both proteins, refolding of both the truncated MP and NSP was required and was completed with success. Further quantification and structural characterization of both these truncated proteins are discussed in Chapter 5.

Chapter 5

Secondary Structural characterization of truncated Movement Protein and Nuclear Shuttle Protein

5.1. Introduction

As previously highlighted in earlier chapters, very little is known about the structure of the nuclear shuttle (NSP) and cell-to-cell movement (MP) proteins of geminiviruses, however, some studies have been done on their functions (Sudarshana et al., 1998). While some research has been performed on the full-length MP of SACMV (Nankoo, 2018), no research focussed on a specific domain has been studied thus far.

In order to fully understand the functions of the SACMV MP pilot domain and the putative MP binding/interacting domain of SACMV MP, the protein structures for these domains were determined. The deduced amino acid sequence of SACMV DNA B and functional studies have informed putative models for both the MP and NSP structural domains, however, little is known about the secondary and tertiary structure particularly of the domains of interest (Fondong, 2013).

The primary structure of a protein comprises the sequence of amino acids which are linked by peptide bonds. The secondary structure of a protein consists of localised folding of a polypeptide chain due to hydrogen bonding. The secondary structure typically includes structures such as α-helices, β-strands and disordered regions. The secondary structure of a protein can be determined using a technique such as circular dichroism (Lesk, 2010; Wilson and Walker, 2012). The tertiary structure of a protein relates to the overall folding of a polypeptide chain. The tertiary structure of a protein can be assessed using spectrophotometric and fluorescence methods such as extrinsic fluorescence with the aid of a fluorophore (Wilson and Walker, 2012).

A spectrophotometric method which permits for the elucidation of secondary of proteins is circular dichroism. Due to the chirality of the proteins peptide backbone, which when exposed to plane polarised light, the left and right handed polarised components are absorbed to different extents (Kelly et al., 2005).

The resulting radiation shifts from the original plane, and this radiation results from the chiral molecule displaying “elliptical radiation” (Kelly *et al*., 2005). Due to prominent existing algorithms, predominant secondary structures, such as α-helices, β-sheets

and random coil structures are associated with characteristic CD spectra in the far UV range (below 250 nm)(Kelly *et al*., 2005).

Extrinsic chromophores which are non-conjugated can be used in order to display a proteins' tertiary structure. These chromophores usually emit a weak fluorescence in polar environments but emit strong fluorescence in non-polar environments, and the fluorescence emission changes with the λmax or blue shift changing with great intensity (Lesk, 2010; Wilson and Walker, 2012).

Proteins possess three intrinsic fluorophores: tyrosine, phenylalanine and tryptophan. Fluorescence emitted from tryptophan is sensitive to the polarity of its local environment and this fluorescence is usually employed for the purpose of conformational monitoring (Vivian and Callis, 2001; Wilson and Walker, 2012).The absorbance of light by a chromophore causes its electrons to be raised to a higher energy state, however, upon returning to a lower energy state, the energy that is lost by the electrons at a longer wavelength will be in a light emitting manner to induce the phenomenon of fluorescence (Sheehan, 2009). Proteins possessing the fluorescence properties of tyrosine and tryptophan residues, owing to their aromaticity, aid as suitable intrinsic probes to verify the tertiary integrity of proteins (Lakowicz, 2006).

This chapter focusses on unravelling the structure of the truncated MP and NSP of SACMV. The elucidation of the secondary structure of these truncated proteins were assessed by the following techniques: Far UV circular dichroism, tryptophan fluorescence, ANS fluorescence, SE-HPLC, Mant-ATP fluorescence and a phosphate functional activity assay.

5.2. Materials and Methods

5.2.1. Protein quantification

Peptide bonds, aromatic amino acids, and to a small extent disulfide bonds, account for the ability for proteins to absorb UV light. This allows for the concentration of proteins to be determined in solution (Wilson and Walker, 2012). Following stepwise dialysis (Chapter 4) each of the truncated proteins were transferred to separate spin filter columns (Vivaspin 20, Sartorius, UK), and centrifuged at 8 000 xg for 35 min where the protein was concentrated down to a final volume of four ml.

Each of the proteins was then prepared in serial dilutions for quantification. A ten times serial dilution was prepared using 100 µl of protein, diluted into 900 µl of refolding buffer, and thereafter serial dilutions were prepared as follows: 20 X, 40 X, 80 X, 160 X and 320 X (using the refolding buffer to dilute the protein). The serial dilutions were analysed at absorbance wavelengths between 340- 280 nm using the Spectra manager programme on the UV/Vis Spectrophotometer Jasco V-630. The Beer-Lambert law was applied as follows:

$$A = \varepsilon cl$$

$A = \varepsilon cl$ where A is the absorbance, ε is the molar extinction coefficient ($M^{-1} cm^{-1}$) at a given wavelength (280 nm for protein detection), c is the molar concentration and l is the pathlength (cm).

$$\text{Conc (mg/ml)} = \frac{(\text{Slope} \times Mr)}{\epsilon}$$

The slope is obtained from the equation of the line from the graph of absorbance vs wavelength. The molar extinction coefficient was derived theoretically from ExPASy ProtParam (Gasteiger et al., 2005) (MP- 21 555 $M^{-1} cm^{-1}$), (NSP- 21 555 $M^{-1} cm^{-1}$).

5.2.2. Far-UV Circular Dichroism

CD is described as the phenomenon when circularly polarised light at a specific wavelength passes through a chiral chromophore such as an amino acid, the left and right components of light are differentially absorbed. In particular, the CD spectra of a protein in the far-UV region is dominated by the $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions of peptide bonds, which are centred around 222 nm and 190 nm, respectively. These spectra are influenced by the geometries of the polypeptide backbone and thereby reflect the various phi (Φ) and psi (Ψ) angles of the peptide bond which subsequently indicate

the secondary structural elements present within a protein such as α-helices, β-sheets and random coils (Kelly et al., 2005; Whitmore et al., 2007). In contrast to β-sheet structural motifs which display a minimum and maximum ellipticity at 218 nm and 195 nm, respectively, proteins which are predominately α-helical, display a trough at 208 nm and 222 nm and a peak at 190 nm (Kelly et al., 2005).

A stock solution of each of the proteins were dialysed 1000 X in a 50 mM Tris-HCl solution (pH 7.4) at 4°C. The dialysed proteins were filtered using a 0.2 µm syringe filter (Sartorius, UK) in order to remove particles which may affect polarisation of light. All measurements were analysed using the Jasco J1500 spectrometer (Jasco, Tokyo, Japan) at 20°C using a wavelength range of 190–260 nm. A 2 mm pathlength quartz cuvette was used. The data pitch was set to 0.5 nm with a scanning speed of 100 $nm.min^{-1}$ response time. All spectra were corrected for background by subtracting the signal of the blank solution from the protein spectra signal. The raw data was converted to mean residue ellipticity [θ] ($mdeg.cm^2$ $dmol^1$) using the following formula

$$[\Theta]_MRE = 100\theta/cnl$$

Ellipticity is the CD milli degree theta, c is the protein concentration, n is the number of residues and l is the pathlength in cm. The quantity of secondary structure of both proteins were estimated the using the CONTIN algorithm on the Dichroweb server (Whitmore and Wallace, 2004).

5.2.3. Intrinsic tryptophan fluorescence

Intrinsic tryptophan fluorescence was measured herein to determine conformational changes of the proteins. The method also confirms if the proteins exists in their native state and are indeed refolded of the truncated MP and NSP which coincidentally contain 2 tryptophan residues each. For each of the proteins a final concentration of 10 µM of native protein was analysed, the denatured protein up to a final concentration of 10 µM (exposed to 8 M urea) was also analysed for both proteins. Both, the truncated MP and NSP proteins in their native and denatured forms were both also analysed in the presence of 5 µM ATP. All spectra were corrected for background interference by the buffer blank (refolding buffer- 0.5 M Tris-HCl, pH 7.5, 0.5 M L-arginine). All samples were analysed using the Jasco FP-6300 spectrofluorimeter with a 1 cm pathlength cuvette. The tryptophan residues were excited at 295 nm and the fluorescence spectra were selectively recorded between a range of 300-600 nm using

an excitation and emission bandwidth of 5 nm. The spectra were recorded at 20°C and at a scan speed of 200 $nm.min^{-1}$. Each spectrum is the average of three accumulations.

5.2.4. Extrinsic ANS fluorescence binding assay

Fluorescence spectroscopy depends upon the intrinsic fluorescence properties of proteins which aid in their structural characterisation, the use of extrinsic fluorescent dyes (fluorophores) may offer additional insights into the structural characterisation of proteins. Similar to tryptophan, a valuable property of many fluorophores is their sensitivity to the surrounding environment. Thereby, they may be used as probes for the determination of their location on a protein. Therefore, the use of extrinsic dyes has become increasingly popular for the use of detection of molten globule intermediates during unfolding events, assess surface hydrophobicitiy and probe the active sites of enzymes (Hawe et al., 2008). One of the most commonly used extrinsic fluorophores used to probe hydrophobic surfaces in proteins is the compound 8-anilinonaphthalene-1-sulfonic acid (ANS). In an aqueous environment ANS has a low quantum yield, however, when it binds non-covalently within a non-polar environment of a protein its fluorescence (Hawe et al., 2008).

A 20 mM stock of ANS (Sigma Aldrich, USA) was made up in a buffer consisting of 100 mM sodium phosphate, pH 6.5, 1 mM EDTA, 0.02 % NaN_2 and stored in the dark. The buffer was allowed to shake for one hour on a shaking platform at approximately 180 rpm. The concentration of ANS was determined using UV/Vis absorbance at A350. Once the concentration of ANS was determined the ANS was added in excess to both proteins in their native and denatured forms. All four samples were analysed using a Jasco FP-6300 spectrofluorimeter. A one cm pathlength cuvette was used. Excitation was set at a wavelength of 380 nm at a slit width of 5 nm at an average of 5 scans. Background (buffer only) was corrected for, and all recordings were recorded between 400 to 600 nm at the scan speed of 200 $nm.min^{-1}$. Fluorescence spectra (recorded at 20°C) were generated from an average of three readings.

5.2.5. Intrinsic Fluorescence MANT-ATP binding Assay

The binding of methylanthraniloyl-ATP (mant-ATP) to a protein may reveal information regarding possible binding sites. Mant-ATP contains a fluorescent probe which emits

fluorescence at ~440 nm when excited at ~355 nm. Therefore, any changes in the tertiary structure alter the fluorescence intensity of mant-ATP (Ni et al., 2000). The fluorescent probe is very sensitive to environmental conditions such as polarity. In non-polar environments, the fluorescent quantum yield is significantly higher than in polar environments. Changes in mant-ATP fluorescence may provide insight into protein tertiary structure, polarity and substrate binding (Ni et al., 2000).

A 10 µM stock of an original 5 mM stock of MANT-ATP (Sigma Aldrich, USA) was made up in refolding buffer (50 mM Tris-HCl pH 7.5, 0.5 M L-Arginine). Per 10 µM of MANT-ATP, 10 µM of protein for native and denatured protein analysis was used (for truncated MP and truncated NSP). Following the addition of MANT-ATP to protein, the samples (protein containing MANT-ATP in the native and denatured form) were briefly placed on a vortex for thirty seconds.
This was followed by incubation on ice and vigorous shaking on a shaking platform for 30 minutes. All samples were analysed on the Jasco FP-6300 spectrofluorimeter. Excitation was set at 360 nm at a slit width of 5 nm at an average of 5 scans. The cuvette pathlength was 1 cm. The average of three accumulations were used to give an average spectrum and all recordings were recorded at wavelengths between 400 and 600 nm. All spectra were recorded at 20°C, background (buffer) corrected for, and take into consideration an average of all three accumulations at the scan speed of 200 $nm.min^{-1}$.

5.2.6. Phosphate assay for determination of functionality

Phosphate assays were employed to measure if the truncated MP and truncated NSP had phosphate activity. To determine this, a phosphate assay kit (Sigma Aldrich, USA(MAK308)), to determine the amount of free phosphate was utilised. Experiments were prepared and measured as per the manufacturer's instructions **(Appendix A)**. All measurements were recorded at 620 nm on a Multiskan spectrophotometric plate reader. All background absorbance was corrected for by subtracting background values from the standard phosphate spectra.

5.2.7. Size exclusion high performance liquid chromatography (SE-HPLC)

Size exclusion chromatography (SEC) is a separation technique which is dependent on the relative size of a macromolecule. A SEC column comprises of a gel of beads with pores which have a defined and a narrow size range. Molecules that fit within the fraction range of the gel will be retained in a manner which is inversely proportional to their hydrodynamic volume, molecules with the smallest size will be retained for longer, while molecules with a size which exceeds the exclusion limit of the gel will pass freely through the column and elute first. In a SE-HPLC system the usage of high-pressure pumps assist the mobile phase of the column to move at a high flow-rate; ensuring rapid separation and optimal resolution (Wilson and Walker, 2010).

Analysis of the molecular weight and consequently the oligomeric state and purity of each target protein by means of SE-HPLC was carried out on a TSKgel SuperSW2000 SEC column that had a resolution range of 5–150 kDa (Tosoh Corporation©, Tokyo, Japan). The SEC column was additionally conjugated with a TSKgel SWXL guard column (Tosoh Corporation©, Tokyo, Japan), this was connected to a Shimadzu ultrafiltration liquid chromatography (UFLC) SPD-20A SE-HPLC system. The SEC column was equilibrated at 20°C with the SE-HPLC buffer (50 mM Tris-HCl, 0.5 M L-Arginine, 400 mM NaCl pH 7.2) at a flow rate of 0.2 $ml.min^{-1}$ and isobaric pressure of ~900 psi. Thereafter, in triplicate, 20 µl of each of the target proteins (in the presence and absence of ATP) (100 µM) were injected onto the column and allowed to pass through the column with a separation period of 30 min. A set of protein standards with a known molecular weight [thyroglobulin (670 kDa), γ-globulin (158 kDa), ovalbumin (77 kDa), myoglobin (17 kDa) and vitamin B12 (1.35 kDa)] were used to construct a standard curve by plotting the logarithm of the known molecular weight for the protein standards as a function of their retention time so that the molecular weight for the target proteins could be determined based on their retention time.

