The recombinant expression and structural characterization of movement protein BC1 from *South African cassava mosaic virus*

School of Molecular and Cell Biology
University of the Witwatersrand
Johannesburg

Nikita Nankoo
389088

Supervised by: Prof M.E.C. Rey

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

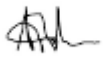
June 2018

Declaration

I, Nikita Nankoo (389088), am a student registered for the degree Master of Science (MSc) in the academic year 2018.

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)

6th June 2018

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 advisor Dr I. A. Achilonu and the University of Witwatersrand Protein Structure, Function and Research Unit for valuable input, support and equipment.

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

Special thanks must go to Miss Frances Ayres for her continued support and aid during the course of this research.

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

Finally I thank loved ones for their support and understanding, most importantly I thank for God for the courage and strength to see this research to completion.

Table of contents

List of tables

List of figures

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-term	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
DLS	Dynamic light scattering
dsDNA	Double stranded DNA
DTT	Dithiothreitol
Dpi	Days post infection
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
HCl	Hydrochloric acid
HSP	Heat shock protein
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl-beta-D-thiogalactopyranoside
IR	Intergenic region
IUPAC	International Union of Pure and Applied Chemistry
Kan	Kanamycin
KCl	Potassium chloride
kDa	Kilo-Dalton
LB	Luria Broth
MANT-ATP	2'(3')-N-methylanthraniloyl-ATP
$MgCl_2$	Magnesium chloride
$MgSO_4$	Magnesium sulphate
MP	Movement protein
N-term	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
PCR	Polymerase chain reaction
PD	Plasmodesmata
PDB	Protein Database
PK	Prokaryotic
PTG	Post transcriptional gene
PTGS	Post transcriptional gene silencing
REP	Replication-associated protein
SACMV	South African cassava mosaic virus
SCOP	Structural classification of proteins
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
SOB	Super optimal broth
SYTA	Synaptotagmin A
SqLCV	Squash leaf curl virus
ssDNA	Single stranded DNA
tet	tetracycline
TMV	Tobacco mosaic virus
TrAP	Transcriptional activator protein

TuMV	Turnip mosaic virus
TYLCV	Tomato yellow leaf curl virus
UV	Ultraviolet
VIGS	Virus induced gene silencing
WAG	Whelan and Goldman
Y2H	Yeast-two-hybrid

Abstract

South African cassava mosaic virus (SACMV) is a circular ssDNA bipartite begomovirus, whose genome comprises of DNA A (encodes six genes) and DNA B (encodes BC1 (cell-to-cell movement protein) and BV1 (nuclear shuttle protein)). Begomoviruses cause cassava mosaic disease (CMD) resulting in substantial root yield losses impacting farmers and potential starch yields for industrial purposes. The structural and physiochemical characteristics of begomoviruses have not been elucidated to date. Additionally, it is crucial to identify protein-protein interactions between SACMV BC1 and cassava host proteins that facilitate SACMV movement. In this study, *in silico* characterization studies for BC1 were undertaken. The FASTA amino acid sequence of BC1 was uploaded onto a webserver (Predict Protein (https://www.predictprotein.org)) and resulted in the predicted secondary structure of BC1 as well as predicted protein binding sites. BC1 was cloned into a pET-28a(+) expression vector and transformed into host cells *E.coli* BL21 (DE3). The optimal conditions for the expression of BC1 protein were found to be induction with 0.25 mM IPTG at an OD600 of ~0.6 expressed at 37 °C for four hours. The protein was refolded using stepwise dialysis. The molecular weight of BC1 was 35 kDa (SDS-PAGE). The secondary structure of BC1 was confirmed to be predominantly β-strands (CD). An ANS (1-anilino-8-naphthalene sulphonate) binding assay confirmed that BC1 possesses hydrophobic pockets with the ability to bind ligands such as ATP. A 2' (3')-N-methylanthraniloyl-ATP (MANT-ATP) assay showed binding of MANT-ATP to BC1. Intrinsic tryptophan fluorescence studies indicated significant conformational change in the denatured form of BC1 in the presence of ATP compared to in the absence of ATP. Literature and the PredictProtein webserver were employed to *in silico* predict suitable cassava host prey proteins for the yeast-two-hybrid (Y2H) system. Cloning of host proteins (HSP70 and Histone H3) into pDEST™22 plasmids was done at GenScript USA. The *in silico* study and structural characterization of BC1 provide insights to the structural characteristics of BC1 and further explains its functionality. Completion of Y2H assays will confirm or refute that cassava host proteins Histone H3 and HSP70 interact with the cell-to-cell movement protein of SACMV.

Chapter One: Literature Review

1.1. Introduction to cassava

Cassava (figure 1.1) (*Manihot esculenta Crantz*) is a woody, tropical, perennial shrub from the family *Euphorbiaceae* and is considered to be one of the most important crops in a number of Asian and African countries due to its ability to provide food security and an income for economically disadvantaged farmers in these regions. Typically cassava plants reach a height of 1-4 metres once they have reached physiological maturity. The tubers in the form of roots are harvested between 6 months to 4 years after planting and contain a dry matter content of between 30-40% (Adekunle *et al*., 2016; Ariyo *et al*., 2006; Legg *et al*., 2015).

In Africa cassava is the third most produced crop, tailing the production of sugar cane and maize (figure 1.3). Cassava is one of the highest suppliers of carbohydrates amongst staple crops, the tubers (figure 1.2) recovered from cassava plants are used for human and animal consumption as they provide an important dietary source (Adekunle *et al*., 2016; Ariyo *et al*., 2006; Berry and Rey, 2001; Blagbrough *et al*., 2010; El-Sharkawy, 2004; FAOSTAT, 2016; Henry, 2000; Legg *et al*., 2006). The leaves recovered from cassava plants are consumed as a green vegetable, prevalently in East Africa, where these leaves provide a source of proteins, minerals and vitamins (Alabi *et al*., 2011). Cassava is classed as an excellent crop as compared to similar crops such as wheat, rice and maize, 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 1.1: Visualisation of cassava leaves (Source: Google images, 2015).

Figure 1.2: Cassava tubers (Source: www.shutterstock.com).

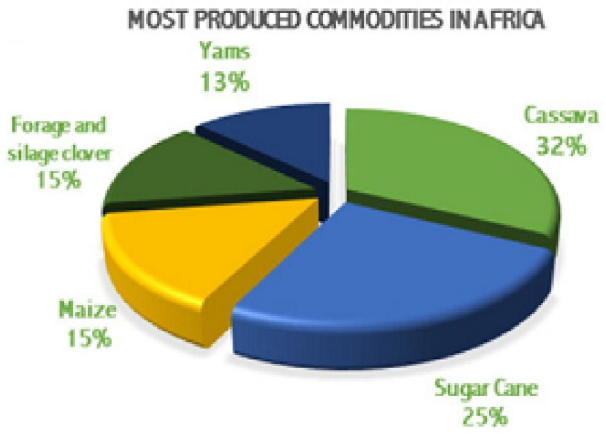


Figure 1.3: A pie chart showing the most produced commodities in Africa (by percentage) (Source: FAOSTAT, 2016).

1.1.1 Cassava production

Since the year 2000 the world average production of cassava has increased by roughly 100 million tonnes (1 tonne = 1.000kg). This was driven by a rising demand 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. An average 500 million people depend on cassava as a major carbohydrate thus making it the third largest carbohydrate source in the world for human food. At present the world average yield of cassava is 12.8 tonnes per hectare (world output of ~290 million tonnes (figure 1.4)), however, there is potential to produce an approximate of 23.2 tonnes of cassava per hectare (equivalent of ~500 million tonnes per year) (FAOSTAT, 2016; FAOSTAT,

2014; Fauquet and Fargette, 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 of 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 1.5) (FAOSTAT, 2016, 2017; 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, 2003).

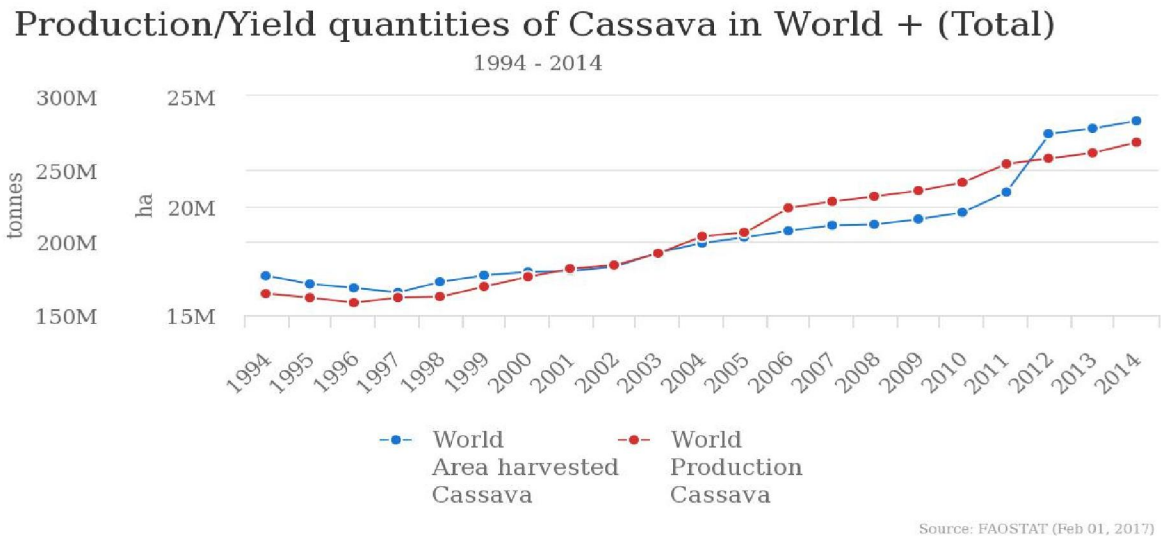


Figure 1.4: A line plot showing the amount of cassava harvested and the production of cassava worldwide between years 1994-2014 (Source: FAOSTAT, 2017).

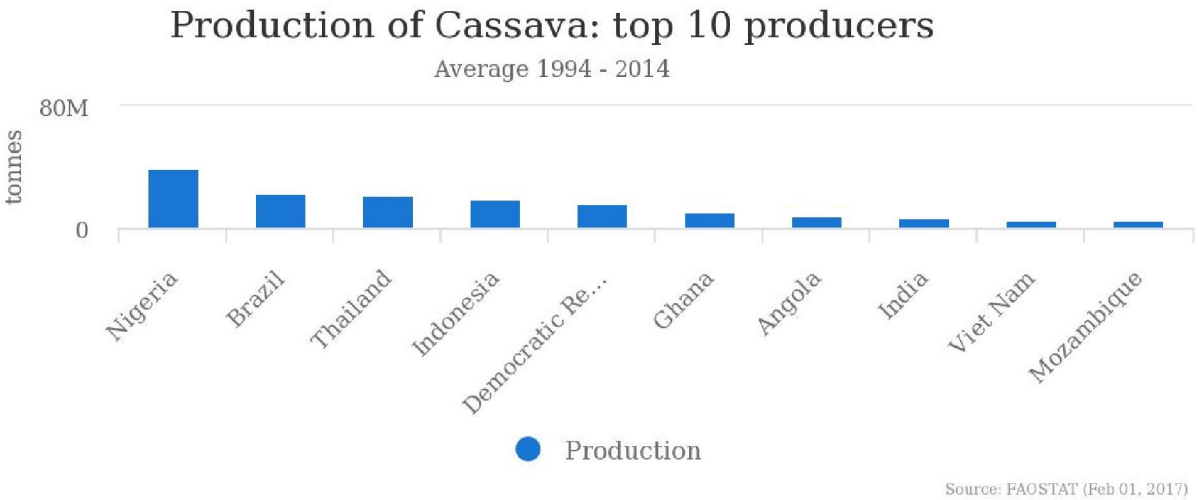


Figure 1.5: A bar graph showing the amount of cassava produced (in tonnes) from the top 10 cassava producers (Source: FAOSTAT, 2017).

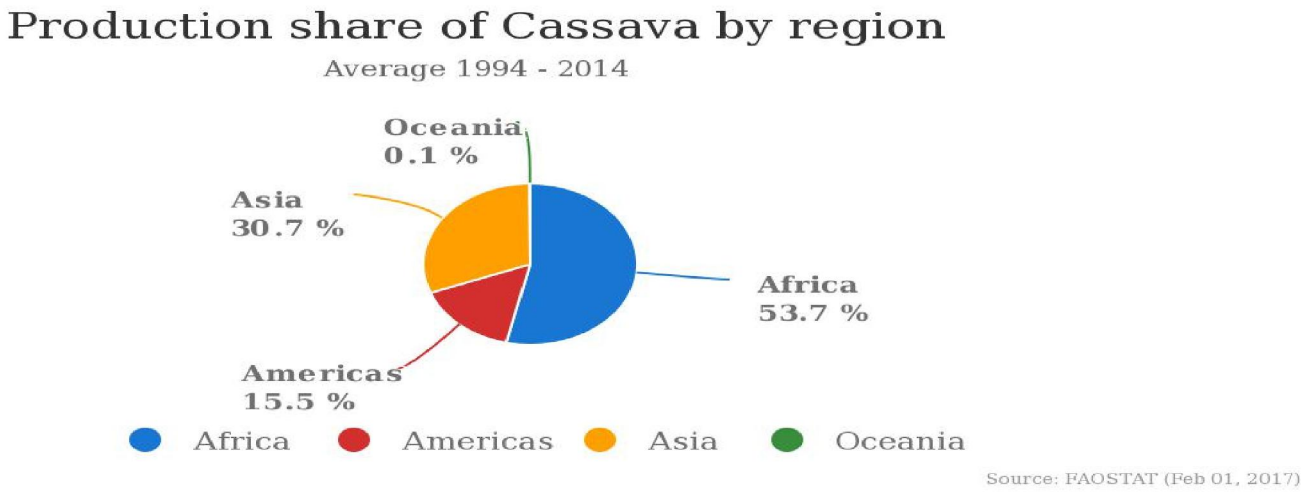


Figure 1.6: A pie chart indicating the production share of cassava by region (by percentage) for the time period 1994- 2014 (Source: FAOSTAT, 2017).

In Africa, cassava is one of the most critical staple food sources for low-income families in rural areas. Cassava is also a primary crop for smallholding farms. The constant expansion of the cassava industry in Africa is primarily driven by the increasing food demand for the crop on the continent (figure 1.7) (Morgan and Choct, 2016). Africa accounts for over 50% of global production of the crop amongst cassava- growing regions of the world (figure 1.6) (Alabi *et al*., 2011). Cassava has an astounding 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 biotic and abiotic stresses. 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).