5.3. Results

5.3.1. Protein quantification

The below linear regression **(*figure* 5.1)** indicates the line of best fit for the scatter plot of corrected absorbance against the dilution factor. The trendline equation displays a gradient of 0.2643 and a R^2 value of 0.99. Using the Beer-Lambert law and the trendline gradient, a concentration of 0.25 mg/ml of truncated MP in refolding buffer (50 mM Tris-HCl pH 7.5, 0.5 M L-arginine) was calculated. The calculation associated with this figure can be found in **Appendix B**.

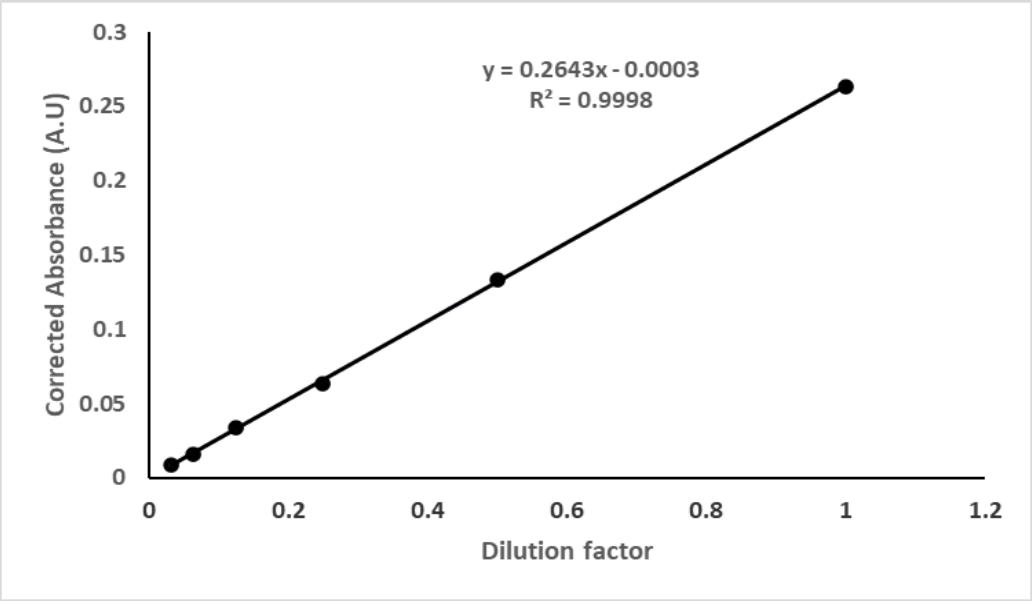


Figure 5.1: Quantification of the truncated MP.

The linear regression below **(*figure* 5.2)** indicates the line of best fit for the scatter plot of corrected absorbance against the dilution factor. The trendline equation displays a gradient of 0.1273 and a R^2 value of 0.99. Using the Beer-Lambert law and the trendline gradient a concentration of 0.07 mg/ml of truncated MP in refolding buffer (50 mM Tris-HCl pH 7.5, 0.5 M L-arginine) was calculated. The calculation associated with this figure can be found in **Appendix B**.

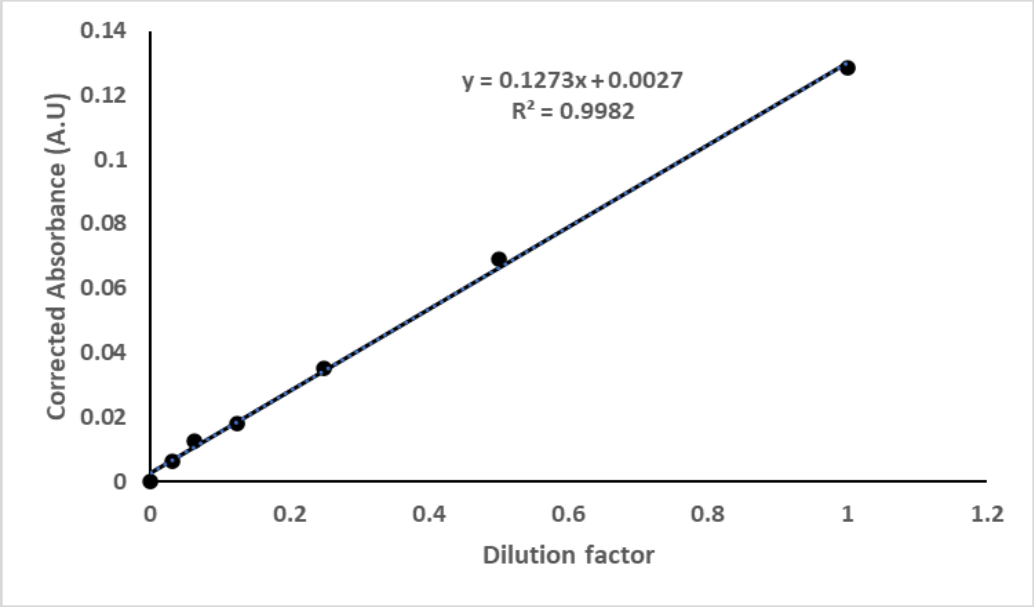


Figure 5.2: Quantification of the truncated NSP.

5.3.2. Far-UV Circular Dichroism

The spectrum below **(*figure* 5.3)** shows the native truncated MP, 5 mM in a 50 mM Tris-HCl solution (pH 7.4). The spectrum shows a peak at ~ 199 nm which is an indicative feature of a protein exhibiting properties of a β-sheet. The spectrum also shows a protruded peak between 210 to 225 nm, indicative of a protein exhibiting random coil/loop properties. This indicates that the truncated MP could possess a strong secondary structure containing a mixture of a β-sheets and random coils. The below table **(*table* 5.1)** shows the results from the Dichroweb server, which confirm the results from the spectra using the CONTIN analysis programme.

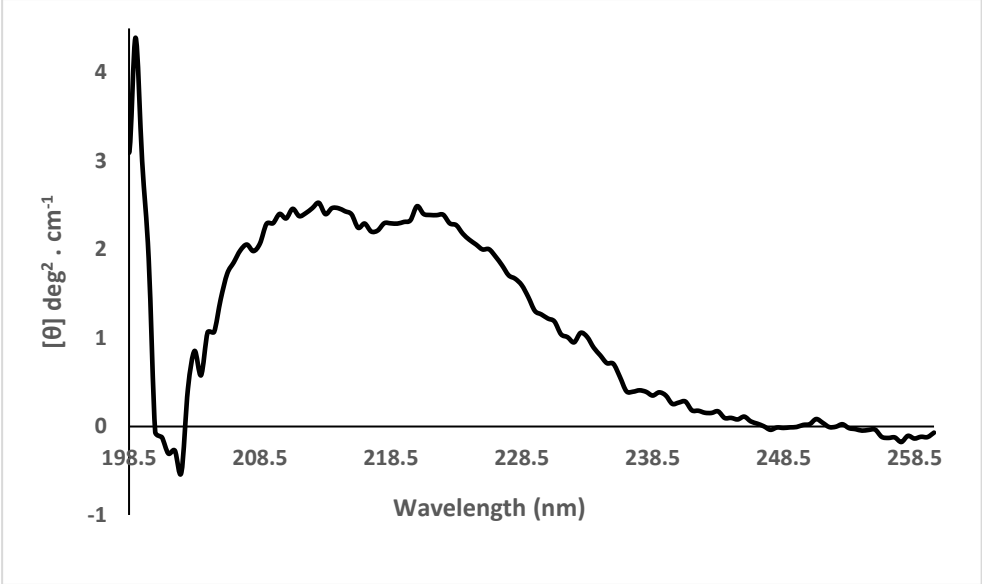


Figure 5.3: Far-UV circular dichroism spectrum for the truncated MP.

Table 5.1: Secondary structure content of the truncated MP as calculated using the CONTIN analysis on the Dichroweb server.

Secondary structure content (%)				
α- helical	**β- sheet**	**β- turn**	**random coil**	**Total**
2	41	18	39	100

The spectra below **(*figure* 5.4)** show the native truncated NSP in [5 mM in a 50 mM Tris-HCl solution (pH 7.4)] and the denatured NSP, [5 mM in a 50 mM Tris-HCl solution (pH 7.4)]. The spectra shows a peak for both the native and denatured NSP between 196 and 200 nm, a slight trough can be seen for the native NSP around 206 nm, this trough is more distinct for the denatured form of the protein, which is an indicative feature of proteins exhibiting properties of a β-sheet. The below table **(*table* 5.2)** shows the results from the Dichroweb server which confirm the results from the spectra using the CONTIN analysis programme.

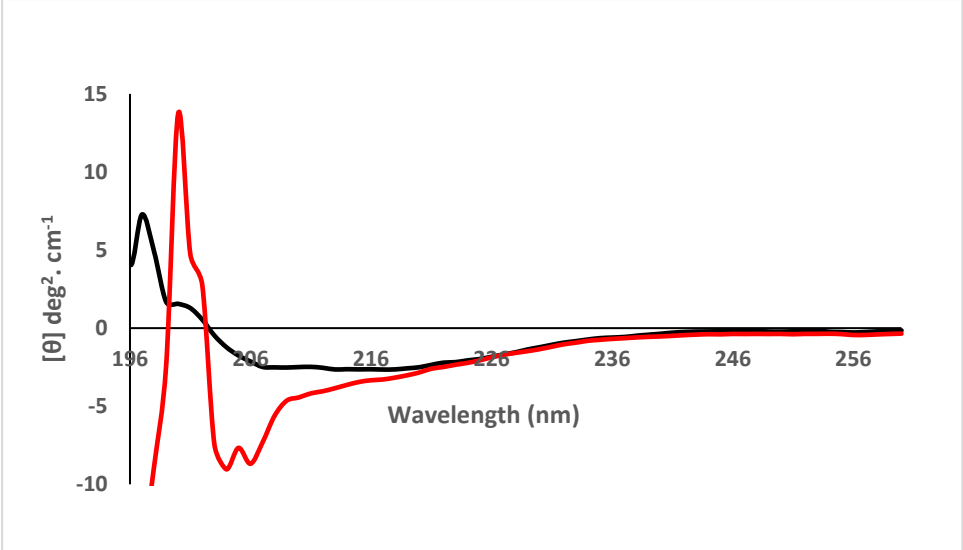


Figure 5.4: Far-UV circular dichroism spectra for the native (black line) and denatured (red line) truncated NSP.

Table 5.2: Secondary structure content as calculated using the CONTIN analysis on the Dichroweb server.

	Secondary structure content (%)				
Protein	**α- helical**	**β- sheet**	**β- turn**	**random coil**	**Total**
Native truncated NSP	5	41	17	37	100
Denatured truncated NSP	6	43	18	33	100

5.3.3. Intrinsic tryptophan fluorescence

The fluorescence spectra below **(*figure* 5.5)** shows native truncated MP (black line) in refolding buffer (50 mM Tris-HCl, pH 7.5, 0.5 M L-arginine) exhibiting high tryptophan fluorescent intensity upon emission at 350 nm. In the presence of ATP in refolding buffer, the native truncated MP (red line), exhibits the same tryptophan fluorescent intensity as without ATP. The denatured truncated MP (blue line) in refolding buffer

(50 mM Tris-HCl, pH 7.5, 0.5 M L-arginine) with 8 M urea exhibited a very low tryptophan fluorescent intensity upon emission at 350 nm. In the presence of ATP in the refolding buffer (green), denatured truncated MP exhibited a similarly low tryptophan fluorescence intensity pattern upon emission at 350 nm. The unfolding of the truncated MP influences the polarity of the environment, and the unfolding of the protein also confirms that the protein is indeed refolded.

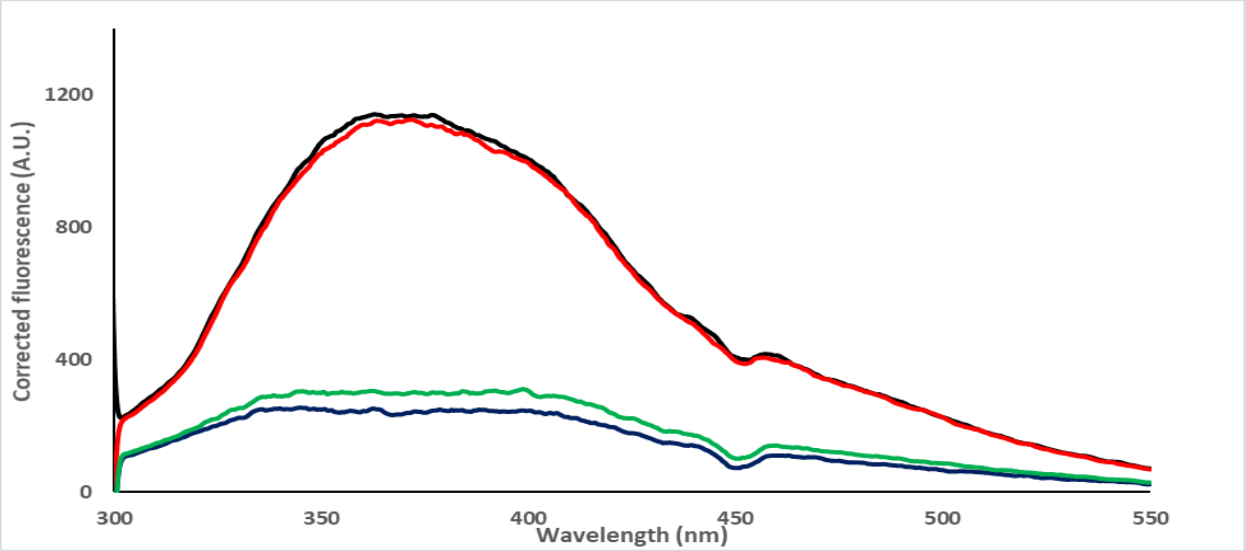


Figure 5.5: Intrinsic tryptophan fluorescence spectra of truncated MP.

The fluorescence spectrum below **(*figure* 5.6)** shows native truncated NSP (black) in refolding buffer (50 mM Tris-HCl, pH 7.5, 0.5 M L-arginine) exhibiting high tryptophan fluorescent intensity upon emission at 350 nm. In the presence of ATP in refolding buffer, the native truncated NSP (green) exhibited a significantly lower tryptophan fluorescent intensity upon emission at 350 nm. The denatured truncated NSP (blue) in refolding buffer with added 8 M urea, exhibited low tryptophan fluorescent intensity upon emission at 350 nm. The denatured truncated NSP (red line) exhibited a similarly low tryptophan fluorescence intensity pattern upon emission at 350 nm in the refolding buffer with added ATP. The unfolding of the truncated NSP influences the polarity of the environment, the unfolding of the protein also confirms that the protein is indeed refolded. The presence of ATP in the native form of the truncated NSP causes possible conformational changes.

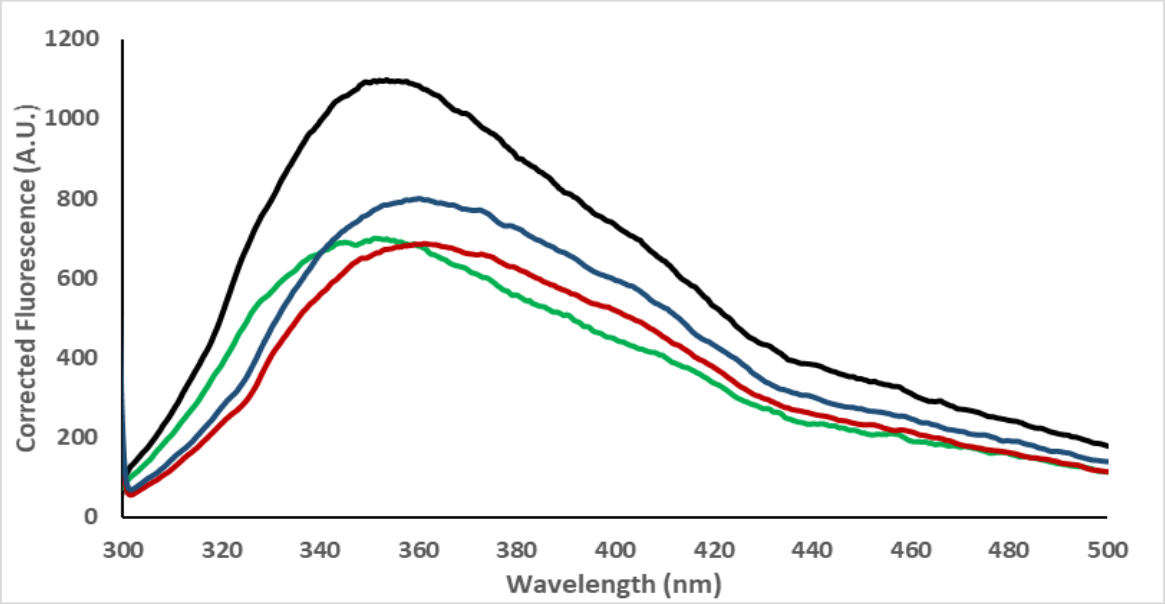


Figure 5.6: Intrinsic tryptophan fluorescence spectra of truncated NSP.

5.3.4. Extrinsic ANS fluorescence binding assay

The corresponding spectra in ***figure* 5.7** represent ANS bound to native truncated MP (red), ANS bound to denatured truncated MP (black) and free ANS in buffer (blue). All samples were prepared in 100 mM sodium phosphate, pH 6.5, 1 mM EDTA and 0.02 % NaN_2. The unfolding of the truncated MP alters the polarity of pockets permitting ANS binding, due to more polar surfaces being exposed.

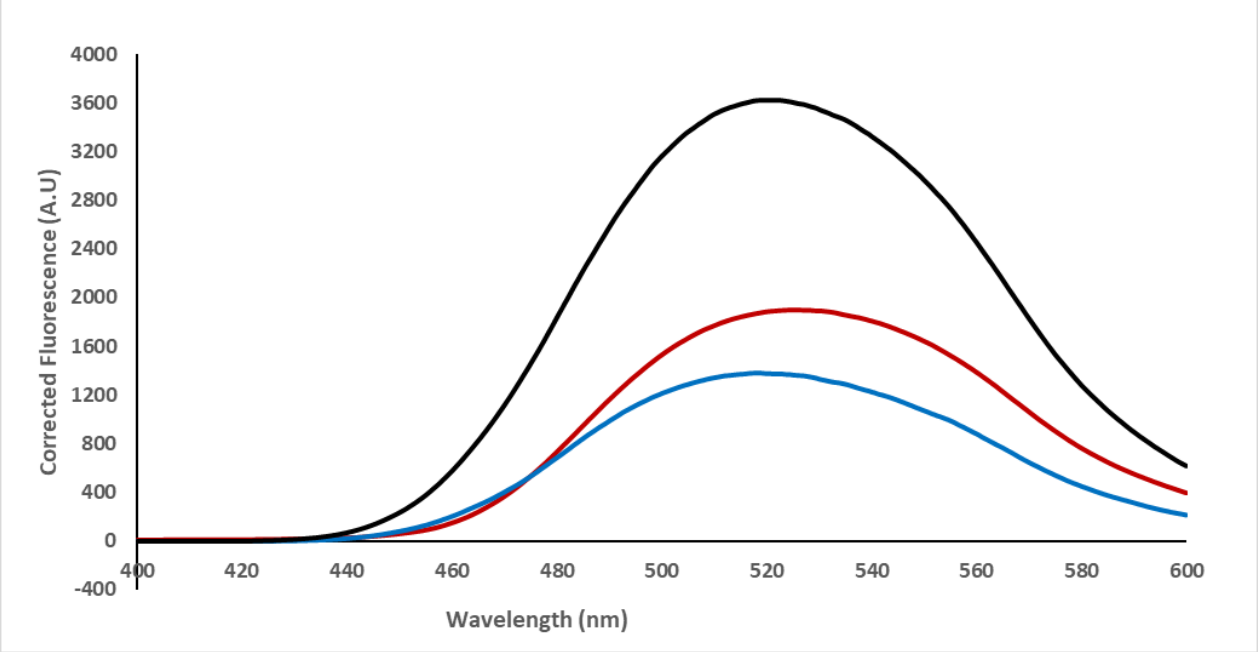


Figure 5.7: Extrinsic ANS fluorescence spectra of truncated MP.

The corresponding spectra in ***figure* 5.8** represents ANS bound to native truncated NSP (red), ANS bound to denatured truncated NSP (black) and free ANS in buffer (blue). All samples were prepared in 100 mM sodium phosphate, pH 6.5, 1 mM EDTA and 0.02 % NaN_2. The results show that the unfolding of the truncated NSP alters the polarity of pockets permitting ANS binding.

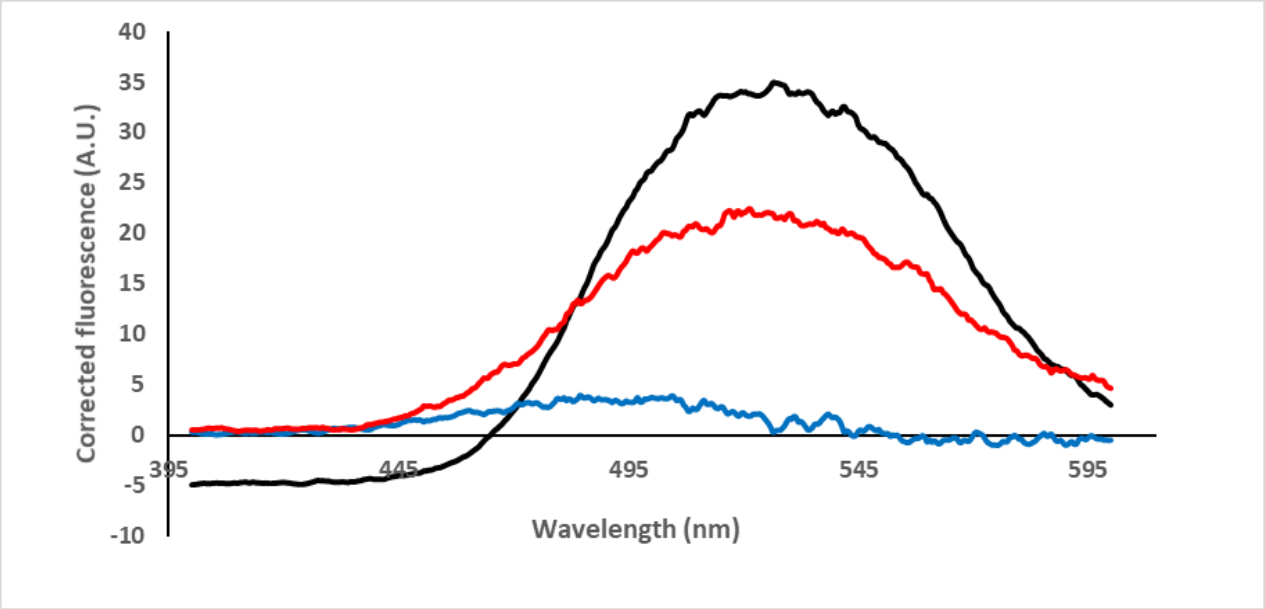


Figure 5.8: Extrinsic ANS fluorescence spectra of truncated NSP.

5.3.5. Intrinsic Fluorescence MANT-ATP binding Assay

The spectra below **(*figure* 5.9)** represent Mant-ATP bound to native truncated MP (red) in refolding buffer (50 mM Tris-HCl pH 7.5, 0.5 M L-Arginine), Mant-ATP bound to the truncated denatured MP (blue) in refolding buffer and free Mant- ATP in refolding buffer (black). All spectra were corrected for by subtracting from the free Mant-ATP spectrum. Results show that both the native and denatured forms of truncated MP both do not bind the Mant-ATP since the truncated MP might not have the affinity for the Mant-moeity.

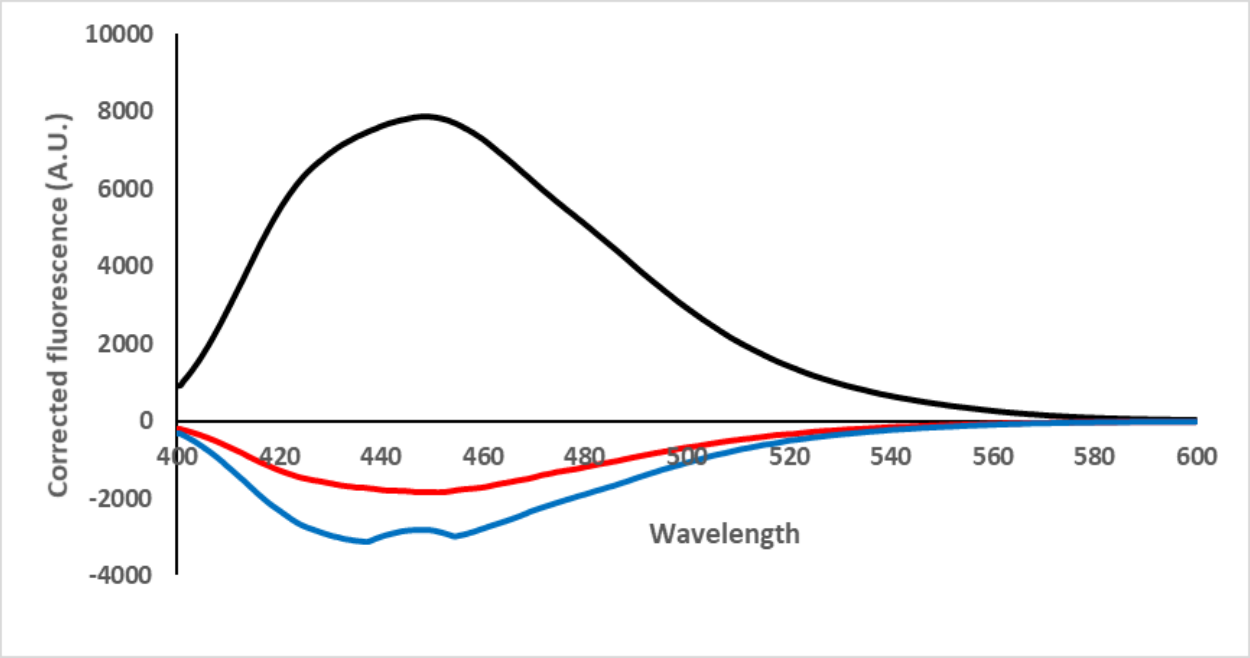


Figure 5.9: Intrinsic fluorescence spectra of truncated MP.

The spectra below **(*figure* 5.10)** represent Mant-ATP bound to native truncated NSP (red) in refolding buffer (50 mM Tris-HCl pH 7.5, 0.5 M L-Arginine), Mant-ATP bound to the truncated denatured NSP (blue) in refolding buffer and free Mant- ATP in refolding buffer (black). All spectra were corrected for by subtracting from the free Mant-ATP spectrum. Results showed that both the native and denatured forms of truncated MP both do not bind the Mant-ATP, as the spectra for both the native and denatured proteins have extremely low fluorescence intensities.

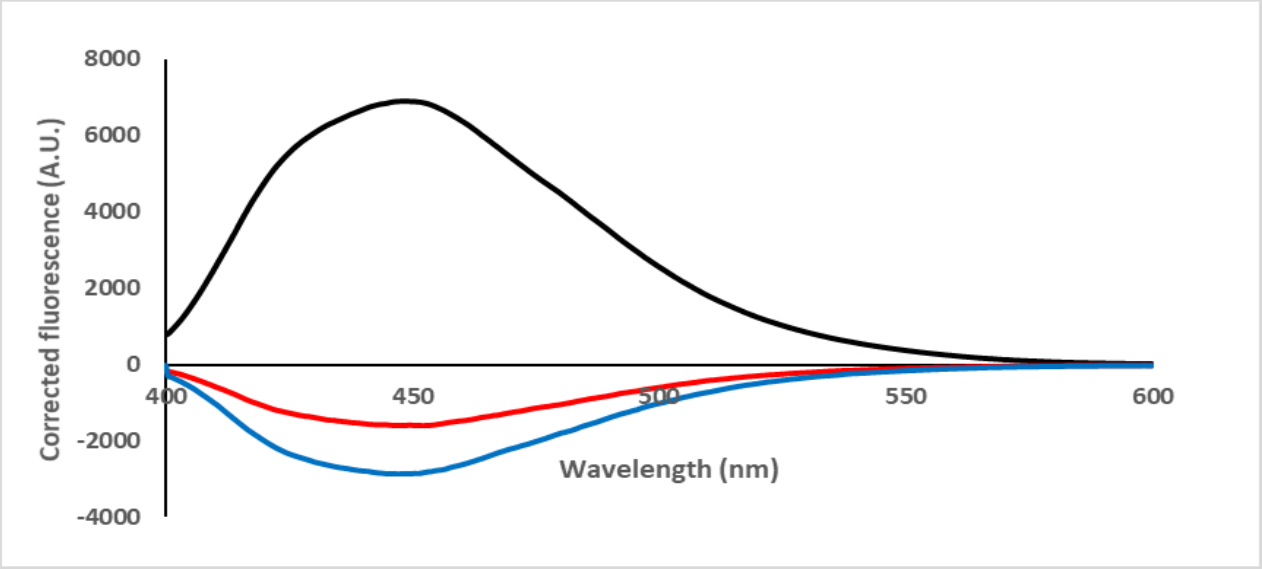


Figure 5.10: Intrinsic fluorescence spectra of truncated NSP.