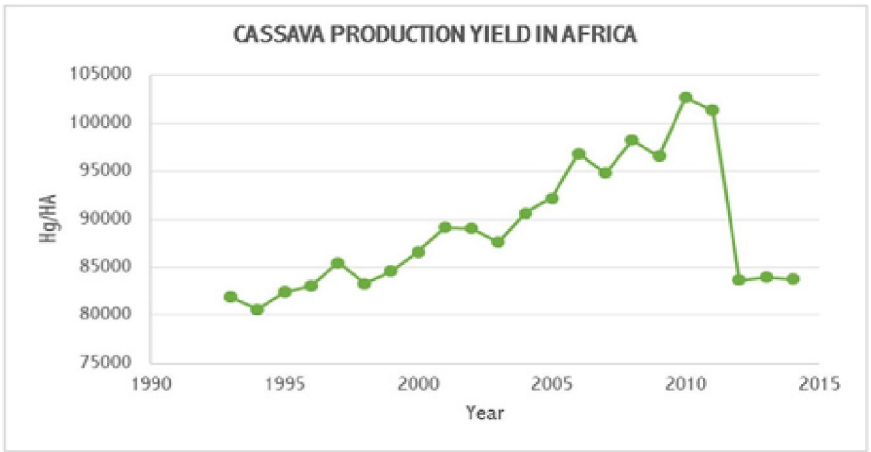


Figure 1.7: A trend line showing cassava yield in Africa over a period of 1994- 2014 (Source: FAOSTAT, 2017).

1.1.2. Cassava products

Cassava plants are cultivated as tuberous root crops. The roots are a key source of dietary starch and are consumed either as fresh or as a number of forms of processed foods. Cassava roots are composed of mostly carbohydrates as well as 1-3% of crude protein. Dried cassava roots are often produced into toasted cassava flour. Recently cassava dried 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. In some parts of Africa, Indonesia, Malaysia and Thailand cassava root chips and pellets are produced as common types of food stock for animal-feed and industry after they have been dried or mashed (Alabi *et al*., 2011; Morgan and Choct, 2016). Cassava is also increasingly becoming highly significant as a feedstock for ethanol fuel production predominantly in 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 (ranging between 16.6-39.9%), they also have high mineral levels and are a valuable source of vitamins B1, B2 and C as well as carotenes. The leaves are also often ground into meal and fed to poultry (Alabi *et al*., 2011; Morgan and Choct, 2016; Veiga *et al*., 2016).

1.1.3. Threats affecting cassava production

There are a number of threats to cassava production, the most important of these being pests and diseases. Most of the deleterious global impacts on cassava production have stemmed from the introduction of insect pests or disease causing pathogens to regions in which they previously did not occur. Some examples of this include: the introduction of arthropod pests such as the cassava mealy bug (*Phenacoccus manihoti* Mat-Ferr), cassava green mite (*Mononychellus tanajoa* (Bondar)) and cassava bacterial blight caused by *Xanthomonas axonopodus* pv. *manihotis*. A diverse set of virus species affect cassava especially in Latin America, the most economically important being the species which gives rise to cassava frogskin disease (CFSD). In Africa and Asia there are fewer virus groups which affect cassava. Cassava mosaic geminiviruses (CMGs) which causes cassava mosaic disease in Africa and Asia and Cassava brown streak viruses (CBSVs) which causes cassava brown streak disease in Africa have large and increasing 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 which infect cassava predominantly in Africa and its associated offshore islands (table 1.1). Of the 15 virus species listed, 11 of them are responsible for the disastrous diseases CMD and CBSD. *Cassava green mottle virus*, Cassava virus C and Cassava Kumi viruses A and B are known to infect cassava, however, they are not well characterized and their significance is still not fully known (Legg *et al*., 2015).

1.2. Cassava Mosaic Disease (CMD)

As depicted in table 1.1, it is apparent that cassava is vulnerable to a wide range of diseases caused by viruses. Cassava mosaic disease is one of these diseases. It is one of the most severe and widespread diseases which limits 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) (Alabi *et al*., 2011, Brown *et al*., 2015 Legg *et al*., 2015, Patil and Fauquet, 2015). 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 an important disease predominantly 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 a*l., 2001). The production of cassava tubers and starch-derived products from the tubers for human and animal consumption are hindered by the disease. The disastrous impact of the disease results in a lack of food security and income generation for farmers of cassava crops since few or no tubers are produced by infected plants depending on the severity of the disease inclusive of the age of the plant at the time of infection (Alabi *et al*., 2011).

Table 1.1: Table showing the viruses affecting cassava in Latin America, Africa and South Asia (Source: Legg *et al*., 2015).

Table 1 The viruses of cassava

Virus name	Genus/Family	References	Sequence(s) available[a]	Diagnostics[b]	Distribution
(a) Latin America					
Cassava common mosaic virus (CsCMV)	*Alphaflexiviridae/ Potexvirus*	Costa (1940), Silva, Kitajima, and Oliveira (1963), and Kitajima, Wetter, Oliveira, Silva, and Costa (1965)	NC_001658	ELISA/ RT-PCR	Colombia, Brazil (isolated cases from Africa, Asia)
Cassava vein mosaic virus (CsVMV)	*Caulimoviridae/ Cavemovirus*	Costa (1940), de Kochko et al. (1998)	NC_001648	PCR	Brazil
Cassava virus X (CsVX)	*Alphaflexiviridae/ Potexvirus*	Lennon, Aiton, and Harrison (1986)	NA[e]	ELISA/ RT-PCR	Colombia
Cassava new alphaflexivirus (CsNAV)	*Alphaflexiviridae/ Potexvirus*	Carvajal-Yepes et al. (2014)	KC505252	RT-PCR	Colombia
Cassava frogskin-associated virus (CsFSaV)	*Reoviridae/ Oryzavirus*	Calvert, Cuervo, Lozano, Villareal, and Arroyave (2008)	DQ139870	RT-PCR	Colombia, Brazil, Costa Rica, Argentina
Cassava polero-like virus (CsPLV)	*Luteoviridae/ Polerovirus*	Carvajal-Yepes et al. (2014)	KC505249	RT-PCR	Colombia, Costa Rica
Cassava torrado-like virus (CsTLV)	*Secoviridae/ Torradovirus*	Carvajal-Yepes et al. (2014)	KC505250, KC505151	RT-PCR	Colombia, Argentina
Cassava symptomless virus (CsSLV)	*Rhabdoviridae/ Nucleorhabdovirus*[c]	Kitajima and Costa (1979)	NA	NA	Brazil
Cassava Caribbean mosaic virus (CsCaMV)	*Alphaflexiviridae/ Potexvirus*[c]	Lennon et al. (1986)	NA	NA	Colombia
Cassava Colombian symptomless virus (CsCSLV)	*Alphaflexiviridae/ Potexvirus*[c]	Lennon et al. (1986)	NA	NA	Colombia
Cassava American latent virus (CsALV)	*Secoviridae/ Nepovirus*[c]	Walter, Ladeveze, Etienne, and Fuchs (1989)	NA	NA	Brazil, Guyana
(b) Africa					
Cassava mosaic disease					
African cassava mosaic virus (ACMV)	*Begomovirus/ Geminiviridae*	Morris, Coates, Lowe, Richardson, and Eddy (1990)	X17095, X17096	PCR and Real-time PCR	SSA[d]
African cassava mosaic Burkina Faso virus (ACMBFV)	*Begomovirus/ Geminiviridae*	Tiendrébéogo et al. (2012)	HE616777, HE616778	PCR and Real-time PCR	Burkina Faso
Cassava mosaic Madagascar virus (CMMGV)	*Begomovirus/ Geminiviridae*	Harimalala et al. (2012)	HE617299, HE617300	PCR and Real-time PCR	Madagascar
East African cassava mosaic Cameroon virus (EACMCV)	*Begomovirus/ Geminiviridae*	Fondong et al. (2000)	AF112354, AF112355	PCR and Real-time PCR	SSA and Comoros
East African cassava mosaic Kenya virus (EACMKV)	*Begomovirus/ Geminiviridae*	Bull et al. (2006)	AJ717580, AJ704965	PCR and Real-time PCR	East Africa, Madagascar, Seychelles, Comoros
East African cassava mosaic Malawi virus (EACMMV)	*Begomovirus/ Geminiviridae*	Zhou, Robinson, and Harrison (1998)	AJ006460, N/A	PCR and Real-time PCR	Malawi
East African cassava mosaic virus (EACMV)	*Begomovirus/ Geminiviridae*	Bull et al. (2006)	AJ717542, AJ704949	PCR and Real-time PCR	SSA
East African cassava mosaic virus-Ugandan Variant (EACMV-UG)	*Begomovirus/ Geminiviridae*	Pita et al. (2001)	AF126804-7	PCR and Real-time PCR	SSA
East African cassava mosaic Zanzibar virus (EACMZV)	*Begomovirus/ Geminiviridae*	Bull et al. (2006)	AJ717562, AJ704942	PCR and Real-time PCR	Zanzibar, Madagascar
South African cassava mosaic virus (SACMV)	*Begomovirus/ Geminiviridae*	Berrie, Rybicki, and Rey (2001)	AF155806, AF155807	PCR and Real-time PCR	South Africa, Madagascar, Zimbabwe
(c) South Asia and minor viruses					
Cassava mosaic disease					
Indian cassava mosaic virus (ICMV)	*Begomovirus/ Geminiviridae*	Malathi, Nair, and Shantha (1985), Hong, Robinson, and Harrison (1993)	NC_001932, NC_001933	PCR	Southern India and Sri Lanka
Sri Lankan cassava mosaic virus (SLCMV)	*Begomovirus/ Geminiviridae*	Saunders et al. (2002)	AJ314737, AJ314738	PCR	Southern India and Sri Lanka
Cassava viruses not linked with any major disease[c]					
Cassava virus C (CsVC) (syn. Cassava Q virus)	*Ourmiavirus/ Unassigned*	Calvert and Thresh (2002), Rastgou et al. (2009)	FJ157981-83	NA[e]	Ivory Coast
Cassava green mottle virus (CsGMV)	*Nepovirus/ Comoviridae*	Lennon, Aiton, and Harrison (1987)	NA	NA	Australasia and Pacific Islands, Solomon Islands
Cassava Ivorian bacilliform virus (CIBV)	*Anulavirus/ Bromoviridae*	Fargette, Roberts, and Harrison (1991), Scott, MacFarlane, McGavin, and Fargette (2014)	NA	NA	Ivory Coast
Cassava Kumi viruses A and B	Uncharacterized	Calvert and Thresh (2002)	NA	NA	Kumi district of Uganda

[a]GenBank accession numbers of reference isolates of viruses provided.
[b]Many methods are available. Most common and current method(s) indicated.
[c]Limited knowledge on disease biology and causal virus, reliable diagnostic tools are yet to be developed.
[d]SSA, sub-Saharan Africa.
[e]NA, not available.

1.2.1. Symptoms of CMD

Symptoms of CMD (figure 1.8, table 1.2) vary in type, extent and severity. CMD produces a diversity of foliar symptoms which include 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 1.8: 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 1.2: A table showing different symptoms caused by CMGs which cause CMD, the genus and family they belong to, the vector used to transmit the virus, and the areas it is predominantly distributed (Source: Alabi *et al*., 2011).

Virus	Genus/Family	Symptoms	Vector	Distribution*
African cassava mosaic virus (ACMV)	*Begomovirus/ Geminiviridae*	Mosaic, leaf distortion and stunting	Whitefly	Africa and India
East African cassava mosaic Cameroon virus (EACMCV)	*Begomovirus/ Geminiviridae*	Mosaic, leaf distortion and stunting	Whitefly	West Africa, Tanzania
East African cassava mosaic Kenya virus (EACMKV)	*Begomovirus/ Geminiviridae*	Mosaic, leaf distortion and stunting	Whitefly	East Africa
East African cassava mosaic Malawi virus (EACMMV)	*Begomovirus/ Geminiviridae*	Mosaic, leaf distortion and stunting	Whitefly	Malawi
East African cassava mosaic virus (EACMV)	*Begomovirus/ Geminiviridae*	Mosaic, leaf distortion and stunting	Whitefly	East Africa
East African cassava mosaic virus-Uganda (EACMV-UG)	*Begomovirus/ Geminiviridae*	Mosaic, leaf distortion and stunting	Whitefly	Sub-Saharan Africa
East African cassava mosaic Zanzibar virus (EACMZV)	*Begomovirus/ Geminiviridae*	Mosaic, leaf distortion and stunting	Whitefly	Zanzibar, Madagascar
Indian cassava mosaic virus (ICMV)	*Begomovirus/ Geminiviridae*	Mosaic, leaf distortion and stunting	Whitefly	India, Sri Lanka, Togo
South African cassava mosaic virus (SACMV)	*Begomovirus/ Geminiviridae*	Mosaic, leaf distortion and stunting	Whitefly	South Africa, Malawi, Madagascar, Zimbabwe
Sri Lankan cassava mosaic virus (SLCMV)	*Begomovirus/ Geminiviridae*	Mosaic, leaf distortion and stunting	Whitefly	India and Sri Lanka

1.3. Geminiviruses

Geminiviruses belong to the family *Geminiviridae* and are increasingly emerging as plant pathogens in a number of tropical and subtropical countries (table 1.3), eliciting a disastrous effect on productivity (Moffat, 1999; Boulton, 2003; Mansoor *et al*., 2003; Vanderschuren *et al*., 2007). Geminiviruses infect 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 increased presence of the viruliferous whitefly *Bemisia tabaci* which is the vector of geminiviruses (figure 1.9). 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).

Table 1.3: List of the economically important viral diseases caused by geminiviruses (Source: 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%, up to 100%
Cassava mosaic disease	*Begomovirus*	Bipartite	Cassava	Africa, India	Overall 15–24%, up to 90%
Cotton leaf curl disease	*Begomovirus*	Monopartite (associated β-component)	Cotton	Pakistan	Average 30%, up to 80%
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%, up to 100%

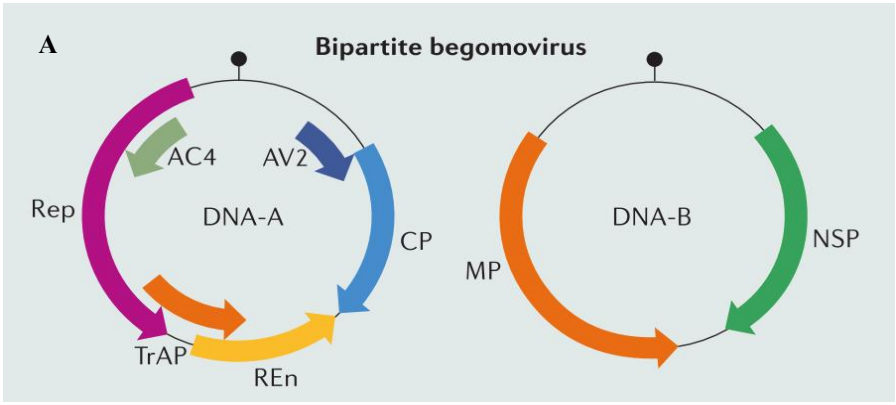
Figure 1.9: The whitefly which transmits geminiviruses (Source: Inoue-Nagata *et al*., 2016).

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 newly added *Capulavirus* and *Grablovirus* (Brown *et al*., 2012; Varsani *et al*., 2017). A large number of geminviruses have bipartite genomes and the two segments

are similar in size. Viruses which belong to the Begomovirus genus, such as SACMV , are transmitted by the whitefly vector *Bemisia tabaci* (Seal *et al*., 2006). 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).

1.3.1. Structure of geminiviruses

All bipartite begomoviruses have two circular DNA molecules, DNA A (2800 nucleotides) and DNA B(2760 nucleotides) (figure 1.10(A)); 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 1.10(B)). 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).



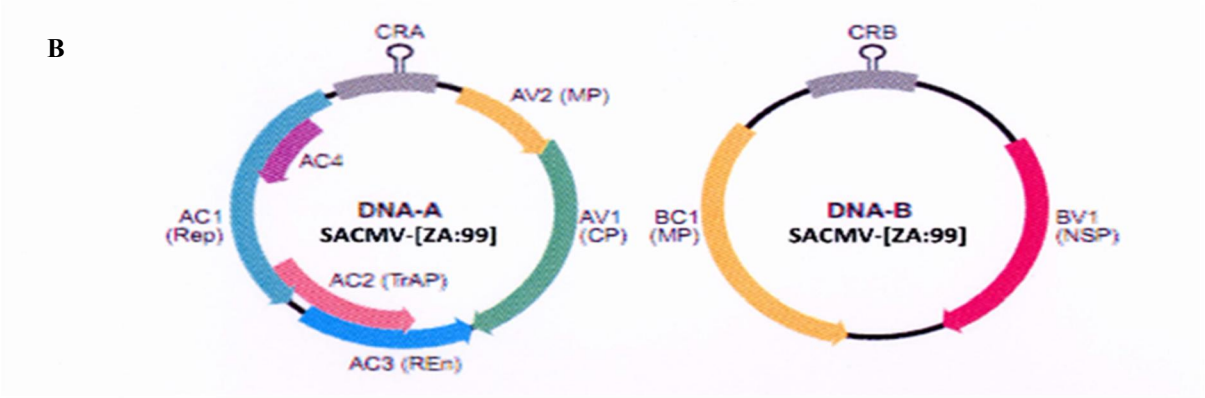


Figure 1.10:(A) A schematic diagram of a bipartite begomovirus, showing both DNA-A and DNA-B and the relative sizes and proteins which they encode for (Source: Hanley-Bowdoin *et al.*, 2013) and **(B)** a detailed illustration of the SACMV bipartite genome, illustrating approximate position of all proteins present (Source: Berrie *et al*., 1998).

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 trnascriptional activator and host RNA silencing surpressor 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).

1.3.2. 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 1.10) 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 for 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 1.11) 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).

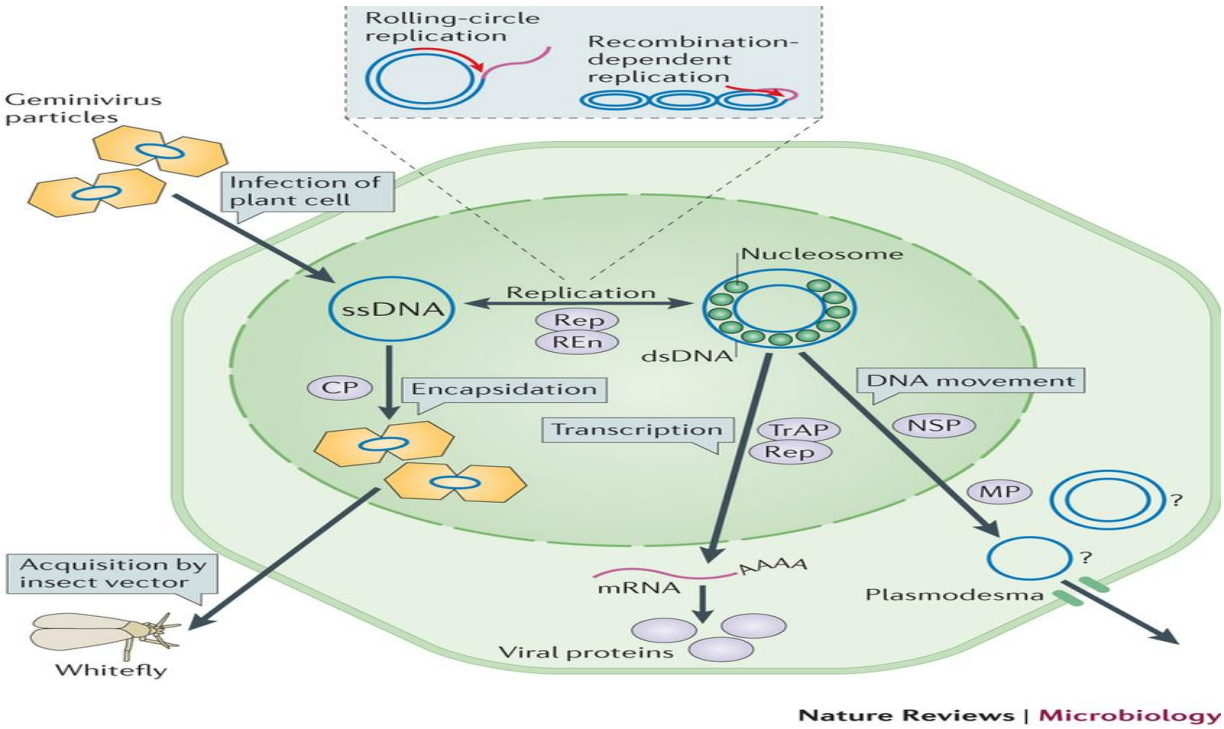


Figure 1.11: The replication of begomoviruses as it begins in a plant cell (host cell) (Source: Hanley-Bowdoin *et al*., 2013).

1.4 Cell-to-cell movement of viruses and BC1 movement protein

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 sucessful 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). 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 accesory 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. BDMV also a bipartite geminivirus has been investigated and can give clues as to how SACMV may behave. Given that all geminiviruses, including SACMV, require functional BC1 and BV1 proteins 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 BC1 protein 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 BC1 and BV1 of SACMV facilitate similar movements and functions (Sudarshana *et al*., 1998).