5.3.6. Phosphatase assay

The phosphate assay was used for the determination of free phosphate in a sample as this indicates a phosphatase function for SACMV MP. For free phosphate to be present, there is a clear to teal green colour change corresponding to the standards **(*figure* 5.11)** (standards information available in **Appendix A**). Results clearly indicated phosphatase activity of the truncated MP **(*figure* 5.11)**. The amount of free phosphate released from the truncated MP was 41.16 µM calculated from the equation of the line of best fit (***figure* 5.12**) using the phosphate concentrations of the standards 1-8 supplied from the kit **(Appendix A)**.

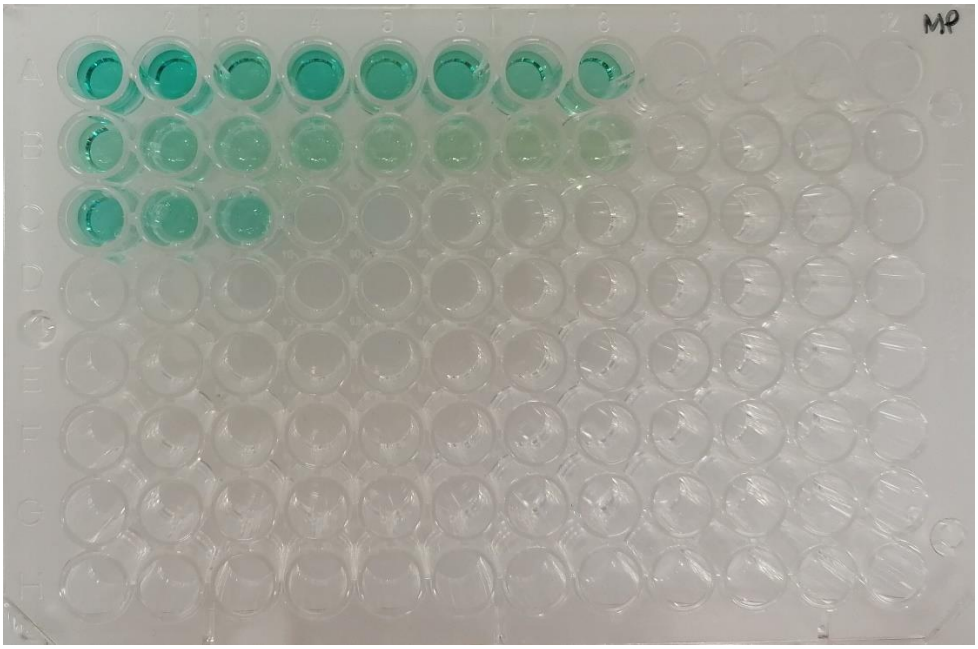


Figure 5.11: 96 well plate showing the colour changes for the phosphate standards and the MP sample (Standards: rows 1-8, columns A-B (according to kit specifics **(Appendix A)** MP sample : rows 1-3, column C).

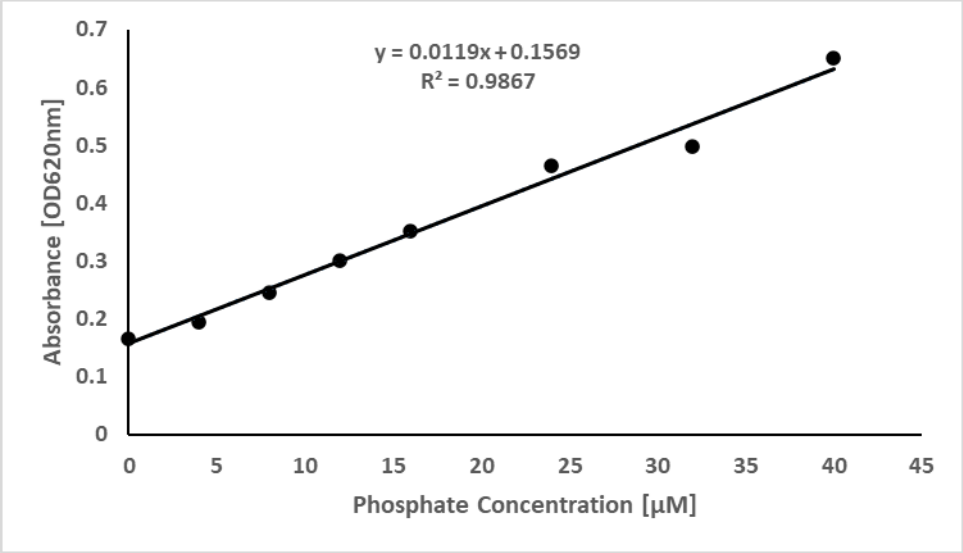


Figure 5.12: Quantification of released phosphate from hydrolysed ATP by the truncated MP.

Phosphate from hydrolysed ATP by the truncated NSP was detected by a clear to teal green colour change corresponding to the standards **(*figure* 5.13)**.

The amount of free phosphate released from the truncated NSP was 46.94 µM, as calculated using the equation of the line of best fit **(*figure* 5.14)** and using the phosphate concentrations of the standards 1-8 supplied from the kit **(Appendix A)**.

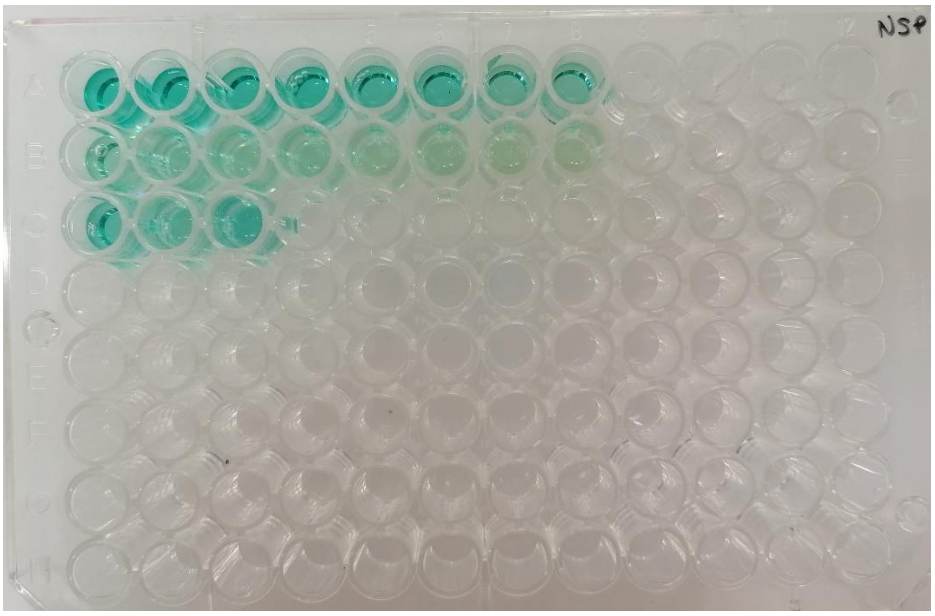


Figure 5.13: 96 well plate showing the colour changes for the phosphate standards and the NSP sample (Standards: rows 1-8, columns A-B (according to kit specifics **(Appendix A)** MP sample: rows 1-3, column C).

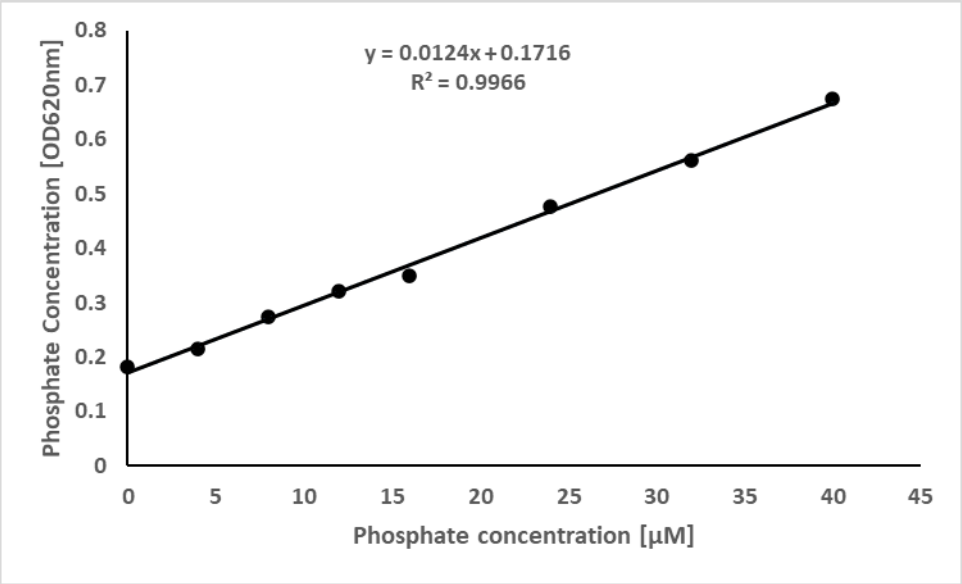


Figure 5.14: Quantification of released phosphate from hydrolysed ATP the by thetruncated NSP.

5.3.7. Size exclusion high performance liquid chromatography (SE-HPLC)

The spectra below **(*figure* 5.15)** showing the native truncated MP eluting at a retention time of 24.8 min in the absence of ATP (blue line), with a corresponding size of 1 kDA which does not correspond to the predicted size of the protein or the size of the protein (23 kDa) when isolated protein overexpression. The native truncated MP in the presence of ATP (red line) (5 mM) elutes at a retention time of 22.6 min, with a corresponding size of 10 kDA which also does not correspond to the predicted size of the protein, and this peak could be indicative of the ATP alone eluting through the high-pressure system. Both samples were plotted in comparison to the HPLC standards (black line).

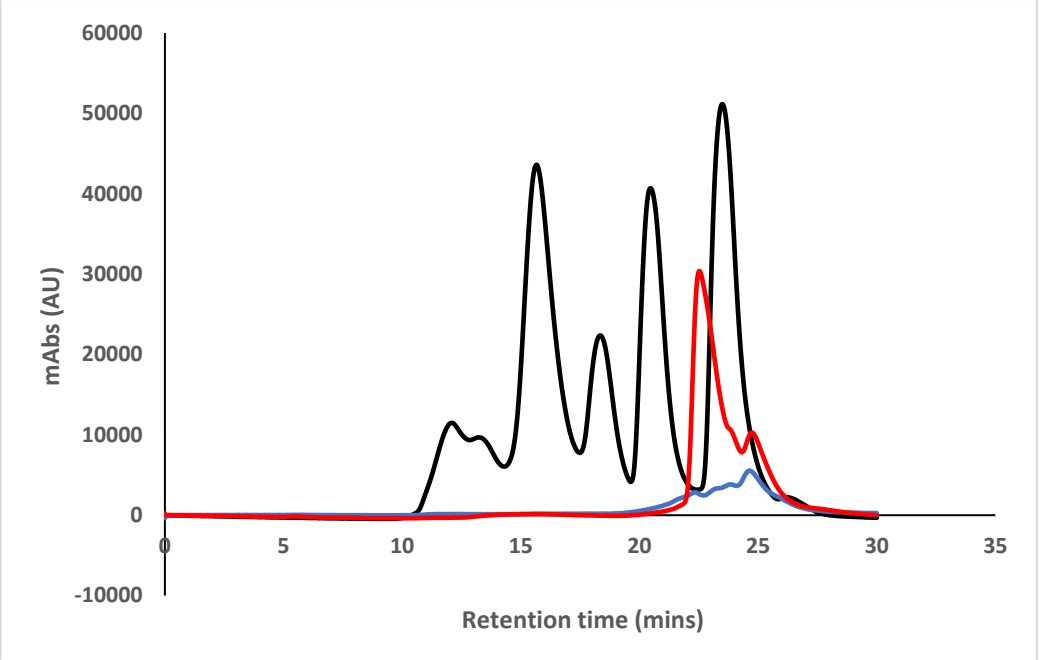


Figure 5.15: SE-HPLC spectra of the native truncated MP in the presence (red line) and absence (blue line) of ATP. The black line represents HPLC standards.

The below spectra **(*figure* 5.16)** shows the native truncated NSP (blue) eluting at a retention time of 24.8 min (no definitive peak), with a corresponding size of 1 kDA which does not correspond to the predicted size of the protein or the size of the protein (13 kDa) when isolated during protein overexpression. The native truncated NSP in the presence of ATP (red) (5 mM) elutes at a retention time of 22.6 min, which is close to the size of the truncated NSP, however, the corresponding size of the peak at 10 kDA, this could be indicative of the ATP alone eluting through the high-pressure system. Both NSP in the presence and absence of ATP were plotted in comparison to the HPLC standards (black).