The movement protein BC1 gene of SACMV is 923 nucleotides and encodes for 307 amino acids (figure 1.12). The various MP domains, their respective functions as well as the predicted secondary, tertiary and molecular structures of geminivirus MPs have not to date been determined. 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.

```
        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 1.12: Amino acid sequence of SACMV BC1 (Source: Genbank: AAF34900.2; Berrie *et al*., 2001).

1.5. Virus-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 (Whitham and Wang, 2004; Jeske, 2009; Gutierrez, 2002). In addition, geminivirus proteins have been implicated in various host-responsive processes such as that of 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 *Bean dwarf mosaic virus* (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.

1.6. Protein-protein interactions

Although virus induced gene silencing (VIGS) may be used to identify the role of candidate host genes, it is not possible to identify whether the viral proteins interact directly with the host proteins or whether transcript alterations are indirect effects from virus infection. To elucidate virus-host protein-protein interactions other approaches must be investigated. In order for proteins to successfully fulfil their role, they must interact with other molecules. This as proteins usually function in complexes. In order to understand a specific proteins function it is important to know which other proteins the specific protein physically and functionally interacts with (Braun *et al*., 2013). Recently protein interactome studies have become a popular method by which to map protein-protein interactions (Braun *et al*., 2013; Ohad *et al*., 2007). Methods commonly used in studies for

protein-protein interaction mapping are: the yeast two-hybrid system (Y2H); bimolecular fluorescence complementation (BiFC) and affinity purification immuno-coprecipitation coupled with mass spectrometry (AP-MS) (Zhang *et al*., 2010; Ohad *et al*., 2007).

The Y2H exploits the system for a GAL4 transcriptional factor protein that binds with a UAS promoter and activates down-stream expression. The GAL protein contains a transcriptional activation domain and a DNA-binding domain. When the bait and prey proteins fuse with the domains of the GAL protein, these domains produce a functional GAL4 transcription factor and activates the UAS-driven expression reporter genes (Zhang *et al*., 2010; Wilson and Walker, 2012). The Y2H has been widely used to identify and map protein interactions and networks particularly in the model plant *Arabidopsis thaliana* (Braun *et al*., 2013; Zhang *et al*., 2010).

BiFC is considered an efficient tool to evaluate protein-protein interactions as they can determine subcellular localization of protein complexes. BiFC is based upon a split between yellow fluorescent protein (YFP), cyan fluorescent protein (CFP) or other green fluorescent protein (GFP) variants to form a functional fluorescent fluorophore. The bait and prey proteins are fused and result in the fusion of two combinatory parts of a fluorescent protein, observed under fluorescent microscopy (Zhang *et al*., 2010; Ohad *et al*., 2007). Similar to the networks mapped using the Y2H approach, the BiFC has been used to identify and map protein interactions and networks in the model plant *Arabidopsis thaliana* (Zhang *et al*., 2010; Ohad *et al*., 2007).

AP-MS uses both affinity purification and mass spectrometry. The bait protein is fused to an affinity tag such as a his-tag for *in vivo* expression. The multi-component protein complex is isolated from the cell lysate by affinity purification and analysed by mass spectrometry (Zhang *et al*., 2010; Braun *et al*., 2013).

1.7. Characterization of proteins

The primary structure of a protein consists of the amino acid residues linked by peptide bonds (Lesk, 2010; Wilson and Walker, 2012). The secondary structure of a protein describes the localised folding of a polypeptide chain due to hydrogen bonding. The secondary structure includes structures such as α-helices, β-strands and disordered regions (Lesk, 2010; Wilson and Walker, 2012). The secondary structure of a protein can be determined using a spectroscopy technique called Circular Dichroism (CD) spectroscopy. CD permits the elucidation of secondary structural features of proteins. Based on

the chirality of amino acids, when exposed to plane polarised light, left and right handed circularly polarised components are absorbed to different extents (Kelly *et al*., 2005). Proteins contain a number of chromophores which can give rise to CD signals in the far UV region (240-180 nm) and this corresponds to peptide bond absorption (Kelly and Price, 2000).

Fluorescence spectroscopy is often used for protein analyses. Extrinsic fluorescent dyes such as 1-anilino-8-naphthalene sulphonate (ANS) can be covalently attached to proteins and therefore provide information on protein characterization (Hawe *et al*., 2008). ANS is a non-conjugating chromophore and emits a weak fluorescence in polar environments. However, in non-polar environments the fluorescence emission significantly increases, making ANS an invaluable tool for the assessment of the degree of non-polarity. ANS is used to monitor the binding of ligands and prosthetic groups to proteins (Wilson and Walker, 2012).

Proteins possess three intrinsic fluorophores: tyrosine, phenylalanine and tryptophan. Phenylalanine yields low quantum and is therefore, considered negligible in its contribution to protein fluorescence emission. Tyrosine possesses lower quantum yields than tryptophan and is highly quenched upon ionisation or when in close proximity to an amino group, carboxyl group or tryptophan residue, this results in the intrinsic protein fluorescence being determined by tryptophan fluorescence where the protein is excited at between 295-303 nm (Hawe *et al*., 2008; Wilson and Walker, 2012).

1.8. Rationale for this study

Cassava is an important food crop in Asia and Sub-Saharan Africa including the southern African region. While providing nutrition to humans and animals, it also provides a source of income for farmers of these regions (Alabi *et al*., 2011). The loss of cassava yields due to virus diseases causes devastating losses for a number of these farmers. 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 is a movement protein and aids cell-to-cell movement of the virus. It is therefore vital to fully understand the functionality of the protein and its interactions with host proteins in the cassava host in order to develop strategies to limit or eliminate SACMV movement once the virus has gained entry into the plant. Since the BC1 protein of SACMV is a novel protein (as it has never been studied before) the protein structure needs to be elucidated. This will be done by structurally characterizing the protein and characterization will include essential parameters such as confirming the molecular weight, secondary structure and

tertiary structure of the protein. Potential selected host proteins that may interact with MP (HSP70 and Histone H3) will be investigated *in silico* using computational databases.

1.9. Specific aims

The specific aims of the study are:

1. *To predict possible functional domains of the movement protein (BC1) of SACMV in silico*

   Since not much is known about BC1, except the amino acid sequences, it is useful to study this protein using a computational approach as this provides theoretical referrals for downstream characterization experiments. This will be done by using a number of internet databases such as 'Predict Protein', 'Encorbio', 'Prot-param' and 'Isoelectric'.

2. *To clone BC1 into a suitable expression vector and to express, purify and refold the protein*

   In order to use the movement protein in downstream applications the gene must be cloned into a suitable expression vector. Once the gene is successfully cloned it will be expressed large-scale and then further purified (after denaturing) and refolded.

3. *To structurally characterize the movement protein*

   Structural characterization will include determining the secondary structure of the protein using CD, determining the molecular weight of the protein using SDS-page gels and determining the extent if at all the protein binds ATP using ANS and MANT-ATP essays to mention the minimum.

4. *Preliminary studies of predictive interactions between cell-to-cell movement protein BC1 of SACMV and host proteins from cassava will be carried out using literature and in silico studies*

   Preliminary studies of predictive interactions between the cell-to-cell movement protein BC1 of SACMV and host proteins from cassava will be carried out using literature and *in silico* studies.

Chapter Two
Understanding movement protein BC1- An *in silico* approach

2.1. Introduction

In order to fully understand any protein's function, and the mechanisms which it employs to serve its function it is essential to understand the biological properties of the protein. Biological properties include hydrophobicity of the protein, the number of domains, ability to bind to biomolecules, molecular weight, and secondary structure conformation (Radivojac *et al*., 2013).

South African cassava mosaic virus (SACMV) belongs to the genus *Begomovirus* which comprises virus genomes of circular ss DNA A and DNA B (bipartite). The DNA A molecule ORF encodes three proteins in the antisense direction namely, the Rep-associated (AC1), replication enhancer Ren (AC2) and transcriptional activator TrAP (AC3) proteins (Patil and Fauquet, 2015). Additionally, an overlapping AC4 ORF encodes a putative host RNA silencing suppressor protein. DNA-B BC1 ORF encodes the movement protein (MP) from the complementary strand whereas 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). While some research (*in silico* and experimental) has been done on proteins encoded by DNA A, very little research has been done on the proteins encoded by DNA B, especially on the MP (Prajapat *et al*., 2011; Hanley-Bowdoin *et al*., 2013).

In recent years there has been a growing availability of computational genomic-level sequence information databases for thousands of virus and other species. Along with high-throughput computational experimental data, and available protein structure and function databases, new opportunities for computational *in silico* protein function prediction have opened up. Several methods have been recommended to manipulate 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).

In silico approaches are used to find predictive models and to provide virtual screening of biological properties of proteins and other compounds (Estrada and Peña, 2000). *In silico* methods are often

favoured in protein modelling as it is often the only way to obtain structural information about novel proteins especially if there is no experimental data is available (Prajapat *et al*., 2010).

Protein sequence databases 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 translated by nucleic acid sequences (Lesk, 2010). These databases are employed by web programs, services and software such as 'Predict Protein' (https://www.predictprotein.org/) and 'IntFOLD' which are webservers which creates alignments and predicts aspects of structure and function based of the annotations listed within the databases (Roche *et al*., 2011; Rost and Liu, 2003). Homolgy modelling programmes such as 'PyMOL' (https://pymol.org/2/) and 'Verify 3D' (https://omictools.com/verify3d-tool) also employ protein databases to generate 3D models of proteins for molecular visualization (Prajapat *et al*., 2010). *In silico* studies have been done on other geminivirus proteins such as Rep of Tomato yellow leaf curl virus (TYLCV), where the study, focussed on homology modelling and docking of Rep with other proteins using '3D JIGSAW' and UCSF Chimera software (Prajapat *et al*., 2011).

Since very little research has been done to date on geminivirus proteins, in particular the MP, it is useful to study the protein *in silico* before attempting experimental approaches to elucidate protein structure. This approach provides a frame of reference when characterizing the protein experimentally.

This chapter is specifically designed to report on the *in silico* study of the movement protein (BC1) from SACMV, a bipartite *begomovirus* from the family *Geminiviruses*. An *in silico* assessment was crucial to attain as BC1 is a novel protein with mostly no comparable information available.

2.2. Materials and Methods

2.2.1. Phylogenetic analysis of geminivirus movement proteins

In order to construct a maximum likelihood phylogenetic tree to compare protein amino acid sequence identities between geminivirus MPs, the amino acid sequences (FASTA format) were mined from Genbank- NCBI database. The amino acid sequence of South African cassava mosaic virus BC1 was retrieved and was used as an input source for most databases (Source: Berrie *et al*., 2001; Genbank: AAF34900.2).

Template sequence for queries: SACMV BC1 Genbank- NCBI **AAF34900.2** (Source Berrie *et al*., 2001).

MDAQFTVTDNNYINSKRTEYALTNDAAPINLQFPGSFEQATMRLKGRCMKIDHIIIEYRNQV
PFNATGSVIVEIRDNRVSLDDAAQAAFTFPIACNVDLHYFSSTYFSISEPSPWRIMYRVEDSN
VIEGVKFASIKAKLRLSSAKHSTDIRFKPPTINILSKGYTKDCIDFWSVEKGETRRRLLNPTPT
ARSQQPITHRPITIHPGETWATRSQIGLPSSLGPAQLEHFRSQSMRMDPSTTPTDLDNDSTEYP
YQKLHRLHTPDLDPGDSISQAQSDSVSRKDLETLLESTINKCLYKIK SDAPRQL

Two hundred and eleven full length geminivirus BC1 amino acid sequences were retrieved from the NCBI database. The FASTA format of the amino acid sequences were deposited into Molecular Evolutionary Genetics Analysis version 7 (MEGA7), a computer software programme which conducts statistical analysis of molecular evolution and constructs phylogenetic trees using a number of different methods (selected against similar programmes based on its simplicity and user friendliness). Any amino acid sequence repeats were removed. This resulted in 15 amino acid sequences remaining which were trimmed and then aligned by using the 'MUSCLE' function and computed. The alignment was then exported in FASTA format. In MEGA7 (https://www.megasoftware.net/) a maximum likelihood tree phylogeny was performed using the ML Heuristic Method- Nearest-Neighbour-Interchange and 1000 bootstrap replicates, Whelan and Goldman model.

2.2.2. Determining matches to non-geminivirus families of proteins

'Inter-ProScan' (https://www.ebi.ac.uk/interpro/) functionally analyses proteins by classifying them into families and predicts 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 were as well as family membership indicated.

2.2.3. Predictive molecular weight, net charge and pI of SACMV BC1

In order to retrieve a predictive molecular weight for the BC1 'Encorbio' (https://www.encorbio.com/protocols/Prot-MW.htm) was selected because the molecular weight is determined as well as the net charge. The atomic weights for each atom used is checked against the International Union of Pure and Applied Chemistry (IUPAC). This database required a FASTA input of the amino acid sequence and subsequently generated a theoretical molecular weight and also generated a net charge for BC1. In order to determine a theoretical pI for BC1, the web service employed was 'Isoelectric' (http://isoelectric.org/) which also required a FASTA input of the amino acid sequence in order to generate the result. This web service estimates the pI of proteins using different sets of dissociation constant (pKa) values (Kozlowski, 2016).

2.2.4. Overall protein characterization prediction analysis

In order to elucidate the *in silico* protein structure of BC1, a web service called 'Predict Protein' (https://www.predictprotein.org/) was employed. This web service used the FASTA input of the BC1 primary amino acid sequence and generated a multitude of theoretical predictions apart from the predictive secondary structure. 'Predict Protein' determined the predicted solvent accessibility (tertiary structure in terms of hydrophobic and hydrophilic regions), protein disorder and flexibility, disordered regions, disulfide bridges as well as protein and nucleotide binding regions.

2.2.5. Building the secondary structure model of BC1

In an attempt to create a predictive model of the secondary structure of BC1, 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 the web 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 protein was entered into the database and the predicted secondary structure was modelled on the web portal using the nearest PDB family header.

2.3. Results and Discussion

2.3.1. Maximum likelihood phylogenetic tree of geminivirus MPs

Phylogenetic studies on the relatedness between geminiviral MPs can also provide clues on functional and structural proteins. The MP BC1 from SACMV was shown to share the highest maximum likelihood of 100% with movement protein BC1 from EACMV based upon the WAG method or the Whelan and Goldman method. The WAG method was used in this study as it is the recommended method for determining the maximum likelihood in phylogenetic studies. This, as it is the most recent method and incorporates similarities found in common with the statistically sound methods of the Jones, Taylor and Thornton Method (JTT) and the Dayhoff methods (Whelan and Goldman, 2001). The method also gives significantly higher likelihoods than any other methods used when being used to assess phylogenies for protein sequences (Whelan and Goldman, 2001). The 100% amino acid sequence identity between SACMV and EACMV (figure 2.1) is expected as previous phylogenetic (nucleic acid sequence identities) anaylses indicate clustering between SACMV DNA B and EACMV DNA B with a 90% similarity. This indicates that SACMV is a strain of EACMV (Berrie *et al*., 2001). Interestingly the cell to cell movement protein did not share a high amino acid sequence identity with African cassava mosaic virus (ACMV). Since EACMV has been suggested to be a recombinant of ACMV (Berrie *et al*., 1998). The movement protein from EACMV has also not to date been studied. Since no information is available about the EACMV or any other geminivirus movement protein it's structural traits cannot be inferred upon the movement protein of SACMV.

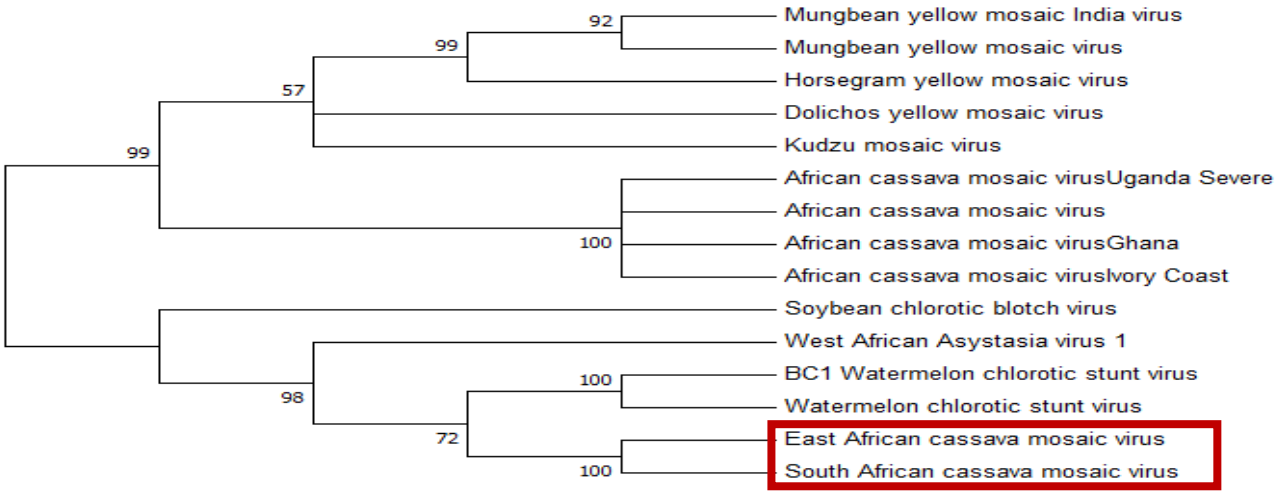


Figure 2.1: A molecular phylogenetic analysis of BC1 movement proteins or similar of geminiviruses using a maximum likelihood method based on the Whelan and Goldman model. (Source: MEGA7, 2017).

2.3.2. Determining matches to non-geminivirus families of proteins

The database Inter-ProScan (figure 2.2.) showed that BC1 of SACMV did not share any amino acid similarity with other non-geminiviral proteins, only that of other geminiviruses. The database also indicated that the protein showed no signs of any possible domains that were shared with any other protein families.

Inter-ProScan was used to examine if BC1 movement proteins share any similarity with any other protein family members which may share similar predicted functions, such as cell-to-cell and long distance movement of viruses (Fondong, 2013), since BC1 has been suggested to move DNA from the nuclear envelope across the plasmodesmata to the next cell (Hanley-Bowdoin *et al*., 2013). However, this *in silico* study did not reveal any sequence matches between SACMV BC1 and other proteins retrieved from 'Inter-ProScan'. No information was available that showed associated protein domains or domain repeats between BC1 and other proteins in the database. Based on a review by Fondong (2013) it is predicted that geminivirus BC1 encoded movement proteins contain a pilot domain located towards the N-terminal domain (required for localization to the cell periphery), a central anchor domain (targets the protein to the cell periphery) and a C-terminal domain or oligomerization domain (facilitates oligomerization).

Figure 2.2: An image taken off the 'Inter-ProScan' site showing the protein family membership and detailed sequence matches of BC1 amongst other non-geminiviral protein families. The image also shows no domains or domain repeats within BC1 (Source: Inter-ProScan, 2015).

2.3.3. Predictive molecular weight, net charge and pI of SACMV BC1

Determining the functional pH or pI (pH at which a protein has no net charge) is essential in protein studies as every biological process is pH dependent (Talley and Emil, 2011; Wilson and Walker, 2012). The isoelectric analysis (figure 2.3) indicated multiple predicted pI values for movement protein BC1 of SACMV. This as the site uses a number of algorithms and sources to calculate the pI using different sets of dissociation constant (pKa) values (Kozlowski, 2016) . While most of the values tend to lean in the same pI region of around 7, the Patrickios predictions seem to deviate at 4.466 (based on the algorithim used to calculate the pI).

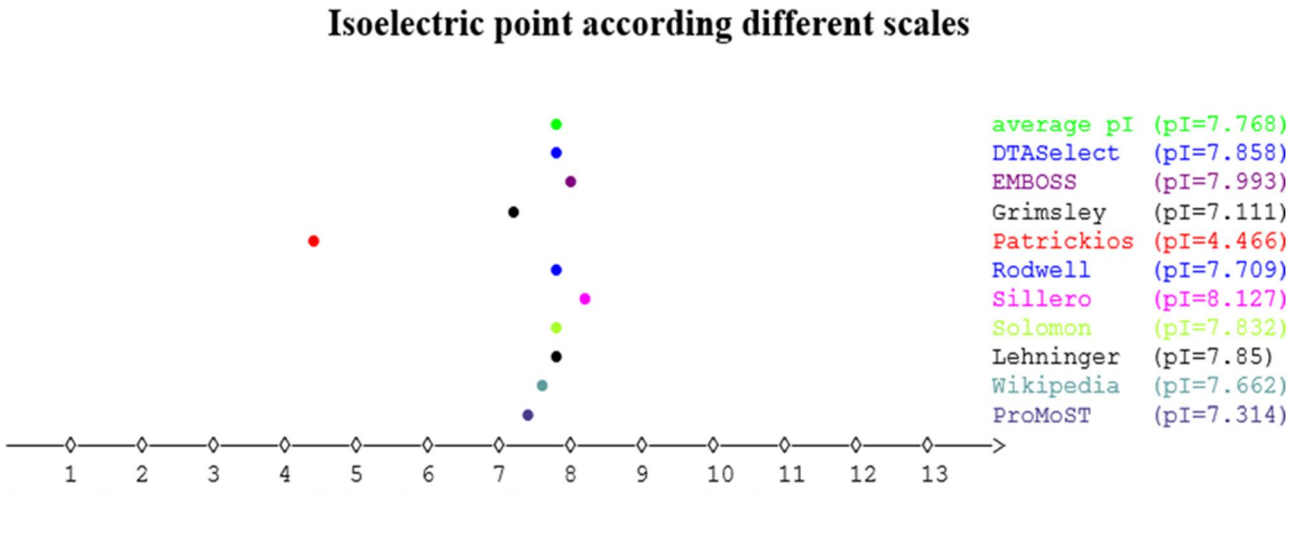


Figure 2.3: 'Isoelectric' image showing the predicted average pI of SACMV BC1 using a number of different scales (as per different alogrithims and sources available to the server) (Source: Isoelectric, 2015).

2.3.4. Overall protein characterization prediction analysis

The secondary structure of a protein is defined by the localised folding of the polypeptide chain due to hydrogen bonding. Known structures include α-helices and ß-pleated sheets (Wilson and Walker, 2012). Determining the secondary structure of a protein is one of the most crucial steps in structurally characterizing a protein. In this study the webserver 'Predict protein' was employed to computationally predict the hypothetical secondary structure of SACMV movement protein BC1 which predicted that the secondary structure of the protein should be mostly disordered (loop). The predictive BC1 movement protein secondary structure consists of a predominantly disordered conformation, and also ß-sheets and α-helices to a lesser extent (figure 2.4. A).

In figure 2.4. B the predicted solvent accessibility of the protein is depicted and this shows what extent of the protein seems to be hydrophobic (38.11%) as well as the extent of the protein which seems to be hydrophilic (51.79%). The figure predicts that the extent of the protein which may be in an intermediate stage is 10.10%. 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 would allow 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 predicted intermediate region on BC1 could account for a region of the protein which may be amphipathic (both hydrophobic and hydrophilic), indicating that this region could function as hydrophilic or hydrophobic based on ligand binding and has tolerance to both polar and non-polar ligands (Campbell and Farrell, 2009).

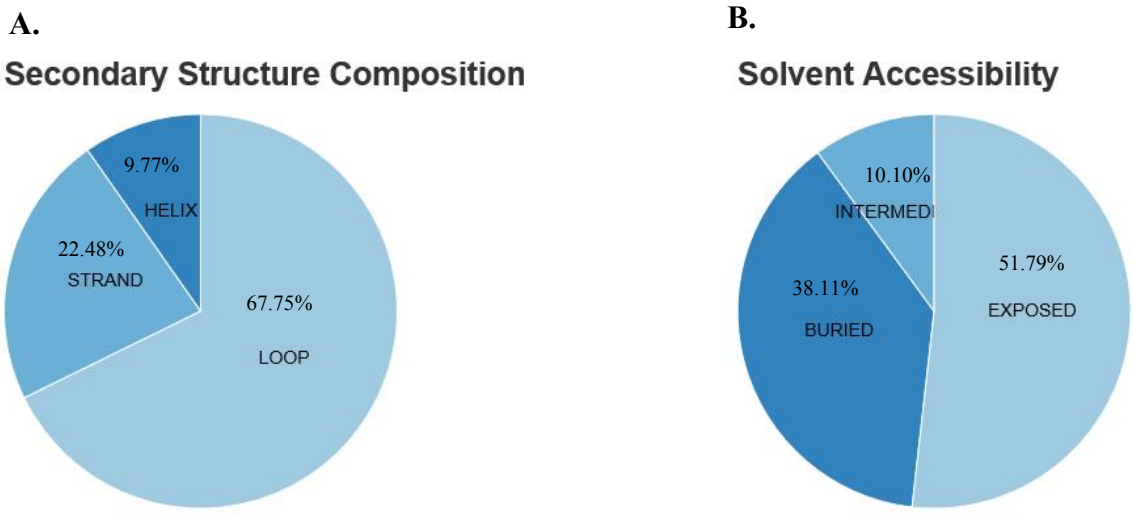