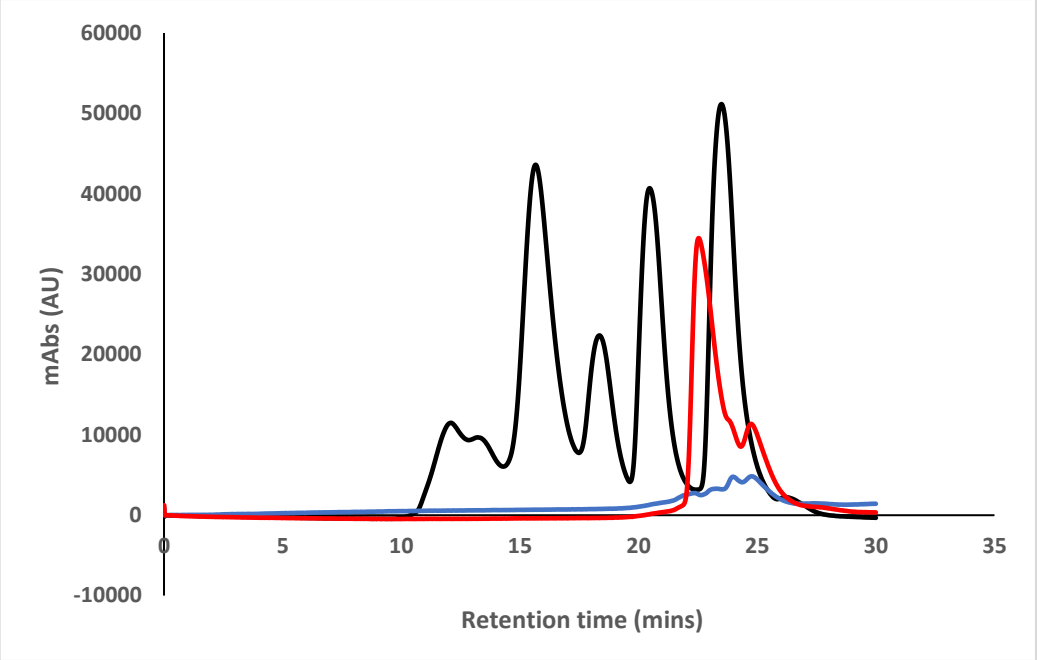


Figure 5.16: SE-HPLC spectra of the native truncated NSP in the presence and absence of ATP.

The standard curve **(*figure* 5.17)** represents the standards Gamma Globulin (158 kDa), Ovalbumin (44 kDa), Myoglobulin (17 kDa) and Vitamin B12 (1.5 kDa). The logs of these standards were used to calculate the molecular weight for the truncated MP and NSP.

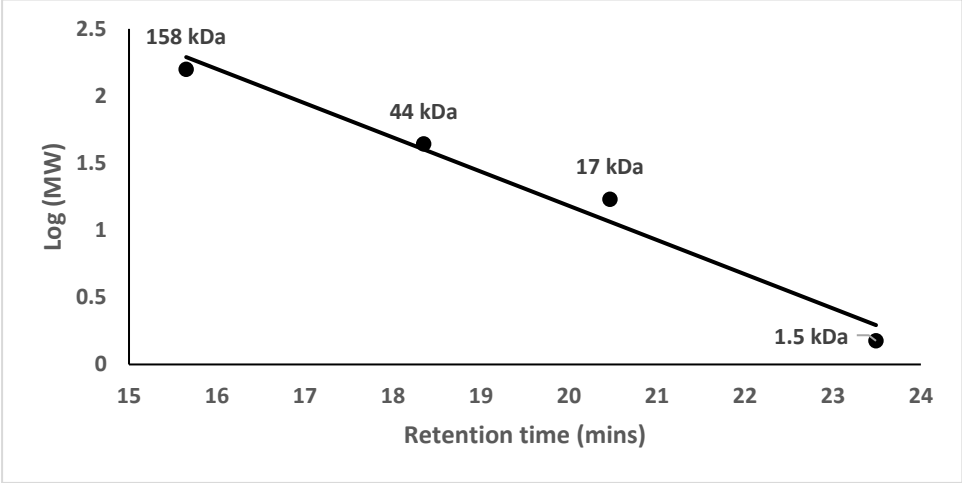


Figure 5.17: Standard curve used for weight determination of truncated MP and NSP.

5.4. Discussion

The secondary structure of an *in vitro* purified truncated SACMV MP and NSP in this study was successfully achieved. These structural characterizations represent novel research as these types of structural studies have not been reported to date in plant geminiviruses. Unravelling the structural components of these proteins will aid in further understanding their function *in vitro*. A few functional studies have been performed on the NSP (Fontes et al., 2004) and the cell-to-cell MP (Rojas et al., 1998) which could infer structural information about these elusive proteins. Studies which have been carried out on the begomovirsues, Cabbage leaf curl virus (CaLCV) and Squash Leaf Curl Virus (SqLCV), have led to the suggestion that MP and NSP of SACMV facilitate similar functions in cell-to-cell movement and shuttling of viral proteins in and out of the nucleus (Sudarshana et al., 1998).So while functional studies have been performed, a great more on the tertiary and quaternary (molecular) structures of geminiviral proteins in future. are required. Ultraviolet visible absorption spectroscopy is a popular tool for the assessment of protein quantification and purity (Wilson and Walker, 2012). The absorbance attained at 340 nm is subtracted from the absorbance attained at 280 nm, this gives a corrected absorbance or correlation factor (Wilson and Walker, 2012). The quantity of truncated MP and NSP attained **(*figures* 5.1 and 5.2)** are in the lower range and this is due to the low yield of pure protein retrieved and refolded **(chapter 4)**. However, based on the R^2 values (0.9998 and 0.9982) of the curve, the purity of the protein is considered to be relatively high (the lower the R^2 value the lower the purity of the protein). However comparisons of protein expression and quantification could not be compared to other studies as they have not been performed.

Attaining pure protein is a pre-requisite in all structural characterization research studies (Wilson and Walker, 2012). Far-UV CD has revealed that the truncated MP **(*figure* 5.3)** possesses a predominant β-sheet secondary structure, however, based on the spectra profile and the confirmation of the secondary structure using the Dichroweb database, it can also be seen that the protein possesses a high random coil structure **(*table* 5.1)**. The β-sheet secondary structure is predominant at 41% **(*table* 5.1)** while the random coil secondary structure composition of the protein tails at 39% **(*table* 5.1)**. This indicates that while the truncated MP contains some stable structure it also possesses a high level of disorder which has been shown to be a trait

of the full length MP due to the large number of intrinsically disordered regions found along the protein length (Nankoo, 2018). The truncated NSP both in its native and denatured forms displayed a predominant β-sheet secondary structure conformations, as observed by the spectra obtained **(*figure* 5.4)**. While the Dichroweb online server confirms this using the CONTIN formula the truncated NSP in its native and denatured form also possesses a high percentage of random coils indicating disorder **(*table* 5.2)**. While the spectra for the native protein display a peak of lower intensity, the denatured protein displays a peak of higher intensity **(*figure* 5.4)**. The SACMV truncated MP and NSP both show evidence of disorder by the high levels random coils present within their secondary structure compositions. This is not uncommon for viral proteins, in particular membrane bound viral proteins which shuttle between the nucleus and the cytoplasm whose protein structures are also principally disordered (Na et al., 2018, Xue et al., 2012).

The intrinsic tryptophan fluorescence indicates that the truncated MP, even in the presence of ATP, displays no conformational changes **(*figure* 5.5)**. In the denatured form the truncated MP displays a very low intrinsic fluorescence intensity and also shows no difference in spectra in the presence and absence of ATP, thereby displaying no conformational changes.In contrast the intrinsic tryptophan fluorescence indicated that the truncated NSP in its native form, displayed a conformational change in the presence of ATP **(*figure* 5.6)**. ATP caused a significant decrease in the intrinsic fluorescence peak at 350 nm, indicating that some of the tryptophan residues of the protein were not exposed for fluorescence detection. (Ghisaidoobe and Chung, 2014). In the denatured form the intrinsic fluorescent peak increased marginally with the presence of ATP, indicating that the conformational change was almost negligible. The clear differences in fluorescent peak intensities between the native and denatured states of both the truncated MP and NSP further indicate that these proteins were refolded after purification as intrinsic tryptophan fluorescence is used as a monitoring tool to assess protein refolding in addition to conformational changes to a protein (Ghisaidoobe and Chung, 2014).

While the extrinsic ANS fluorescence binding assay showed that both the NSP and MP proteins displayed hydrophobic surfaces for the binding of ANS, the denatured forms have a greater ability to bind ANS due to the increased hydrophobic surface area made available due to denaturation (Hawe, 2007). The binding of ANS to both forms of NSP caused a noisy spectrum and this is possibly due to the properties of the ANS fluorophore causing instability to the truncated NSP. Confirmation of refolding of both proteins was demonstrated since the ANS fluorescence binding assay also displayed a clear difference in spectra between native and denatured forms of both proteins. Fluorescent ANS binding assays are often used to monitor protein folding/unfolding (Hawe et al., 2007).

It was proposed that NSP and MP would bind ATP because based on the ability of the proteins to bind ligands, however there has been no evidence in literature of ATP binding to either protein. However, the intrinsic fluorescent Mant-ATP binding assays indicated that neither the truncated MP nor the truncated NSP were able to bind the Mant-ATP fluorophore **(*figures* 5.9 and 5.10)** due to the negative spectra seen for both these proteins in the native and denatured forms. While the Mant-ATP fluorophore did not bind to the proteins, it does not unequivocally dispute the ATP binding ability of these proteins. Mant is a fluorescent moiety bound to ATP, and it is proposed that the Mant moiety was unable to bind to the proteins due to the properties of the proteins which prevent interaction of the Mant fluorophore with the proteins (Ni et al., 2000).

The phosphatase assays displayed a distinct colour change for the truncated native MP **(*figures* 5.11 and 5.13)**, indicating that the refolded protein has functional activity since it was able to hydrolyse ATP to form ADP and AMP, releasing free phosphate (Wilson and Walker, 2012). The demonstration of the ability of both these SACMV truncated proteins to hydrolyse ATP to simpler adenosine forms releasing free phosphate, represents their functional requirement for energy to function. Adenosine 5′-triphosphate (ATP) is involved in a variety of cellular processes, including the virus life cycle, and cell-to-cell movement in which ATP-dependent reactions are catalysed by viral-encoded proteins or complexes consisting of viral and host-cell proteins (Ando et al., 2012). Heat shock proteins such as heat shock protein 70 (HSP70) have also been shown to possess ATPase activity, and since HSP70 chaperones form interactive complexes with geminivirus MPs and NSPs (Fontes et al., 2004; Krentz et

al., 2012), they may additionally perform an ATPase function during movement. It is proposed that the ATPase activity of SACMV MP and NSP and putative interactions of the SACMV DNA-MP-NSP complex with HSP70 would facilitate cytoplasmic movement from the nucleus to the cell periphery and across the plasmodesmata (Krenz et al., 2012).

The HPLC results for both truncated proteins (MP and NSP) did not elute at the expected retention times. Since the SE-HPLC was unsuccessful for both these truncated proteins (data not shown), and since there was an absence of multiple peaks for either of the proteins, it could not be confirmed that these proteins exist as oligomers and the size of both truncated proteins could not be confirmed. Since fluorescence studies indicate that both these truncated proteins contain hydrophobic regions, this could explain why these proteins elute at later than expected retention times as they bind to the beads in the stationary phase of the column for longer periods of time rather than passing through with the mobile phase. It has also been well documented that those insoluble proteins are more problematic to run through a SE-HPLC column with successful results than soluble proteins (Wilson and Walker, 2012; Yap et al., 2010).

5.5. Conclusions

Experimental results in this study have revealed, for the first time, secondary structural characteristics of two SACMV truncated proteins, MP (N-ter) and NSP (C-ter). Both these proteins have been found to possess a predominant β-sheet secondary structure in their native form, while the denatured form of the truncated NSP shared the same secondary structure profile. The truncated MP displayed no changes in conformation in its native or denatured states and in the presence of ATP, while the truncated NSP highlighted a conformational change in its native state in the presence of ATP. However no conformational change was demonstrated in the NSP denatured state even in the presence of ATP. Extrinsic fluorescence studies for both proteins, indicated that more hydrophobic patches were exposed for the binding of the ANS fluorophore in the denatured form. Both the extrinsic fluorescent study and the intrinsic tryptophan fluorescent study confirmed that both truncated proteins were refolded *in vitro*. Based on intrinsic Mant-ATP studies carried out on both truncated proteins, it can be confirmed that the Mant moiety fluorescent fluorophore does not bind to either protein in both the native and denatured forms. This however does not unequivocally disprove that ATP does not bind to these proteins, and further studies need to be undertaken in this regard. Both the truncated MP and the truncated NSP display functionality in the form of phosphatase activity, indicated by the release of free phosphate by hydrolysing ATP to simpler forms of adenosine and free phosphate. Possibly due to both truncated proteins containing wide hydrophobic surface regions and being insoluble in their native form, SE-HPLC was unsuccessful.

Chapter 6

Conclusions

Many functional studies have been performed on the geminiviral nuclear shuttle proteins (Fontes et al., 2004) and cell-to-cell movement proteins (Rojas et al., 1998), which could infer or predict secondary structure. However little is known about the structure of these elusive proteins. MP and NSP of phloem-limited geminiviruses, such as Squash leaf curl virus, have been shown to co-localize with tubules (Lazarowitz, 1999). The intracellular and intercellular transport of geminiviral DNA-protein infers association with the cytoskeleton-endoplasmic reticulum highways in the cell. These studies could possibly point to the highly insoluble nature of MP and NSP. The purification of the truncated NSP in the study proved more difficult than the truncated MP, due to its insolubility, likely due to the predicted intrinsically disordered regions in the truncated protein and the high composition of random coils in the secondary structure. The truncated MP while easier to purify, was also insoluble due to its predicted disordered regions reflecting the predicted and experimental secondary structure results retrieved, indicating the high presence of random coils in the secondary structure, which mirrors insolubility based on protein disorder (Linding et al., 2003). This may explain why to date no *in vitro* expression or purification studies have been reported for full length or truncated domains of these proteins. It is addtionally interesting to note that that no proteins in any database could be found with any appreciable homology to either of these proteins except from the same family of *Geminiviridae.* The origins of these geminiviral nuclear shuttle and cell-to-cell proteins therefore remain unsolved, and also cannot offer some possible function and structure clues. Certainly the highly disordered nature of these geminivirus, and other virus, proteins is indicative of their requirement to interact with several host proteins to perform their functions. Identification of other plant host interacting proteins (Krapp et al., 2017) may provide further insight in future studies into the structure-function evolution of these interesting proteins.