Figure 2.4: 'Predict Protein' images showing the predicted **(A)** secondary structure and **(B)** predicted secondary structure solvent accessibility (exposure of hydrophobic regions to solvent) of SACMV movement protein BC1 (Source: Predict protein, 2016).

The predicted disordered regions of a protein are indicative of the regions of the protein that could 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. This allows the disordered segment to perform relatively unhindered conformational searches when binding to their target molecules (Voet *et al*., 2013).. Four different prediction methods were used in

order to determine the disordered regions by the PredictProtein webserver. Each of these prediction methods generated a different result, each indicated by a different colour. The yellow bars depicted in figure 2.5 indicate the disordered and flexible protein regions as predicted by the PROFbval software programme which identifies flexible regions of a protein based on the lack of a 3D structure of the protein in that specific region (Schlessinger and Rost, 2005). The grey bars depicted in figure 2.5 indicate the disordered and flexible regions as predicted by the NORSnet software programme, this software specializes in distinguishing between well-folded proteins and very long contiguous segments with non-regular secondary structures making this software more sensitive to determining disordered and flexible regions than PROFbval and Ucon software programmes (Schlessinger *et al*., 2007). The olive green bars depicted in figure 2.5 indicate the disordered and flexible regions of the protein as predicted by the DISOPRED2 (Disorder Prediction server 2) server which checks the protein sequence against around 750 non-redundant sequences with high resolution X-ray structures, this could result in a false result of a disordered region and would likely be the least reliable prediction (Ward *et al*., 2004). However, upon comparison amongst all four prediction methods (NORSnet software being the most reliable) shown in figure 2.5 it seems that there is consensus that the most disorder and flexibility occurs at the C-terminus. This indicates that the predicted oligomerization domain postulated by Fondong, 2013 could be lacking a fixed 3D structure.

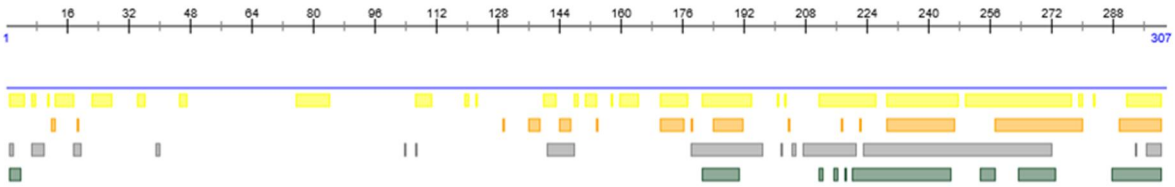


Figure 2.5: 'Predict Protein' image showing the predicted protein disorder for the protein BC1 (Source: Predict Protein, 2016). The different colours (colour bars represents disorder and flexibility, lack of colour represents protein folded regions) represent the different prediction methods used to predict the protein disorder regions: yellow, orange, grey and olive green represents the PROFbval, the Ucon, the NORSnet and the DISOPRED2 prediction methods, respectively.

A protein disulfide bond is a covalent link between the sulfur atoms of the thiol groups (–SH) in two cysteine residues. Disulfide bridges are covalent bonds between sulfur atoms of the thiol groups in cysteine residues only. Disulfide bridges play a crucial role in stabilizing the folding process of protein folding (Voet *et al*., 2013). The results in figure 2.6 indicate that we predict BC1 to be rather stable during protein folding since it has a disulfide bridge.

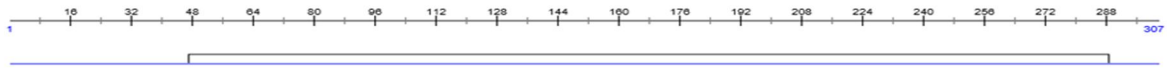


Figure 2.6: 'Predict Protein' showing the predicted disulfide bridge for SACMV BC1.

In order for a protein to function successfully the protein must interact with other biological molecules such as nucleotides (DNA and/or RNA) and proteins and in order for them to interact with these molecules they must be able to bind to them. The predicted binding sites on SACMV BC1 displayed by the 'Predict protein' database indicate a number of protein binding sites spanning across the protein, however, there appears to be more protein binding sites towards the C-terminal end of the protein. This correlates with literature by Voet *et al*., (2013) where they propose that in regions where proteins are disordered, those regions are more flexible and are able to bind to targets compared to ordered regions. From figure 2.7, it is clear that the BC1 protein has many predictive protein binding sites, predominantly at the C-terminus, while only two predictive nucleotide binding sites occur between amino acid positions 128 and 144. It is expected that BC1 would have many predictive protein binding sites at the C terminus as this is where the protein has the most disorder (figure 2.5). Geminivirus MPs have been known to bind DNA with high affinity since they form a complex with the DNA and shuttle the DNA across the plasmodesmata (Fondong, 2013; Rojas *et al*., 1998). This explains the DNA binding sites between amino acid positions 128 and 144 where it can be proposed that BC1 will bind DNA.

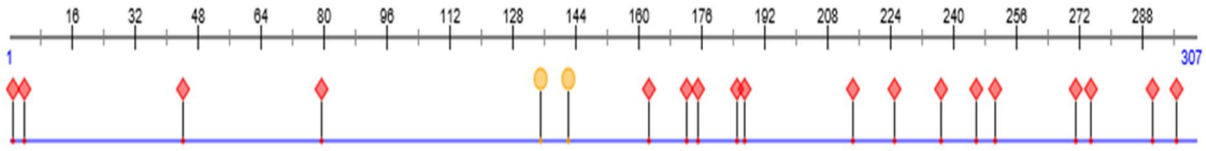


Figure 2.7: 'Predict Protein' image showing the predicted SACMV BC1 protein binding sites (indicated by the red diamonds) and nucleotide binding sites (indicated by yellow circles) (Source: Predict Protein, 2016).

2.3.5. Building the secondary structure model of BC1

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). Figure 2.8

shows that only 51% of the protein was able to be modelled by Phyre 2. This is not sufficient for a reliable model. The ability of Phyre 2 to only model 51% of the protein correlates with the results attained in figure 2.2 where the Inter-ProScan server was unable to establish any sequence or domain similarity with any other protein families. Figure 2.8 illustrates that research on BC1 is still novel as the PDB was not able to establish any sequence or structure similarity with another protein.

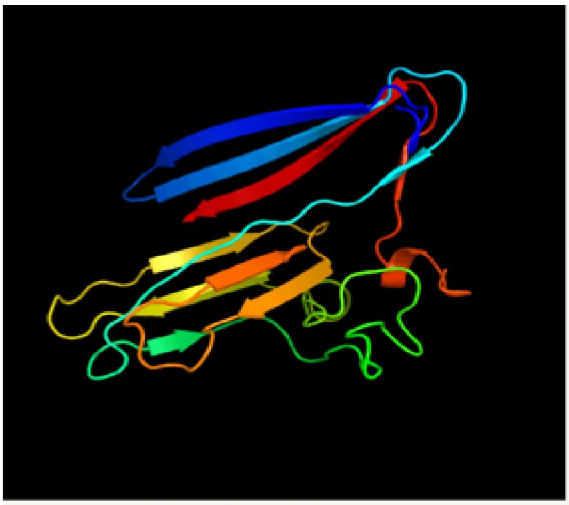


Image coloured by rainbow N → C terminus

Model dimensions (Å): **X**:51.936 **Y**:46.251 **Z**:40.627

PDB header:isomerase
Chain: A: **PDB Molecule:**aldose 1-epimerase;
PDBTitle: crystal structure of a putative galactose mutarotase (bsu18360) from2 bacillus subtilis at 1.45 a resolution

Confidence and coverage

Confidence: 65.9% Coverage: 51%

157 residues (51% of your sequence) have been modelled with 65.9% confidence by the single highest scoring template.

Phyre alarm

You may wish to submit your sequence to Phyrealarm. This will automatically scan your sequence every week for new potential templates as they appear in the Phyre2 library.

Please note: You must be registered and logged in to use Phyrealarm.

Figure 2.8 An image, taken from 'Phyre 2' web portal for protein modelling, prediction and analysis, showing the predicted homology model of SACMV movement protein BC1 where 51% of the sequence was modelled with 65.9% confidence.

2.4. Conclusions

Since little information is available on the structure of geminivirus cell-to-cell movement proteins encoded by BC1 ORF on DNA B, predictive *in silico* characterization, using a wealth of internet programs, web portals and databases (Roche *et al*., 2011; Prajapat *et al*., 2010; Prajapat *et al*., 2011; Rost and Liu, 2003), can be useful in accompanying in vitro experimental analyses. The use of a computational/ *in silico* approach to predict novel protein structure or function can provide some useful insight as to what to expect when structurally characterizing the protein experimentally *in vitro*, using a number of publically available databases or programs such as 'Predict protein' (Roche *et al*., 2011; Rost and Liu, 2003) , 'Encorbio' and 'Phyre 2' (Kelley *et al*., 2009).While the results obtained during in silico studies may occasionally be disputed most results tend to be validated *in vitro* and *in vivo*. A computational approach is a useful tool when studying novel proteins.

Chapter 3
Purification and characterization of the cell-to-cell movement protein of *South African cassava mosaic virus*

3.1. Introduction

South African cassava mosaic virus (SACMV), belonging to the family *Geminiviridae* and genus *Begomovirus* whose members infect a wide range of plant hosts (Berrie *et al*., 2001; Rey *et al*., 2012), is composed of a bipartite genome (single stranded circular DNA A and DNA B components) (Berrie *et al*., 2001) . DNA A ORFs of bipartite begomoviruses encodes three proteins in the antisense direction namely, the Rep-associated (AC1), replication enhancer Ren (AC2) and transcriptional activator TrAP (AC3) proteins (Patil and Fauquet, 2015). Additionally, an overlapping AC4 ORF encodes putative host RNA silencing suppressor protein. Conversely, in the sense direction the structural coat protein (CP) is encoded by AV1 (Patil and Fauquet, 2015), and AV2, a reported RNA silencing suppressor (Vanitharani *et al*., 2004). DNA-B BC1 ORF encodes for the movement protein (MP) from the complementary strand whereas the nuclear shuttle protein (NSP) is encoded from the sense strand by the BV1 ORF (Patil and Fauquet, 2015).

In plants viruses have to travel through the cell wall via plasmodesmata (Rojas *et al*., 1998), and the MP is required for cell-to-cell and long-distance movement through its interaction the NSP (Ward *et al*., 1997). Geminiviruses also utilize host membranes such as the endoplasmic reticulum (ER) to traffic from the nucleus to the plasmodesmata (Zhang *et al*., 2002). There are two models suggested for the active role of geminivirus MP and NSP during cell-to-cell transport. The first or 'Couple-Skating' model suggests that the MP binds to the ssDNA/dsDNA-NSP complex on the cytoplasmic side of plasma membranes or microsomal vesicles and subsequently transfers the nucleoprotein complex into the next cell along the plasma membrane or through the endoplasmic reticulum (Krentz *et al*., 2012; Fontes *et al*., 2004). The second or 'Relay-race' model suggests after NSP-mediated nuclear export, viral ssDNA or dsDNA is taken by the MP, which furthers transports the viral DNA into an adjacent cell, likely with the assistance of other accessory proteins (Krentz *et al*., 2012; Fontes *et al*., 2004).

The deduced amino acid sequence of SACMV DNA B (Berrie *et al*., 2001) and functional studies have informed a putative model of the BC1 protein structural domains (Fondong, 2013), but little is currently known on its secondary and tertiary structure. The MP contains a putative N-terminus domain required for localization to the cell periphery, a central anchor that also targets proteins to cell membranes, and a C-terminal oligomerization domain. The nuclear shuttle protein binding site is predicted to locate in the central region of the BC1 protein (Fondong, 2013). The primary structure of a protein consists of the amino acid sequence 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).

Affinity chromatography is a preferred method for purification of recombinant proteins of interest. Since it relies on the reversible interactions between the protein to be purified and that of the affinity ligand coupled to the chromatographic immobilized matrix. One of the most popular techniques incorporated into affinity chromatography is the incorporation of a nickel-nitrilotriacetic acid (Ni-NTA) resin as an immobilized matrix in the affinity chromatography column. This resin forms a complex with the 6x His-tag which is commonly incorporated and is expressed at either the N or C terminal end of the protein. The formation of this complex allows the protein to be adsorbed while all other proteins that may have also been expressed during the expression phase are washed off the column. Introducing a competing chelating agent such as EDTA or imidazole aids in severing the complex between the immobilized Ni-NTA resin and the His-tags and retrieving the pure protein (Parikh and Cuatrecasas, 1975).