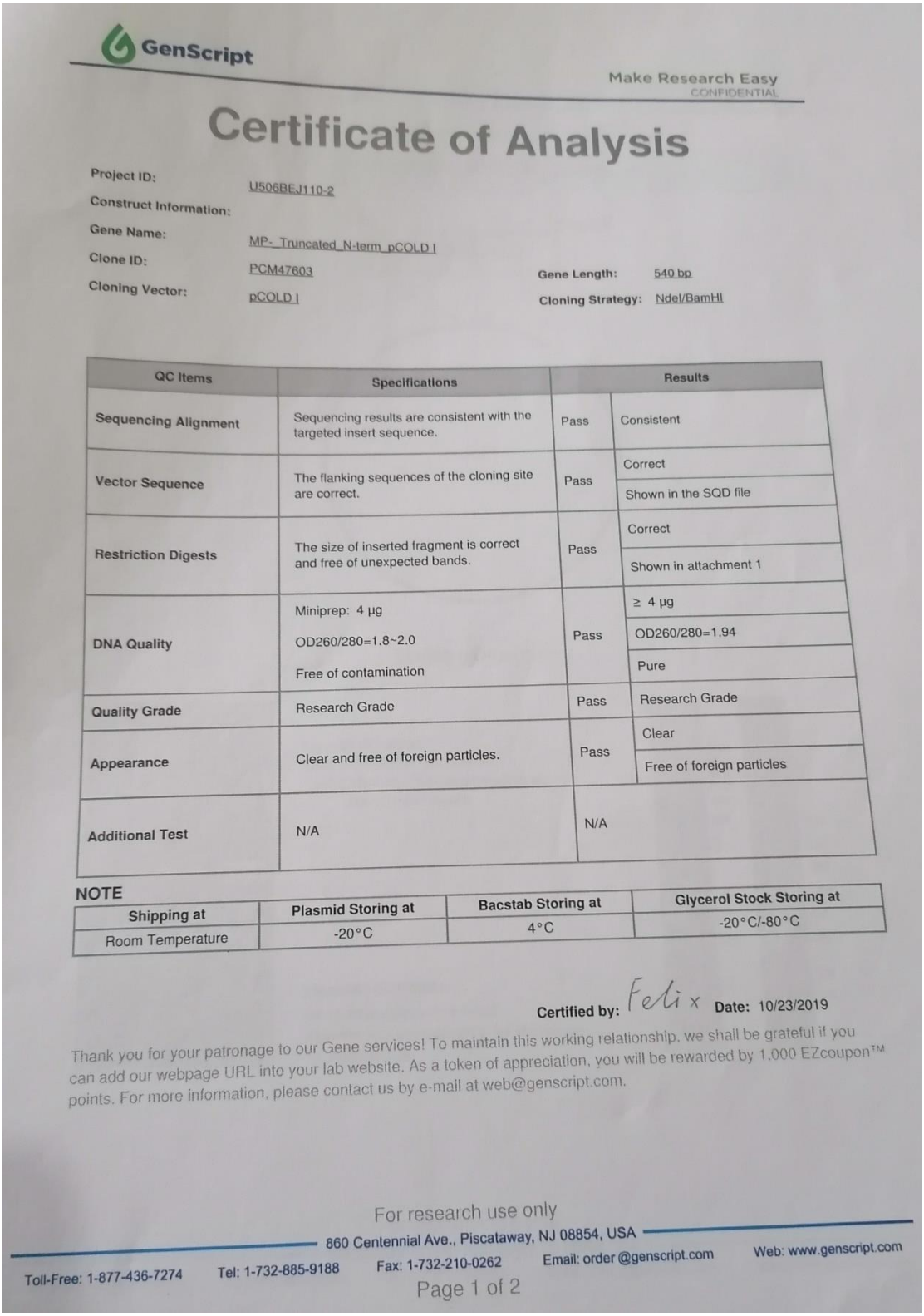
Using Far-UV-CD, predominant protein secondary structures were elucidated for both proteins, and these structures correspond to the results attained from the computational analysis. The extrinsic 8-Anilinonaphthalene-1-sulfonic acid fluorescent assay showed that both proteins were able to bind ligands with better affinity in the denatured state as there were more hydrophobic surfaces available for binding

compared with the native state. This assay also confirmed the truncated proteins were refolded as the assay is often used as a folding/unfolding tool. The integrity of the refolded proteins was further confirmed with the intrinsic tryptophan assays which confirmed that both proteins were indeed refolded. The intrinsic tryptophan fluorescence also highlighted that the presence of ATP caused a conformational change of the NSP in the native form, highlighting that possibly ATP could bind to the NSP C ter. However, the Mant-ATP derivative was unable to bind to either truncated protein in both the native and denatured states of the protein, indicating that ATP was unable to bind to both truncated proteins. Phosphatase assays indicated that both proteins possessed functional activity as they were able to convert ATP to ADP and AMP releasing free phosphate. This however could mean that only the Mant-moiety does not bind and possibly hinders the binding to the proteins. Finally, while the molecular weights of the truncated proteins were estimated by SDS-PAGE, to be 23 and 13 kDa respectively (truncated MP and NSP) further confirmation by size exclusion high pressure liquid chromatography was unsuccessful, possibly due to the insoluble nature of the proteins of the proteins.

In conclusion, the results of this research have contributed to the first of the expression, purification, and structural characterization of the pilot domain of the SACMV MP and the MP-binding domain of the SACMV NSP,using *in vitro* and computational approaches. Further research is still required to fully structurally characterize these truncated proteins. ATP binding studies using a different derivative may provide different results than those attained in the study. In future research mass spectrometry could be employed to replace the size exclusion high pressure liquid chromatography method for size confirmation of the proteins. Further research based on outcomes from this study could be to structurally characterize more domains of both these proteins in order to contribute to the understanding of the functions of these proteins.

Appendix A

Codon-optimized MP cloned into pCOLDI expression vector: Certificate of analysis



GenScript

Make Research Easy

CONFIDENTIAL

Certificate of Analysis

Project ID: U506BEJ110-2

Construct Information:

Gene Name: MP-_Truncated_N-term_pCOLD I

Clone ID: PCM47603

Gene Length: 540 bp

Cloning Vector: pCOLD I

Cloning Strategy: NdeI/BamHI

QC Items	Specifications	Results	
Sequencing Alignment	Sequencing results are consistent with the targeted insert sequence.	Pass	Consistent
Vector Sequence	The flanking sequences of the cloning site are correct.	Pass	Correct Shown in the SQD file
Restriction Digests	The size of inserted fragment is correct and free of unexpected bands.	Pass	Correct Shown in attachment 1
DNA Quality	Miniprep: 4 μg OD260/280=1.8~2.0 Free of contamination	Pass	≥ 4 μg OD260/280=1.94 Pure
Quality Grade	Research Grade	Pass	Research Grade
Appearance	Clear and free of foreign particles.	Pass	Clear Free of foreign particles
Additional Test	N/A	N/A	

NOTE

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

Certified by: Felix **Date:** 10/23/2019

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

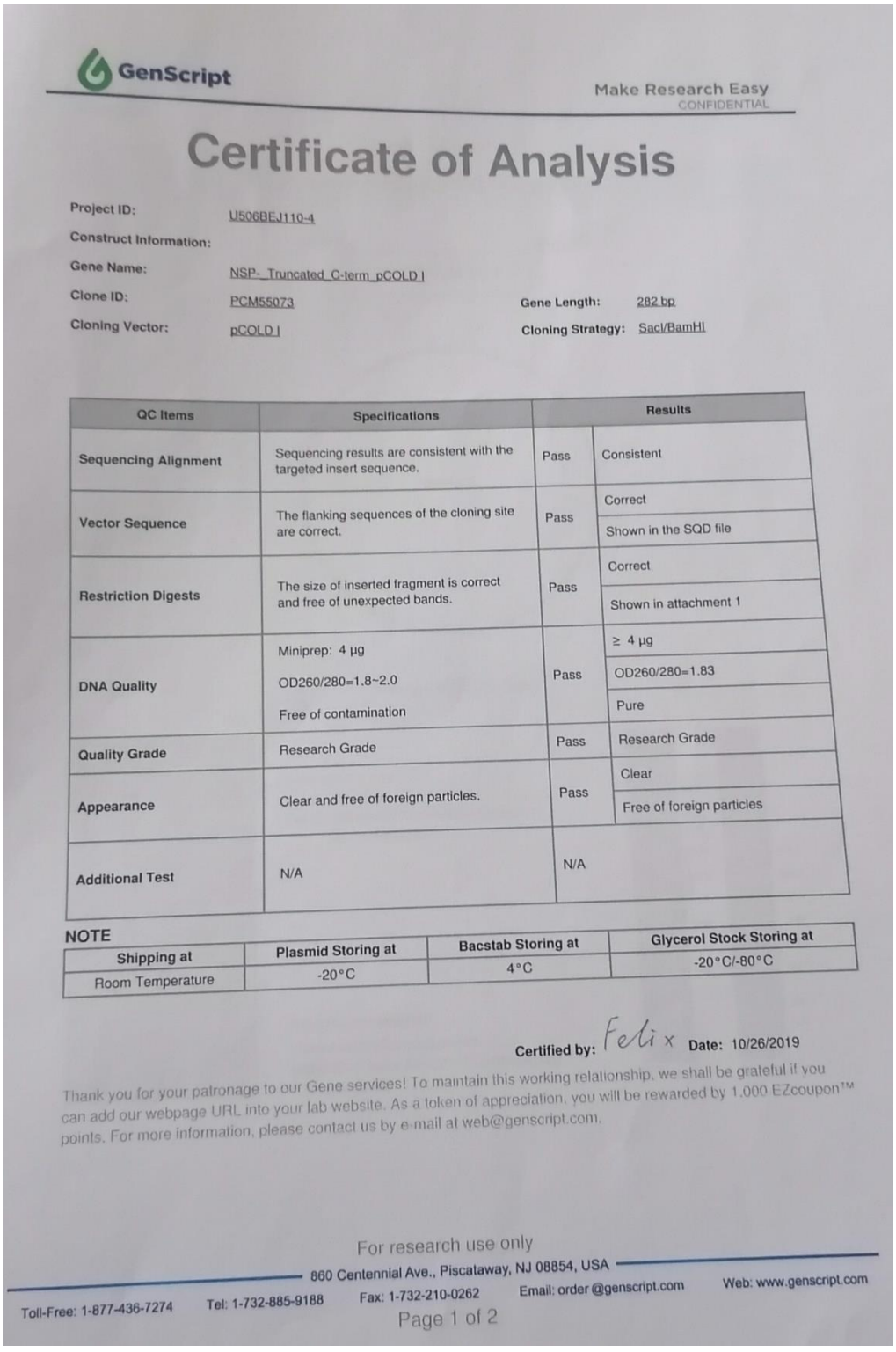
For research use only

860 Centennial Ave., Piscataway, NJ 08854, USA

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

Page 1 of 2

Codon-optimized NSP cloned into pCOLDI expression vector: Certificate of analysis



GenScript

Make Research Easy
CONFIDENTIAL

Certificate of Analysis

Project ID: U506BEJ110-4

Construct Information:

Gene Name: NSP-_Truncated_C-term_pCOLD I

Clone ID: PCM55073 **Gene Length:** 282 bp

Cloning Vector: pCOLD I **Cloning Strategy:** SacI/BamHI

QC Items	Specifications	Results	
Sequencing Alignment	Sequencing results are consistent with the targeted insert sequence.	Pass	Consistent
Vector Sequence	The flanking sequences of the cloning site are correct.	Pass	Correct Shown in the SQD file
Restriction Digests	The size of inserted fragment is correct and free of unexpected bands.	Pass	Correct Shown in attachment 1
DNA Quality	Miniprep: 4 μg OD260/280=1.8~2.0 Free of contamination	Pass	≥ 4 μg OD260/280=1.83 Pure
Quality Grade	Research Grade	Pass	Research Grade
Appearance	Clear and free of foreign particles.	Pass	Clear Free of foreign particles
Additional Test	N/A	N/A	

NOTE

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

Certified by: Felix **Date:** 10/26/2019

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

For research use only

860 Centennial Ave., Piscataway, NJ 08854, USA

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

Page 1 of 2

Cloning verification truncated MP into pCOLDI expression vector

GenScript

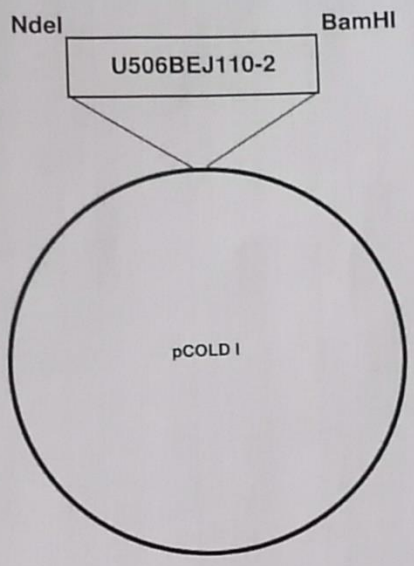
Make Research Easy

CONFIDENTIAL

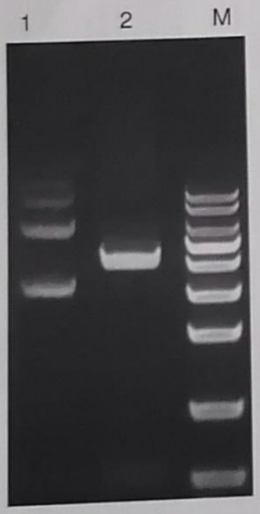
Attachment 1

Plasmid Construct Map

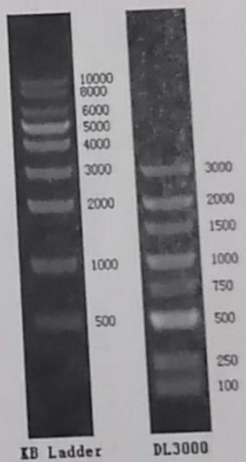
The gene was cloned in pCOLD I by NdeI/BamHI .