Circular dichroism is a spectroscopic technique that permits for the elucidation of secondary structural features of proteins. Due to the chirality of the peptide backbone, when exposed to plane polarised light, left and right handed polarised components are absorbed to a different extents (Kelly *et al*., 2005). The resulting radiation shifts from the original plane, the radiation which results from the chiral molecule displays “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).

Non-conjugated extrinsic chromphores can be used in order to display a proteins tertiary structure. These chromophores normally emit a weak fluorescence in polar environments but emit strong fluorescence in non-polar environments, 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 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).

A derivative of the nucleotide ATP called 2'(3)-N-methylanthraniloyl adenosine-5'-triphosphate (MANT-ATP) is used to study nucleotide binding to proteins using fluorescence spectrophotometry (Pisareva *et al*., 2006; Ni *et al*., 2000). Upon binding to a protein the MANT moiety attached to the polar ATP molecule allows for fluorescence upon excitation at a wavelength of 349 nm (Pisavera *et al*., 2006).

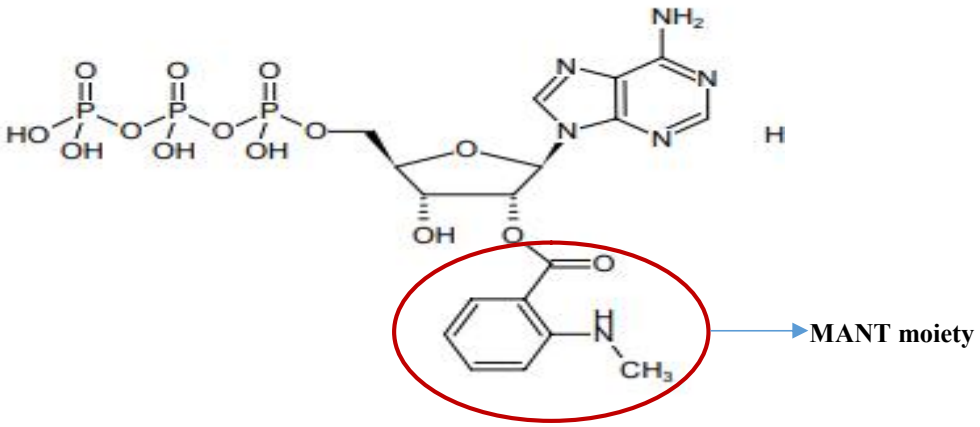


Figure 3.1: Depiction of MANT-ATP (Source: JenaBIOSCIENCE.com, 2018)

This chapter reports on the expression, purification, refolding and structural characterization of the cell-to-cell movement protein

3.2. Materials and methods

3.2.1. Cloning of the BC1 gene into the pET28a(+) expression vector

The BC1 gene (923 bp) was sent to GenScript USA for synthesis (codon-optimized), and cloning into the expression vector system pET28a(+) using *Nde*I and *Not*I restriction sites (figure 3.2).

ATGGACGCCCAATTTACCGTCACAGACAATAATTACATCAACAGCAAACGCACCGAGTA
CGCATTAACAAACGATGCTGCACCAATCAATCTCCAATTCCCAGGCTCATTCGAGCAGG
CTACCATGCGACTCAAGGCCGATGTATGAAAATCGACCACATTATAATTGAATACCGAA
ACCAGGTCCCATTTAACGCAACTGGGTCGGTTATCGTAGAAATCAGAGACAATCGCGTC
AGCCTTGACGACGCAGCTCAGGCAGCATTCACTTTCCCCATCGCTTGCAACGTGGACCT
CCACTACTTCTCTTCTACATATTTTTCGATTTCTGAACCTTCCCCTTGGAGAATCATGACA
GAGTCGAAGACTCAAACGTCATAGAAGGCGTGAAATTCGCATCCATCAAGGCCAAGCT
CCGATTATCATCGGCCAAACATTCCACGGACATACGTTTCAAACCCCCAACAATTAACA
TCTTATCCAAGGGATACACAAAAGACGCATAGACTTCTGGTCCGTGGAAAAAGGAGAA
ACAAGACGACGATTATTAAATCCCACTCCAACTGCTCGTAGTCAACAACCCATAACCCA
CAGGCCCATCACCATCCATCCAGGCGAAACATGGGCCACAAGGTCTCAGATTGGGCTAC
CAAGCTCATAGGCCCAGCACAGCTGGAACACTTTCGTTCACAGTCCATGAGAATGGACC
CATCGACAAACCAACGGACTTAGACAACGACTCCACAGAATATCCTTACCAGAAACTAC
ACAGATTACACACACCCGATTTGACCCAGGGGACTCGATATCACAGGCCCAATCCGACT
CCGTCTCCAGAAAGGACCTCGAGACACTGCTTGAGAGTACCATAAACAAGTGTCTCTAC
AAAATAAAATCCGACGCACCAAGGCAATTGTAA

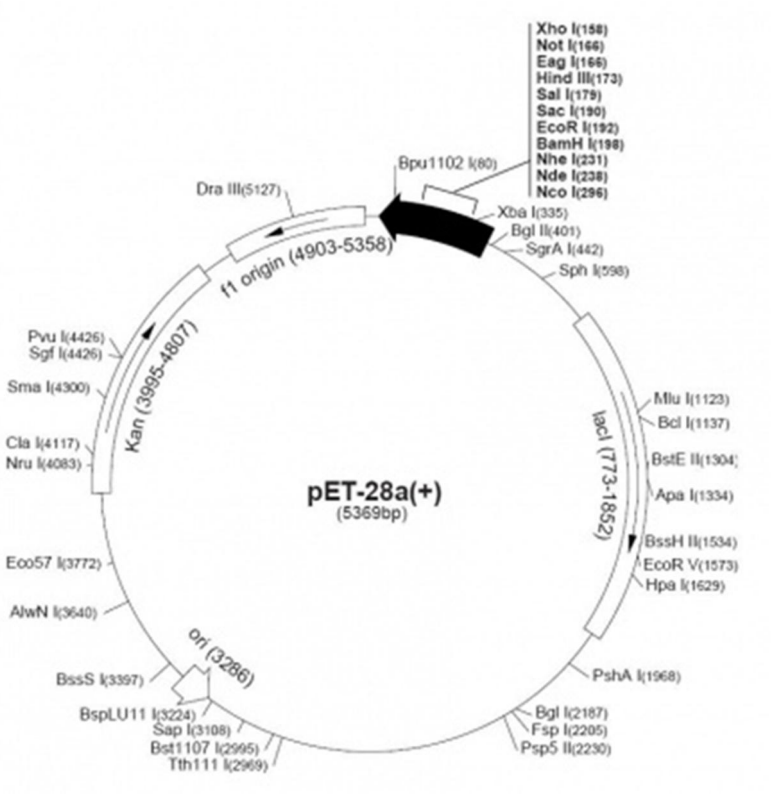


Figure 3.2: Vector map of expression vector pET-28a(+) size 5369 bp containing kanamycin resistance and a multiple cloning site as well as multiple restriction sites (Source: Novagen online, 2017).

3.2.2. Transformation of BC1 gene insert into *E.coli* BL21(DE3) competent cells and gene insert validation

The lyophilized cells (4 µg) (GenScript USA) was dissolved in sterile nuclease free water making up a working concentration of 100 ng/µl of cells, which was transformed via heat shock transformation (GenScript transformation protocol) into competent BL21(DE3) *E.coli*. The transformed colonies were selected and grown up in liquid SOB media [5 % (w/v) yeast extract (Sigma Aldrich, USA), 20 %(w/v) tryptone (Sigma Aldrich, USA), 0.5 %(w/v) NaCl (Sigma Aldrich), 1 M KCl (Merck, Germany)] 1 M $MgCl_2$(Merck, Germany),1 M $MgSO_4$(Merck, Germany) and 50 µg.mL^{-1} kanamycin (Sigma Aldrich, USA) overnight for further plasmid extraction and glycerol stock preparation. From prepared glycerol stocks the gene insert was validated by a restriction digest, using enzymes *NdeI* and *NotI* (New England Biolabs Inc).

Table 3.1: Table indicating reagents used for gene validation of BC1 insert.

Reagent	Volume
Nuclease free distilled water (Sigma Aldrich, USA)	38.5 µl
10 x Cut smart buffer (New England Biolabs Inc)	5 µl
Ndel	1 µl
Notl	1 µl
Plasmid DNA (making up a concentration of 1 µg).	4.5 µl
Total	50 µl

3.2.3. Protein overexpression

An overnight culture of transformed *E. coli* BL21(DE3) was prepared in a 50 ml falcon tube (Lasec-SA) in SOB [5 % (w/v) yeast extract, 20 %(w/v) tryptone, 0.5 %(w/v) NaCl, 1 M KCl, 1 M $MgCl_2$,1 M $MgSO_4$] medium supplemented with 50 µg.mL^{-1} kanamycin . The culture was placed in a 37 ^{0}C incubator (United Scientific) and allowed to shake at 200 rpm overnight for no longer than 16 h. Following incubation the overnight culture was centrifuged at 12 000 xg for 12 min. The supernatant was discarded and the pellet re-suspended in fresh SOB + 50 µg.mL^{-1} kanamycin (100 ml). The OD_{600} of the culture was verified on the Nanodrop 1000 (Inqaba Biotech) and was measured around ~0.1-0.2. The culture was then placed in a 37 ^{0}C incubator and allowed to shake at 200 rpm for ~1 h until the OD_{600} reached between 0.4-0.6. The culture was then induced with 0.25 mM of IPTG (Sigma Aldrich, USA) and placed at 37 ^{0}C to shake at 200 rpm for 4 h to allow the protein to express. Following four hours the cultures were centrifuged at 15 000 xg for 15 min. The supernatants were discarded and the pellets air dried and stored at -80 ^{0}C for further use.

3.2.4. Protein denaturation, purification and refolding

Preceding protein purification, the protein had to undergo protein denaturation. The following diagram (figure 3.3) indicates the steps involved in preparing the protein for denaturation and solubilisation prior to undergoing purification.

Figure 3.3: A list indicating the steps to be followed during protein denaturation and solubilisation of the movement protein BC1.

Purification of the BC1 protein was completed using a gravity flow column containing a nickel charged NTA Profinity™ IMAC resin (Bio-rad). The steps described in figure 3.4 were carried out to ensure removal of the His tag protein:

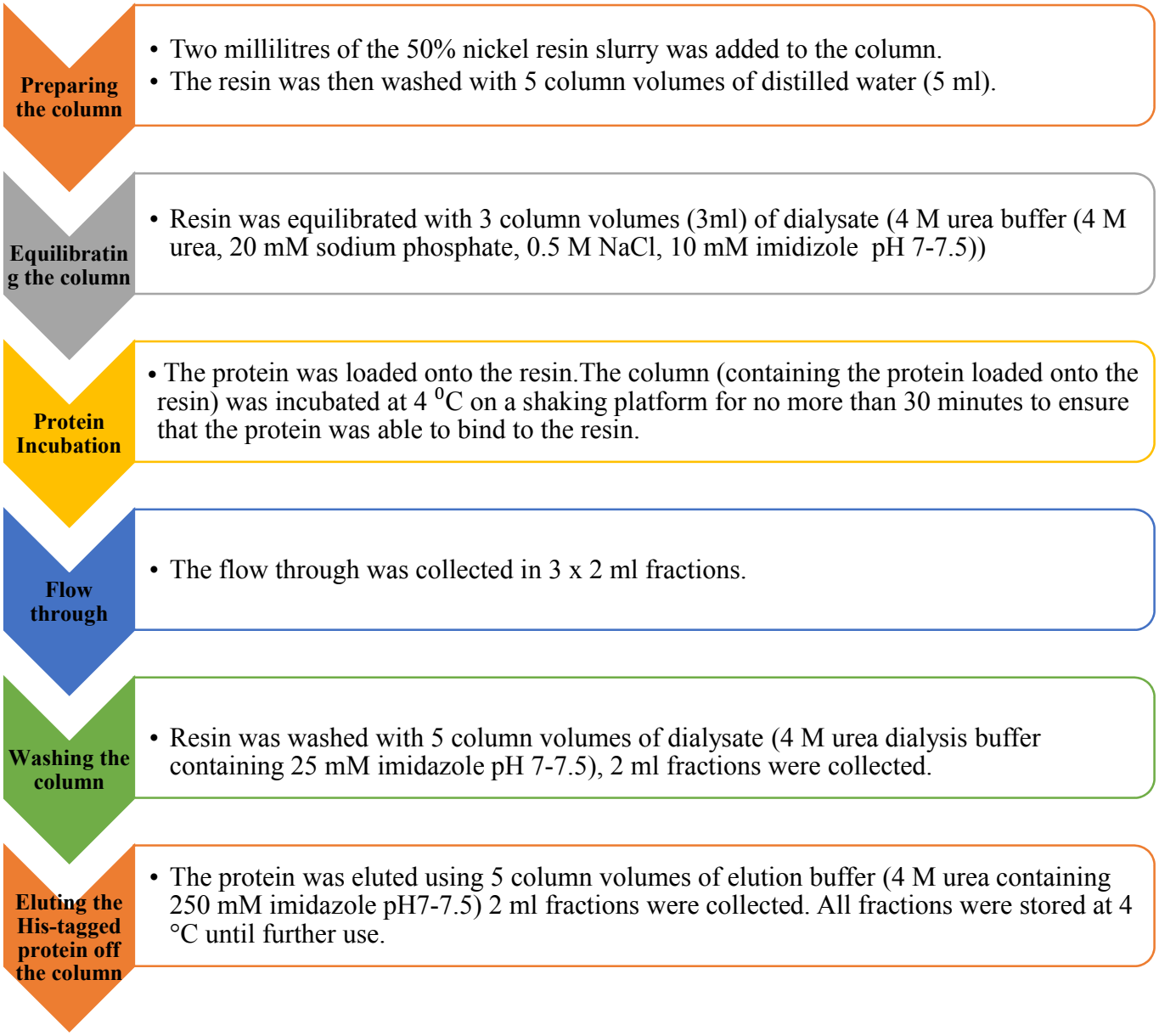


Figure 3.3: A flow diagram indicating the steps involved in purifying BC1 protein on a gravity flow nickel charged affinity chromatography column.

All the eluted protein fractions were prepared with 2X SDS buffer (added in a 1:1 ratio) and run on a 12 % SDS-PAGE mini-PROTEAN gels (TGX Stain-Free™ Fastcast™ acrylamide kit Bio-Rad Laboratories, Inc). Based on the SDS-PAGE results, column fractions containing the purified protein (eluted fractions 1-3) were pooled together. The pooled fractions were diluted into 60 ml of 4 M urea (4 M urea, 20 mM sodium phosphate, 0.5 M NaCl, pH 7-7.5). This solution was transferred into

dialysis tubing and placed into 2 L of refolding buffer (50 mM Tris-HCl pH 10, 0.5 M L-Arginine (Sigma Aldrich, USA), 2 % (v/v) Triton X-100 (Merck, Germany), 10 mM DTT (Roche, USA)) overnight. The dialysis tubing was then transferred into a new refolding buffer (50 mM Tris-HCl pH 10, 0.5 M L-Arginine, 10 mM DTT) for two overnight dialysis cycles. Following the refolding procedures a SDS-PAGE gel was run in order to check if any significant aggregation had occurred during refolding.

3.2.5. Protein quantification

Following protein refolding, the protein was concentrated using sucrose (to minimise the risk of aggregation the protein was diluted in 60 ml of 4 M urea as stipulated above, therefore concentration with sucrose was preferred to permit for the large volume to be concentrated). The closed dialysis bag containing the protein was placed into a glass bowl on top of a bed of sucrose granules (Sigma-Aldrich, USA (99.0%)), additional sucrose was then poured on top of the dialysis bag. The bowl was covered with cling film and placed at 4 °C. After 5 hours the dialysis bag was removed from the sucrose and the remaining concentrated protein was transferred to a spin filter column (Vivaspin 20, Sartorius, UK), and centrifuged at 8 000 xg for 35 minutes where the protein was concentrated down to a final volume of 3 ml.

The protein 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:

$$A = \varepsilon c l$$

$A = \varepsilon c l$ 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).

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

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) (BC1- 31 650 M^{-1} cm^{-1}).

3.2.6. Far-UV Circular Dichroism

A stock solution of the protein was dialysed overnight against (Milli-q water) at 4°C. The dialysed solution was 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 in a 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 (Milli-q water) 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 BC1 was estimated the CONTIN algorithm in the Dichroweb server (Whitmore and Wallace, 2004).

3.2.7. Extrinsic fluorescence ANS binding assay

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 ratio of protein to ANS was also determined for the assay (indicated in table 3.2 below).

Table 3.2: Table indicating the volumes and ratios of protein to ANS and denaturant (8 M urea) for extrinsic ANS fluorescence binding assay.

	ANS+Buffer (µl)	Protein (µl)	8M Urea (µl)	Ratio per tube	Total per tube (µl)
Tube 1	300	200	0	3:2:0	500
Tube 2	300	0	200	3:0:2	500
Tube 3	200	200	100	2:2:1	500

All three samples and two blanks (ANS and 8 M Urea) were analysed using a Jasco FP-6300 spectrofluorimeter. A 1 cm pathlength cuvette and 200 $nm.min^{-1}$ scan speed 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.

3.2.8. Intrinsic Fluorescence MANT-ATP binding Assay

A 100 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 10, 0.5 M L-Arginine) and in 8 M Urea (8 M Urea in 50mM Tris- HCl, pH 8.5, 0.05 % NaN_3 pH 7-7.5). Per 100 µM of MANT-ATP, 35 µM of protein was used (ratio 20:7). 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 and scan speed used was 200 $nm.min^{-1}$. 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 $nm.min^{-1}$.

3.2.9. Intrinsic Tryptophan Fluorescence

Intrinsic tryptophan fluorescence was measured herein to determine the polarity of the tryptophan residues of BC1. Components in sample preparation are shown in table 3.3 below.

Table 3.3: Table showing the presence and absence of components in sample preparation for intrinsic tryptophan fluorescence analysis.

Component			
Refolding Buffer (50 mM Tris-HCl pH 7.5 and 0.5 mM L-Arginine)	**Protein BC1 10 µM**	**ATP 10 µM**	**8 M Urea Solution**
✓	X	X	X
✓	✓	X	X
✓	✓	✓	X
X	X	X	✓
X	✓	X	✓
X	✓	✓	✓

The six samples were analysed using the Jasco FP-6300 spectrofluorimeter with a 1 cm pathlength cuvette. The tryptophan residues were excited at 280 nm and the fluorescence spectra were selectively recorded between a range of 300-500 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 in 50 mM Tris-HCl pH 7.5 containing 0.5 mM L-Arginine or 8 M Urea. Background fluorescence (free unbound tryptophan residues) was corrected for by subtracting background values from the actual protein spectra (protein bound to tryptophan residues).

3.3. Results and Discussion

3.3.1. Synthesis of the BC1 gene and cloning into the pET-28a(+) expression vector

SACMV BC1 codon-optimized gene was successfully cloned into the expression vector pET 28a(+) and met all the quality controls (certificate in appendix)

3.3.2. Transformation of BC1 cells into *E.coli* BL21 (DE3) competent cells and gene insert validation