Enzyme Digestion



Lane M: KB Ladder
Lane 1:U506BEJ110-2 plasmid
Lane 2:U506BEJ110-2 plasmid digested by NdeI and BamHI



Digestion Conditions:
About 300ng plasmid digested
Digestion in water-bath,37°C for 40 minutes
1% Agarose Gel

For research use only
860 Centennial Ave., Piscataway, NJ 08854, USA
Toll-Free: 1-877-436-7274 Tel: 1-732-885-9188 Fax: 1-732-210-0262 Email: order @genscript.com Web: www.genscript.com

Cloning verification truncated NSP into pCOLDI expression vector

GenScript

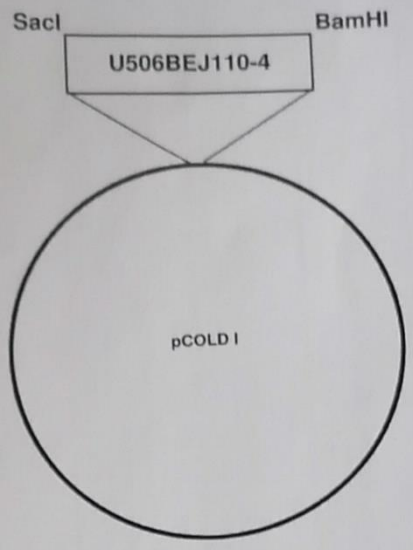
Make Research Easy

CONFIDENTIAL

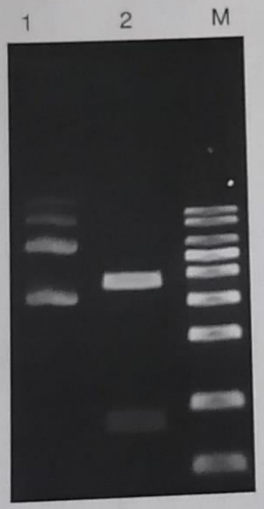
Attachment 1

Plasmid Construct Map

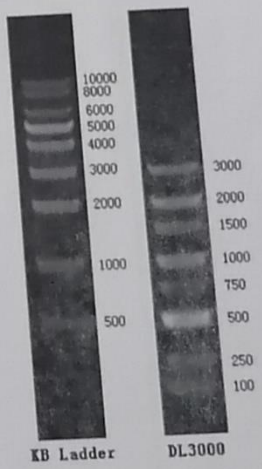
The gene was cloned in pCOLD I by SacI/BamHI .



Enzyme Digestion



Lane M: KB Ladder
Lane 1:U506BEJ110-4 plasmid
Lane 2:U506BEJ110-4 plasmid digested by SacI and MluI



Digestion Conditions:
About 300ng plasmid digested
Digestion in water-bath,37°C for 40 minutes
1% Agarose Gel

Phosphatase assay Protocol and Standard specification

Preparation Instructions
The components of this kit are ready to use. Briefly centrifuge vials before opening. To maintain reagent integrity, avoid repeated freeze/thaw cycles.

Storage/Stability
The kit is shipped at ambient temperature. Store all components at 2–8 °C upon receiving.

Procedure
All samples and standards should be run in duplicate. Use ultrapure water for the preparation of samples and standards.

Note: The Malachite Green Reagent must be brought to room temperature and well shaken before use. The Malachite Green Reagent is highly sensitive to phosphate. It is important that all enzyme preparations and assay buffers not contain free phosphate. Lab detergents may contain high levels of phosphate. Make sure that labware is washed thoroughly with ultrapure water and free from contaminating phosphate.

1. Dilution of phosphate standards. Prepare 1,000 µL of 40 µM phosphate Premix solution by mixing 40 µL of 1 mM phosphate standard with 960 µL of ultrapure water. Number the tubes. Prepare concentration standards by diluting the Premix as shown in Figure 1.

Figure 1
Preparation of Standards

#	Premix + water	Final Vol (µL)	Phosphate Conc (µM)	pmoles Phosphate in 50 µL
1	200 µL + 0 µL	200	40	2,000
2	160 µL + 40 µL	200	32	1,600
3	120 µL + 80 µL	200	24	1,200
4	80 µL + 120 µL	200	16	800
5	60 µL + 140 µL	200	12	600
6	40 µL + 160 µL	200	8	400
7	20 µL + 180 µL	200	4	200
8	0 µL + 200 µL	200	0	0

Transfer 50 µL of prepared standard in duplicate to wells in a clear-bottom 96 well plate. Store diluted standard at 2–8 °C for future use.

2. Transfer 50 µL of test sample (e.g., enzyme reaction) in duplicate into wells of the microplate. In the case of enzyme reactions, the reaction may be terminated by either adding a specific inhibitor, or can be stopped directly by the addition of the Malachite Green Reagent. Reaction buffer can be added as a blank control for the samples.

Notes: Precipitation may occur at high concentrations of phosphate (>100 µM), or in the presence of high concentrations of proteins and metals. If precipitation occurs, dilute samples in ultrapure water and repeat the assay.

Because any exogenous free phosphate would interfere with the assay, it is important to ensure the protein preparation, the reaction buffer, and labware employed in the assay do not contain free phosphate. This can be conveniently checked by adding the Malachite Green Reagent to the buffer and measuring the color formation.

3. Add 100 µL of the Malachite Green Reagent to each well. Mix by tapping the plate.

4. Incubate for 30 minutes at room temperature for color development.

5. Measure absorbance at 600–660 nm (620 nm) on a plate reader.

For cuvette assays, add 800 µL of Malachite Green Reagent to 400 µL of sample and standards. Perform the assay as described for the microplate assay.

Results
Calculation
Plot $OD_{620\ nm}$ versus phosphate standard concentrations. Use linear regression analysis to determine amount of free phosphate in the test samples.

Appendix B

Calculation for the protein concentration of truncated MP **(*figure* 5.1)**

$$\frac{\textbf{mg/ml = slope x MW (daltons)}}{\textbf{ϵ x path length}}$$

$$= \frac{0.2643 \times 20061.81}{21\ 555}$$

$$= 0.25 \text{ mg/ml}$$

Calculation for the protein concentration of truncated NSP **(*figure* 5.2)**

$$\frac{\textbf{mg/ml = slope x MW (daltons)}}{\textbf{ϵ x path length}}$$

$$= \frac{0.1273 \times 10297.71}{21\ 555}$$

$$= 0.07 \text{ mg/ml}$$

References

Adekunle, A., Orsat, V and Raghavan, V. (2016). Lignocellulosic bioethanol: A review and design conceptualization study of production from cassava peels. Renew. Sustain. *Energy Rev*, 64: 518–530.

Alabi, O.J., Kumar, P.L and Naidu, R.A. (2011). Cassava Mosaic Diesease: *A Curse to Food Security in Sub- Saharan Africa.* Online. APSnet Features. Doi10.1094/APSnetFeature-2011-0701.

Alabi, O.J., Kumar, P.L and Naidu, R.A. (2008). Multiplex PCR method for the detection of African cassava mosaic virus and East African cassava mosaic Cameroon virus in cassava. *Journal of Virological Methods*, 154: 111-120.

Allie, F., Pierce, E.J., Okoniewski, M.J and Rey.C. (2014). Transcriptional analysis of South African cassava mosaic virus-infected susceptible and tolerant landraces of cassava highlights differences in resistance, basal defense and cell wall associated genes during infection. *BMC Genomics*, 15:1006.

Amari, K., Vasquez, F and Heinlein, M. (2012). Manipulation of plant host susceptibility: an emerging role for viral movement proteins. *Frontiers in Plant Science*, 3(10): 1-7.

Ando, T., Imamua, H., Suzuki, R., Aizaki, H., Watanabe, T., Wakita, T and Suzuki, T. (2012). Visualization and Measurement of ATP levels in Living Cells Replicating Hepatitus C Virus Genome RNA. *Plos Pathogen*, 8(3): e1002561.

Ariyo, O. A., Dixon, A. G. O and Atiri, G. I. (2005). Whitefly Bemisia tabaci (Homoptera: Aleyrodidae) infestation on cassava genotypes grown at different ecozones in Nigeria. *J. Econ. Entomol*, 98: 611–617.

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

Baneyx, F. (1999). Recombinant protein expression in Escherichia coli. Current *Opinion in Biotechnology*,10: 411–421.

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

Berry, S and Rey, M.E.C. (2001). Molecular evidence for diverse populations of cassava-infecting begomoviruses in southern Africa Brief Report. *Achival Virology*, 46: 1795–1802.

Blagbrough, I.S., Bayoumi, S.A.L., Rowan, M.G and Beeching, J. (2010). Cassava: An appraisal of its phytochemistry and its biotechnological prospects. *Phytochemistry*, 71: 1940-1951.

Boevink , P and Oparka, K.J. (2005). Virus-host interactions during movement process. *Plant Physiol*, 138: 4–6.

Boulton, M. (2003). Geminiviruses: major threat to world agriculture. *Annals of Applied Biology*, 142(2).

Braun, P., Aubourg, S., Van Leene, J., De Jaeger, G and Lurin, C. (2013). Plant Protein Interactomes. *Annu. Rev. Plant Biol*, 64: 161–187.

Briddon, R.W and Stanely, J. (2006). Subviral agents associated with plant single-stranded DNA viruses. *Virology*, 344: 198-210.

Brown, J. K., Fauquet, C. M., Briddon, R. W., Zerbini, M., Moriones, E and Navas-Castillo, J. (2012). *Geminiviridae*. In Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses, ed. AMQ King, MJ Adams, EB Carstens, EJ Lefkowitz, pp. 351–73. London: Elsevier.

Chen, X., Xia, J., Xia, Z., Zhang, H., Zeng, C., Lu, C., Zhang, W and Wang, W. (2015). Potential functions of microRNAs in starch metabolism and development revealed by miRNA transcriptome profiling of cassava cultivars and their wild progenitor. *BMC Plant Biology*, 15(33).

Clark, E.D. (2001). Protein refolding for industrial processes. *Curr. Opin. Biotechnol*, 12: 202–207.

Czosnek, H., Eybistz, A., Sade, D., Gorovits, R., Sobol, I., Bejarano, E., Rosas-Díaz, T and LozanoDurán, R. (2013). Discovering Host Genes involved in the Infection by the Tomato Yellow Leaf Curl Virus Complex and in the Establishment of Resistance

to the Virus Using Tobacco Rattle Virusbased Post Transcriptional Gene Silencing. *Viruses*, 5:998-1022.

Desbiez, C., David, C., Mettouchi, A amd Gronenborn, B. (1995). Rep protein of tomato yellow leaf curl geminivirus has an ATPase activity required for viral DNA replication. *PNAS*, 92(12).

El-Sharkawy, M.A. (2004). Cassava biology and physiology. *Plant molecular biology*, 53: 621-641.

Esterhuizen, L.L., Mabasa, K.G., van Heerden, S.W., Czosnek, H., Brown, J.K., van Heerden, H and Rey, M.E.C. (2013). Genetic identification of members of the Bemisia tabaci cryptic species complex from South Africa reveals native and introduced haplotypes. *Journal of applied Entomology*, 137(1):122-135.

FAOSTAT. (2021). *Cassava world commodities*. Available at https:// www.faostat.com/cassava/indices (Accessed 20th October 2021).

Fauquet, C and Fargette, D. (1990). African Cassava Mosaic Virus: Etiology, Epidemiology, and Control. *Plant Disease*, 76(4): 404-411.

Fauquet, C.M., Briddon, R.W., Brown, J.K., Moriones, E., Stanley, J., Zerbini, M and Zhou, X. (2008). Geminivirus strain demarcation and nomenclature, *Archives of Virology*, 153: 783-821.

Fields, S., Kohara, Y and Lockhart, D.J. (1999). Functional genomics. *Proclaimed National Academy of Science USA*, 96: 8825-8826.

Fontes, E.P.B., Santos, A.A., Luz, D.F., Waclawovsky, A.J and Chory, J. (2004). The geminivirus nuclear shuttle protein is a virulence factor that suppresses transmembrane receptor kinase activity. *Genes and Development*, 18: 2545-2556.

Gaberc-Porekar, V and Menart, V. (2001). Perspectives of immobilized-metal affinity chromatography. *Journal of Biochemical and Biophysical Methods*, 49: 335–360.

Gafni, Y and Epel, B.L. (2002). The role of host and viral proteins in intra- and inter-cellular trafficking of geminiviruses. *Physiol. Mol. Plant Pathol*, 60: 231–241.

Garcia, M., and Dale, N. (1999). Cassava Root Meal for Poultry. *J. Appl. Poult. Res*, 8: 132–137.

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

Ghisaidoobe, A.B.T and Chung, S.J. (2014). Intrinsic Tryptophan Fluorescence in the Detection and Analysis of Proteins: A Focus on Főrster Resonance Energy Transfer Techniques. *International Journal of Molecular Sciences*, 15: 22518-22538.

Gutierrez, C. (2000). DNA replication and cell cycle in plants: learning from geminiviruses. *EMBO Journal*, 19: 792-799.

Gutierrez, C. (2002). Strategies of geminivirus DNA replication and cell cycle interference. *Physiological and Molecular Biology of Plants*, 60: 19-230.

Guitierrez, C., Ramirez-Parra, E., Mar Castellano, M., Sanz-Burgos, A.P., Luque, A and Missich, R. (2004). Geminivirus DNA replication and cell cycle interactions. Veterinary Microbiology, 98: 111-119.

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

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

Hanley-Bowdoin, L., Bejarano, E.R., Robertson, D and Mansoor, S. (2013). Geminiviruses: masters at redirecting and reprogramming plant processes. *Nature Reviews:Microbiology*, 10: 1-12.

Harries, P.A and Nelson, R.S. (2008). Movement of viruses in Plants.Encyclopedia of Virology, 3rd Edition. pp. 348-355. Elsevier Press.

Hehnle, S., Wege, C., and Jeske, H. (2004). Interaction of DNA with the Movement Proteins of Geminiviruses Revisited. *J. Virol*. 78, 7698–7706.

Henry, G. (2000). Global Cassava End Uses and Market, Production, Perspective and Future prospects FAO Bulletin. Rome, 85:242.

Hohn, T and Vazquez, F. (2011). RNA silencing pathways of plants: Silencing and its suppression by plant DNA viruses. *Biochimica et Biophysica Acta*, 1809: 588-600.

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

Hui, P. (2012). Next Generation Sequencing: Chemistry, Technology and Applications. *Topics in Current Chemistry*, 10: 1007-1028.

Inoue-Nagata, A. K., Lima, M. F and Gilbertson, R. L. (2016). A review of geminivirus diseases in vegetables and other crops in Brazil: current status and approaches for management. *Horticultura Brasileira*, 34(1): 8-18.

Irvine, G.B. (1997). Size-exclusion high-performance liquid chromatography of peptides: a review. *Analytical Chimical Acta*,352:387–397.

Jeske, H. (2009). Geminiviruses. In TT Viruses: The Still Elusive Human Pathogens, (Heidelberg: Springer Verlag), pp. 185–226.

Kaur, J., Kumar, A and Kaur, J. (2018). Strategies for optimization of heterologous protein expression in *E.coli*: Roadblocks and reinforcements. *International Journal of Biological Macromolecules*, 106:803-822.

Kelley, L.A and Sternberg, M.J.E. (2009). Protein structure prediction on the web: A case study using the phyre server. *Nat. Protoc,* 4: 363–373.

Kelly, S.M., and Price, N.C. (2000). The Use of Circular Dichroism in the Investigation of Protein Structure and Function. 349–384.

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

Kleinow, T., Nischnag, M., Beck, A., Kratzer, U., Tanwir, F., Priess, W., Kapp, G and Jeske, H. (2009). Three C-terminal phosphorylation sites in the Abuliton mosaic virus movement protein affect symptom development and viral DNA accumulation. *Virology*, 390(1):89-101.

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

Krentz, B., Jeske, H and Kleinow, T. (2012). The induction of stromule formation by a plant DNAvirus in epidermal leaf tissues suggests a novel intra- and intercellular macromolecular trafficking route. *Frontiers in Plant Science*, 3(291): 1-12.

Krenz, B., Windeisen, V., Wege, C., Jeske, H and Kleinow, T. (2010). A plastid-targeted heat shock cognate 70kDa protein interacts with the Abutilon mosaic virus movement protein. *Virology*, 401: 6–17.

Kuo, W.H.K and Chase, H.A. (2011). Exploiting the interactions between poly-histidine fusion tags and immobilized metal ions*. Biotechnol. Lett*, 33: 1075–1084.

Lakowicz, J.R. (2006). Plasmonics in biology and plasmon-controlled fluorescence. *Plasmonics*, 1: 5–33.

Lazarowitz, S.G and Beachy, R.N. (1999). Viral movement proteinsas probes for intracellular and intercellular trafficking in plants*. Plant Cell*, 11: 535-548.

Legg, J.P., Owor, B., Sseruwagi, P and Ndunguru, J. (2006). Cassava mosaic virus disease in East and Central Africa: epidemiology and management of regional pandemic. *Advances in Virus Resistance*, 67:355-418.

Legg J.P., Lava Kumar P., Makeshkumar T., Ferguson M., Kanju E., Ntawuruhunga P., Tripathi L Cuellar W. (2015). Cassava virus diseases: biology, epidemiology and management. *Adv. Virus Res*, 91:85–142.

Lesk, A.M. (2010). *Introduction to Protein Science*, 2nd Edition. Chapters 2, 3, 7 and 8. Oxford University Press. USA.

Levy, A and Tzfira, T. (2010). Bean dwarf mosaic virus: a model system for the study of viral movement. *Molecular Plant Pathology*, 11(4): 451-461.

Lewis, J.D and Lazarowitz, S.G. (2009). Arabidopsis synaptotagamin SYTA regulates endocytosis and virus movement protein cell-to-cell transport. *Proclaimed National Academy of Science USA* , 107(6): 2491-2496.

Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., and Russell, R.B. (2003). Protein Disorder Prediction. *Structure*,11: 1453–1459.

Mansoor, S., Briddon, R.W., Zafar, Y and Stanley,J. (2003). Geminivirus disease complexes: an emerging threat. *Trends in Plant Science*, 8: 128-134.

Mariano, A.C., Andrade, M.O., Santos, A.A., Carolino, S.M.B., Oliveira, M.L., Baracat-Pereira, M.C., Brommonshenkel, S.H and Fontes, E.P.B. (2004). Identification of a novel receptor-like protein kinase that interacts with a geminivirus nuclear shuttle protein. *Virology*, 318:24-30.

Mayer, M.P. (2010). Gymnastics of molecular chaperones. *Molecules and Cells*, 39:321-331.

McPherson, A. (1990). Current approaches to macromolecular crystallization. *European Journal of Biochemistry*, 189: 1–23.

Mertens, H.D.T., and Svergun, D.I. (2010). Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J. Struct. Biol*. 172, 128–141.

Micsonai, A., Wien, F., Kernya, L., Lee, Y.-H., Goto, Y., Réfrégiers, M and Kardos, J. (2015) Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc. Natl. Acad. Sci*, 112: 3095–3103.

Min, B. E., Martin, K., Wang, R., Tafelmeyer, P.,Bridges, M and Goodin, M. (2010). A Host- Factor Interaction and Localization Map for a Plant-Adapted Rhabdovirus Implicates Cytoplas- Tethered Transcription Activators in Cell-to-Cell Movement. *Molecular Plant-Microbe Interactions*, 23: 1420- 1432.

Moffat, A.S. (1999). Geminiviruses emerge as serious crop threat. *Science*, 286: 1835.

Montagnac, J. A., Davis, C. R and Tanumihardjo, S. A. (2009). Nutritional Value of Cassava for Use as a Staple Food and Recent Advances for Improvement. *Comprehensive Reviews in Food Science and Food Safety*, 8: 181–194.

Na, J.-H., Lee, W.-K., and Yu, Y. (2018). How Do We Study the Dynamic Structure of Unstructured Proteins: A Case Study on Nopp140 as an Example of a Large, Intrinsically Disordered Protein. *Int. J. Mol. Sci*. 19, 381.

Nankoo, N. (2018). ‘The recombinant expression and structural characterization of movement protein BC1 from South African cassava mosaic virus’, MSc, University of the Witwatersrand, South Africa.

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

Ohad, N., Shichrur, K and Yalovsky, S. (2007). The analysis of protein-protein interactions in plants by bimolecular fluorescence complementation. *Plant Physiology*, 145: 1090–1099.

Palmer, I and Wingfield, P.T. (2004). Preparation and Extraction of Insoluble (Inclusion-Body) Proteins from Escherichia coli. *Current Protocols in Protein Science*,1:1–25.

Parikh, I and Cuatrecasas, P. (1975). Affinity Chromatography, Principles and Applications. *Methods Protein Sep*: 3–29.

Patil, B.L and Fauquet, C.M. (2009). Cassava mosaic geminiviruses: actual knowledge and perspectives. *Molecular Plant Pathology*, 10(5): 685-701.

Patil, B.L and Fauquet, C.M. (2015). Studies on differential behaviour of cassava mosaic geminivirus DNA components, symptom recovery patterns, and their siRNA profiles. *Virus Genes*, 10.

Pavelić, K., Kralj, M., Kraljević, S and Sedić, M. (2008). Functional Genomics and proteomics. *Global approach to Biomedicine*, 21-30.

Pavlopoulos, G.A., Oulas, A., Lacucci, E., Sifrim, A., Moreau, Y., Schneider, R., Aerts, J and Iliopoulos, I. (2013). Unraveling genomic variation from next generation sequencing data. *BioData*, 6(13).

Pierce, E.J and Rey, M.E.C. (2013). Assessing Global Transcriptome Changes in Response to South African Cassava Mosaic Virus [ZA-99] Infection in Susceptible Arabidopsis thaliana. *PLOS ONE*, 8:6.

Reichel, C., Más, P and Beachy, R.N. (1999). The role of the Endoplasmic Reticulum and cytoskeleton in plant viral trafficking. *Trends in plant science*, 4(11): 458-462.

Rey, C and Vanderschuren, H. (2017). Cassava Brown Streak Diseases: Current Perspectives and Beyond. *Annual Reviews in Virology*, 4(1):429-452.

Roberts, A.G. and Oparka, K.J. (2003). Plasmodesmata and the control of symplastic transport. *Plant Cell Environment*, 26: 103-124.

Rojas, M., Donahue, J.P., Tan, Z and Lin, Y.-Z. (1998). Genetic engineering of proteins with cell membrane permeability. *Nat. Biotechnol*, 16: 370.

Rojas, M.R., Noueiry, A.O., Lucas, W.J., and Gilbertson, R.L. (1998). Bean Dwarf Mosaic Geminivirus Movement Proteins Recognize DNA in a Form- and Size-Specific Manner. *Cell*, 95: 105–113.

Seal, S.E., vandenBosch, F and Jeger, M.J. (2006). Factors influencing begomovirus evolution and their increasing global significance: implications for sustainable control. *Critical reviews in Plant Science*, 25: 23-46.

Sels, J., Mathys, J., De Conick, B.M.A, Cammue, B.P.A and De Bolle, M.F.C. (2008). Plant pathogenesis- related (PR) proteins: A focus on PR peptides. Plant *Physiology and Biochemistry*, 46: 941-950.

Shackelford, G.E., Haddaway, N.R., Usieta, H.O., Pypers, P., Petrovan, S.O and Sutherland, W.J. (2018). Cassava farming practices and their agricultural and environmental impacts: a systematic map protocol. *Environmental Evidence*, 7(30).

Shi, F., Telesco, S.E., Liu, Y., Radhakrishnan, R and Lemmon, M. A (2010). ErbB3/HER3 intracellular domain is competent to bind ATP and catalyze autophosphorylation. *Proclaimed National Academy of Science*. U. S. A,107: 7692–7697.

Shimizu, T., Yoshii, A., Sakurai, K., Hamada, K., Yamaji, Y., Suzuki, M., Namba, S and Hibi, T. (2009). Identification of a novel tobacco DnaJ-like protein interacts with the movement protein of tobacco mosaic virus. *Archives in Virology*, 154: 959-967.

Soellick, T., Uhrig, J.F., Bucher, G.L., Kellmann, J.W and Schreier, P.H. (2000). The movement protein NSm of Tomato spotted wilt tospovirus (TSWV): RNA binding, interaction with TSWV N protein, and identification of interacting plant proteins. *Proclaimed National Academy of Science USA*, 97:2373-2378.

Stanely, J., Bisaro, D.M., Briddon, R.W., Brown, J.K, Fauquet, C.M., Harrison, B.D., Rybicki, E.P and Stenger, D.C. (2005). Geminiviridae, In: Ball, L.A. (Ed), *Virus Taxonomy. VIIIth Report of the International Committee on Taxonomy of Viruses*. Elsevier/Academic Press, London, 301-326.

Sudarshana, M.R., Wong, H.L., Lucas, W.J and Gilbertson, R.L. (1998). Dynamics of Bean Dwarf Mosaic Geminivirus Cell-to-Cell ad Long-Distance movement in Phaseolus vulgaris Revealed, Using the Green Fluorescent Protein. *The American Phytopathological Society*, 11(4): 227-291.

Thresh, J.M and Cooter, R.J. (2005). Strategies for controlling cassava mosaic virus disease in Africa. *Plant Pathology*, 54: 587-614.

Unseld, S., Hőhnle, M., Ringel, M and Frischmuth, T. (2001). Subcellular targeting of the coat protein of African Cassava Mosaic Geminivirus. *Virology*, 286:373-383.

Vanderschuren, H., Stupak, M., Fütterer, J., Gruissem, W and Zang, P. (2007). Engineering resistance to geminiviruses- review and perspectives. *Plant Biotechnology Journal*, 5: 207-220.

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): 9487-9498.

Varsani, A., Roumagnac, P., Fuchs, M., Navas-Castillo, J., Moriones, E., Idris, A., Briddon, R.W., Rivera-Bustamante, R., Murilo Zerbini, F., and Martin, D.P. (2017). Capulavirus and Grablovirus: two new genera in the family Geminiviridae. *Arch. Virol*. 162, 1819–1831.

Vivian, J.T., and Callis, P.R. (2001). Mechanisms of tryptophan fluorescence shifts in proteins*. Biophys. J.* 80, 2093–2109.

Verchot, J. (2012). Cellular chaperones and folding enzymes are vital contributors to membrane bound replication and movement complexes during plant RNA virus infection. *Frontiers in Plant Science*, 3: 1-15.

Voet, D., Voet, J.G and Pratt, C.W. (2013). *Principles of Biochemistry*, 4th Edition. Chapters 4, 5 and 6. John Wiley and Sons Inc. USA

Waigmann, E., Ueki, S., Trutnyeva, K and Citovsky, V. (2004). The ins and outs of non-destructive cell-to-cell and systemic movement of plant viruses. *Critical Reviews in Plant Science*, 23: 195-250.

Ward, B.M., Medville, R., Lazarowitz, S.G and Turgeon, R. (1997). The geminivirus BL1 movement protein is associated with endoplasmic reticulum–derived tubules in developing phloem cells. *J. Virol*, 71: 3726–3733.

Wei, T., Zhang, C., Hou, X., Sanfacon, H and Wang, A. (2013). The SNARE Protein Syp71 Is Essential for Turnip Mosaic Virus Infection by Mediating Fusion of Virus-Induced Vesicles with Chloroplasts. *PLOS: Pathogens*, 9(5): 1-11.

White, K.P. (2001). Functional Genomics and The Study of Development, Variation and Evolution. *Nature Reviews: Genetics*, 2: 528-537.

Whitham, S.A and Wang, Y. (2004). Roles for host factors in plant viral pathogenicity. *Current Opinions in Plant Biology*, 7: 365-371.

Wilson, K and Walker, J. (2012). *Principles and Techniques of Biochemistry and Molecular Biology*, 7thEdition. Chapters 8,10 and 12. Cambridge University Press. USA.

Xue, B., Dunker, A.K., and Uversky, V.N. (2012). Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J. Biomol. Struct. Dyn*. 30, 137–149.

Yap, L., Sancheti, H., Ybanez, M.D, Garcia, J., Cadenas, E and Han, D. (2010). Chapter 6- Determination of GSH, GSSG, and GSNO Using HPLC with Electrochemical Detection. *Methods in Enzymology*, 413:137-147.

Zerbini, F.M., Briddon, R.W., Idris, A., Martin, D, P., Moriones, E., Navas-Castillo, J., Rivera-Bustamanta, R., Roumagan, P and Varsani, A. (2017). ICTV Virus Taxanonmy Profile: Geminiviridae. *Journal of General Virology*, 98(2):131-133.

Zhang, S.C., Ghosh, R and Jeske, H. (2002). Subcellular targeting domains of Abutilon mosaic geminivirus movement protein BC1. *Arch Virology*, 147: 2349-2363.

Zhang, Y., Gao, P and Yuan, J. (2010). Plant Protein-Protein Interaction Network and Interactome. *Curr. Genomics*, 11: 40–46.

Zhou, Y., Rojas, M.R., Park, M., Seo, Y., Lucas, W.J and Gilbertson, R.L. (2011). Histone H3 interacts and colocalizes with the Nuclear Shuttle Protein and the Movement Protein of a Geminivirus. *American Society for Microbiology*, 85(22): 11821-11832.