The *E. coli* strain BL21 (DE3) was selected for transformation and expression as the strain is able to host a wide variety of vectors for protein expression (Baneyx, 1999). For the expression of recombinant proteins it is important to select a host system which is amenable to the type of protein being expressed and allows for post translational modification found in eukaryotes. While *E.coli* has been shown to be one of the preferred organisms used to express proteins, as its genome has been substantially studied and its mechanism is largely understood with regards to transcription, translation and protein refolding (Baneyx, 1999), there are drawbacks in expression of eukaryotic-derived proteins in prokaryotic organisms. Yeast is another well-studied eukaryotic system used for expression of proteins (Wilson and Walker, 2012) but is more difficult to work with because cell growth is slower and the level of expression is not always high and post-translational modifications may 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 (Nasser *et al*., 2003).

Successful cloning of BC1 into the pET28A expression vector was achieved by restriction digest (figure 3.5). Figure 3.5 (lanes 2 and 4) indicate the correct BC1 gene size (923 bp) (Berrie *et al*, 2001). The lack of presence of any other bands besides those indicating the plasmid and the gene of interest it shows a clean digest. There is no evidence that there is digestion through the gene itself.

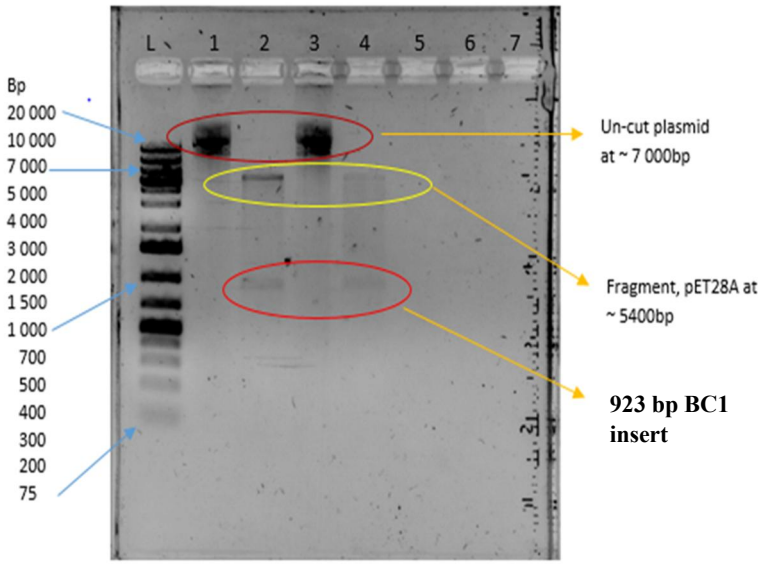


Figure 3.5: A 1% agarose gel showing the digestion products from restriction digests carried out on the pET28a(+) vector using restriction enzymes *NdeI* and *NotI*. L: Ladder Generuler 1 KB plus (Thermo Scientific), Lanes 1: Un-digested; 2: digested with restriction enzymes; 3: Un-digested; 4: digested with restriction enzymes; 5; 6; 7: empty.

3.3.3. Protein Expression

From this study, it was demonstrated that BC1 is predominantly insoluble (fig 3.6A and B) and was expressed under non-induced and induced conditions. In order to ascertain the optimal expression conditions of a recombinant protein, a number of variables have to be tested. These include often manipulating incubation time and temperatures, the addition of chaperones and varying concentrations of IPTG (Clark, 2001; Papaneophytou and Kontopidis, 2014; Wilson and Walker, 2012). It was established that the optimal expression conditions for the BC1 insoluble protein was 37 °C for 4 h and induction by 0.25 mM IPTG (lane 7 fig3.6B). While low temperatures are suggested for the expression of soluble proteins (Clark 1999, 2001; Papaneophytou and Kontopidis, 2014) this was not the case for the insoluble SACMV BC1. While lower concentrations of soluble protein were present at 37° C (lanes; 2, 6: figure 3.6.A; lanes; 1, 5; figure 3.5.B) this is not sufficient for downstream applications. Due to the inability to retrieve a soluble fraction at a lower temperature or with using various concentrations of IPTG, it was decided that the BC1 protein had to be denatured and folded for further characterization.

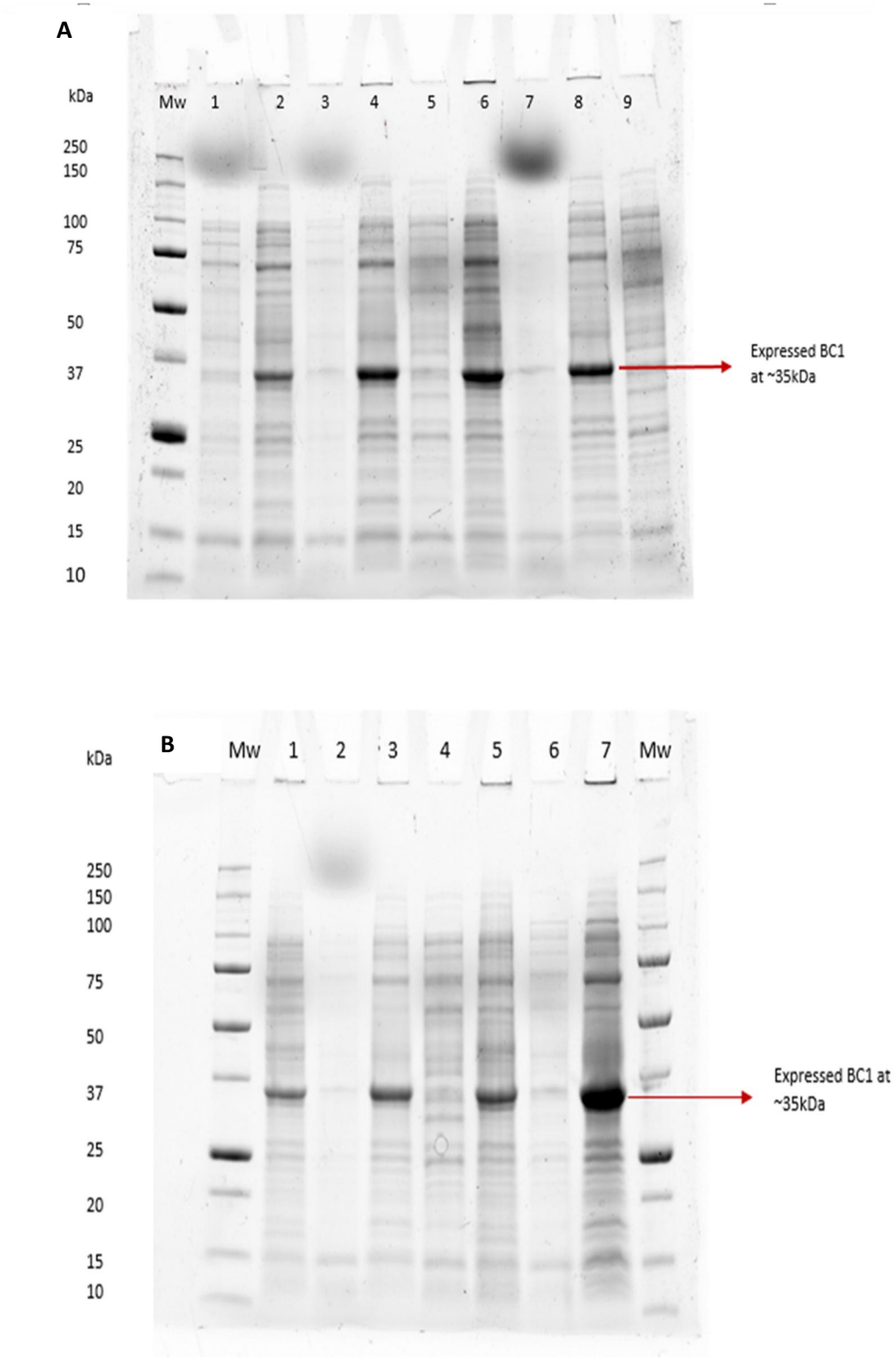


Figure 3.6: SDS-PAGE gels (12%) showing expression of protein BC1 (35 kDa (including the His tag)) in non-induced and induced (0.25 mM IPTG at 37 °C) fractions for 1-2 h (**A**) and 3-4 h (**B**) (A) MW: Molecular weight marker; Lanes 1: Soluble non-induced 1 h; 2: soluble induced 1 h; 3: insoluble non-induced 1 h; 4: insoluble induced 1 h; 5: soluble non-induced 2 h; 6: soluble induced 2 h; 7: insoluble non-induced 2 h; 8: insoluble induced 2 h; 9: soluble non-induced 3 h. (B) MW:

Molecular weight marker; Lanes 1 soluble induced 3 h; 2: insoluble non-induced 3 h; 3: insoluble induced 3 h; 4: soluble non-induced 4 h; 5: soluble induced 4 h; 6: insoluble non-induced 4 h; 7: insoluble induced 4 h; MW: Molecular weight marker.

3.3.4. Protein denaturation, purification and refolding

Affinity chromatography is a technique commonly selected for the purification of proteins. A number of His-tagged proteins are often purified by means of affinity chromatography (Campbell and Farrell, 2009). In this study, his-tagged BC1 was bound to the affinity column using nickel before refolding (figure 3.7). Eluted unfolded BC1 protein was collected in fractions 6 and 9 (figure 3.7 lanes 6-9). While the purification subsequent to denaturation (gel) indicated fractions in fractions; 6, 7 and 8 it must also be taken into consideration that the column resin could have been overloaded and that there may have been an abundance of the protein present due to the 30 minute incubation.

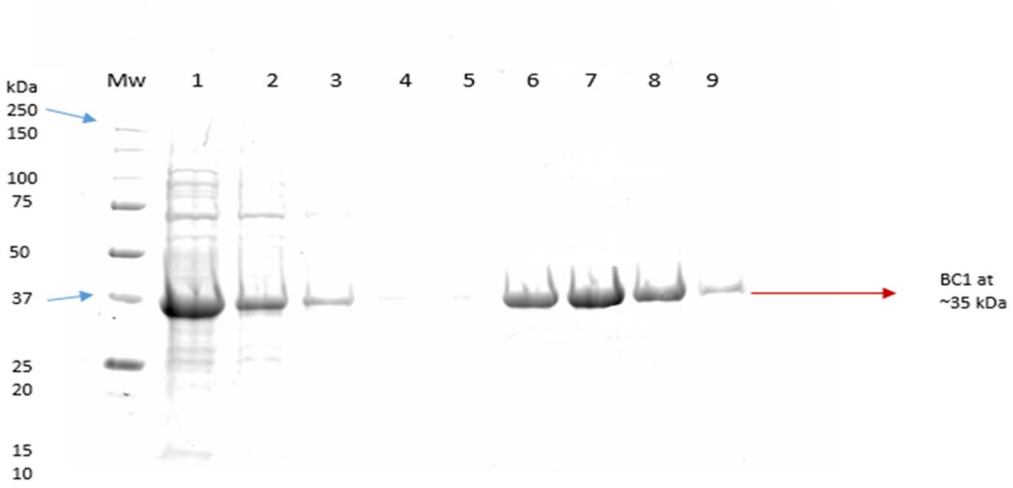


Figure 3.7: A 12% SDS-PAGE gel showing the elution fractions of BC1 from an IMAC column after denaturation before refolding. MW: Molecular weight marker protein standards unstained, Lanes 1: Protein fraction after solubilisation with 4 M Urea; 2: Flow through fraction; 3: Wash fraction 1; 4: Wash fraction 2; 5: Wash fraction 3; 6: Protein elution fraction 1; 7: Protein elution fraction 2; 8: Protein elution fraction 3; 9: Protein elution fraction 4.

The purification of refolded protein showed that the protein elutes in the flow through (lane 2) and first two wash elutions (lanes 3-4) (figure 3.8). This indicates that the his-tag may have been obscured stepwise refolding steps as ideally the protein should have eluted in the final elution steps with the 250 mM concentration of imidazole, similar to the initial purification (subsequent to denaturation). This is possible as the addition of a poly-histidine tag has the ability to change the

structural conformation of a protein (Kuo and Chase 2011). This indicates that during refolding the protein could have incorporated the poly-histidine tag into its tertiary structure burying the poly-histidine tag preventing binding to the nickel on the column. This would imply that the poly-his tags are only accessible to the nickel in the denatured form of BC1.

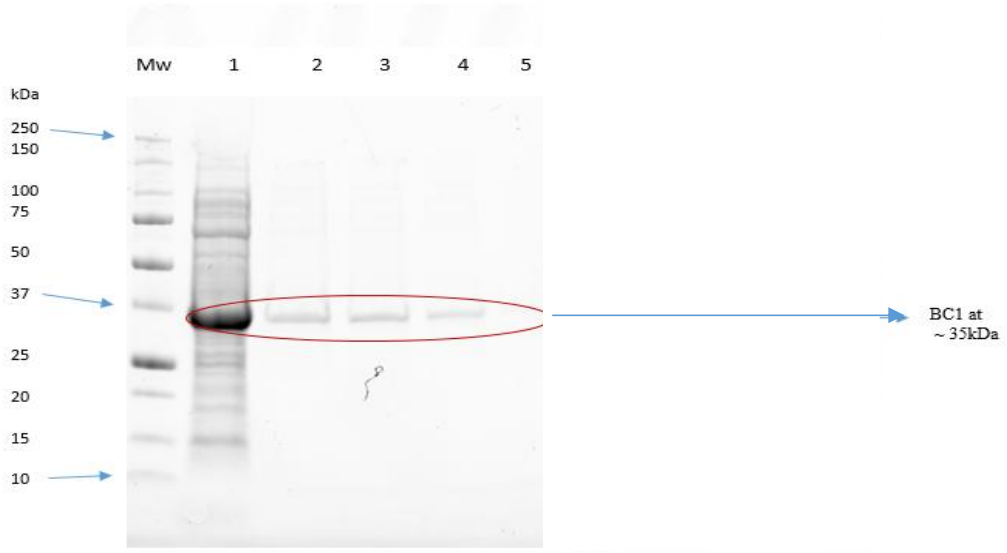


Figure 3.8: A 12% SDS-PAGE gel showing the different eluted fractions from an IMAC column retrieved after the refolding of BC1. MW: Molecular weight marker Protein Standards unstained (Bio-rad), Lanes 1: Protein fraction after solubilisation with 4 M Urea; 2: Flow through fraction; 3: Wash fraction 1; 4: Wash fraction 2; 5 : empty.

3.3.5. Protein quantification

Protein quantification was determined by evaluating the concentration of the protein by means of a serial dilution and UV spectroscopy (figure 3.9). BC1 was quantified to be 31.42 µg/ul, which was sufficient to continue with downstream structural applications. Ultraviolet visible absorption spectroscopy is a suggested tool for assessing protein quantification and purity (Wilson and Walker, 2012). Absorbance at 340 nm was subtracted from the absorbance at 280 nm in order to give a corrected absorbance or correction factor this is suggested by literature (Wilson and Walker, 2012). The correction factor takes into account any interference by means of light or noise which may impact the result.

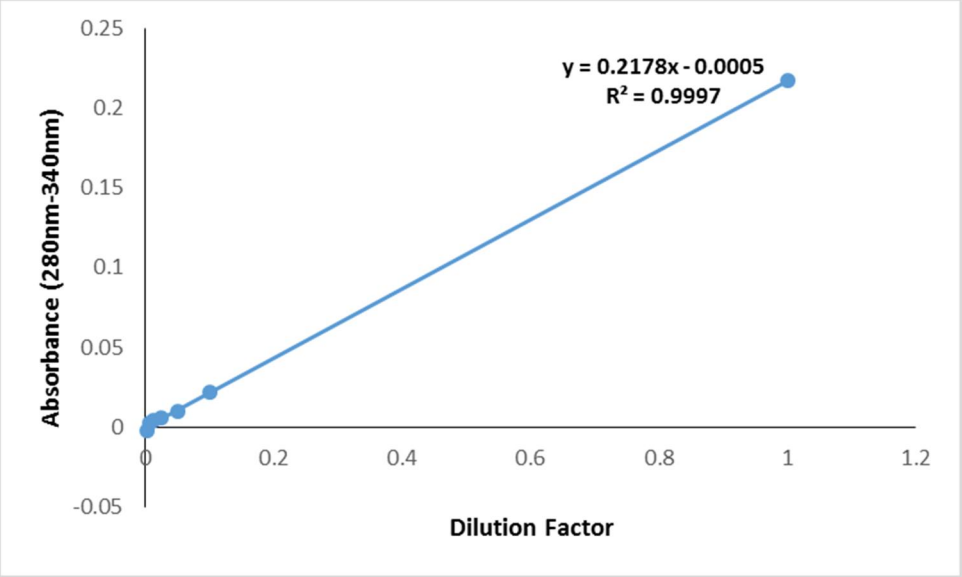


Figure 3.9: A graph of corrected absorbance against inverse dilution factor. The R^2 value was determined to be 0.9999 and the slope of the equation (which is 0.8826) was used to determine the concentration of BC1 (substituting into the Beer- Lambert law).

3.3.6. Far- UV Circular Dichroism

Far UV Circular dichroism (CD) spectroscopy (between 180-250 nm) estimates the secondary structure content of proteins and peptides, and determines the fraction of α-helices and β-strands with respect to the amino acids involved in unordered forms in an amide backbone (Micsonai *et al*., 2015; Wilson and Walker, 2012). In this study, far UV CD was utilized in order to determine the secondary structure of BC1 with the assistance of the Dichroweb server applying the CONTIN algorithm. The CD spectrum (figure 3.10) confirmed that the BC1 secondary structure is predominantly (57.2%) β-strand this as the pattern of the spectrum corresponds with the reference spectra as per Kelly and Price, 2000 (peak at ~200 nm and a trough at ~ 210 nm). The data analysed and retrieved from Dichroweb further shows this (table 4). This result does not correlate with the *in silico* prediction (chapter 2; fig. 2.4) which indicated that a higher (65%) predominate secondary structure of BC1 would be disordered and a lower percentage (25%) would be β- strands. Nonetheless, a high level of secondary structure disorder from this CD data (37.2%) was predicted (table 3.4), and could represent the C terminus of BC1 which is unordered, as indicated by the *in silico* study (chapter 2; figure 2.5). While other structural characterization analyses were carried out using the refolding buffer this analysis was carried out using milli-q water as this was the only solution which did not interfere with the signal. The refolding buffer contains tris, which has a high absorbance value below 200 nm (195 nm) which is not recommended for Far-UV CD spectroscopy as it gives off a high background noise (Micsonai *et al*., 2015; Wilson and Walker, 2012).

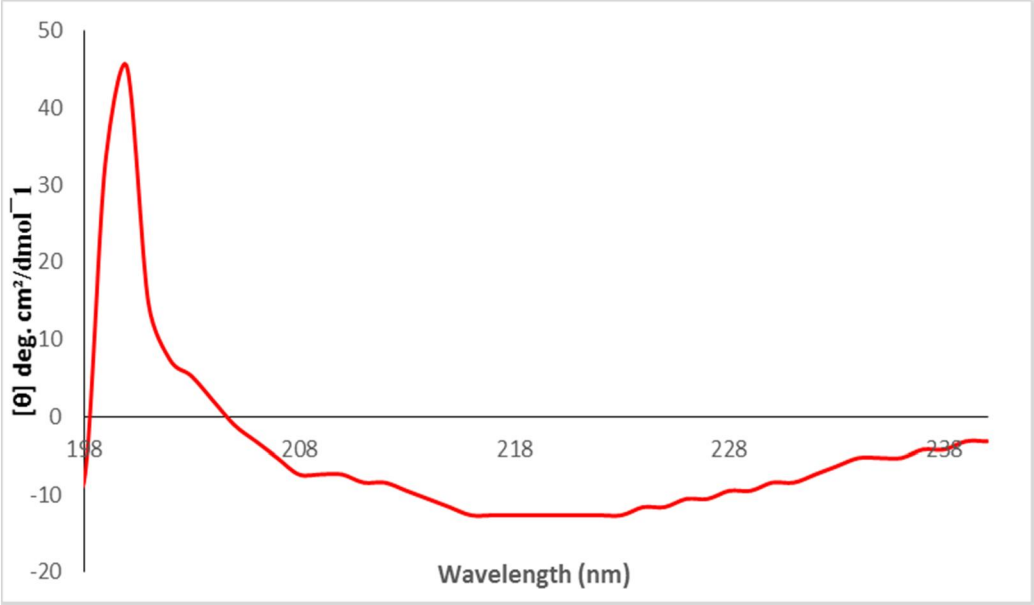


Figure 3.10: Far UV-CD spectrum of BC1. Signals were collected using 5 µM of protein in Milli- q water. The spectrum indicates that the protein displays a secondary structure which is predominantly β- strands.

Table 3.4: Summary of the secondary structure content using the CONTIN alogothrim implemented in the Dicroweb server.

Secondary Structure content (%)			
α-helix	β-strand	Unordered	Total
5.2	57.2	37.2	100

3.3.7. Extrinsic ANS fluorescence binding assay

ANS (1-anilino-8-naphthalene sulphonate) is a common non-conjugating extrinsic chromophore. It emits a weak fluorescence in polar environments (aqueous solutions) and when in a non-polar environment, fluorescence increases significantly. An increase in quantum yield and an increase simultaneously in the blue shift of the emission wavelength λmax from 510 nm to 480 nm is indicative of the protein binding to the ANS (Hawe *et al*., 2007; Schönbrunn *et al*., 2000; Wilson and Walker, 2012). In this study, the tertiary structure of BC1 was examined with the aim of identifying possible hydrophobic pockets within the protein. Figure 3.11 shows a significant shift in quantum

yield and simultaneous blue shift (λmax) from 510 nm to 480 nm for BC1 (native form). This indicates that BC1 provides ANS a non-polar environment upon binding, suggesting that BC1 possesses hydrophobic binding pockets in the native form. Figure 3.11 also shows that in the denatured form BC1 showed a significant shift in quantum yield but did not show a blue shift (λmax) from 510 nm to 480 nm (λmax remained at 510 nm). This indicates that the denatured form of BC1 has no binding pockets available to bind ANS as compared to its native counterpart. Since studies on BC1 are novel it is not known if ATP binds to the protein. However, since BC1 is involved in the cell-to-cell movement of a ss/ds DNA complex as suggested by the proposed 'couple-skating' and 'Relay-race' models it can be assumed that BC1 would bind ATP. This as in order to move complexes either to adjacent cells or transfer a complex to a cell moving along the plasma membrane energy would be required. In order to confirm the presence of hydrophobic binding pockets, an intrinsic MANT-ATP assays was undertaken.

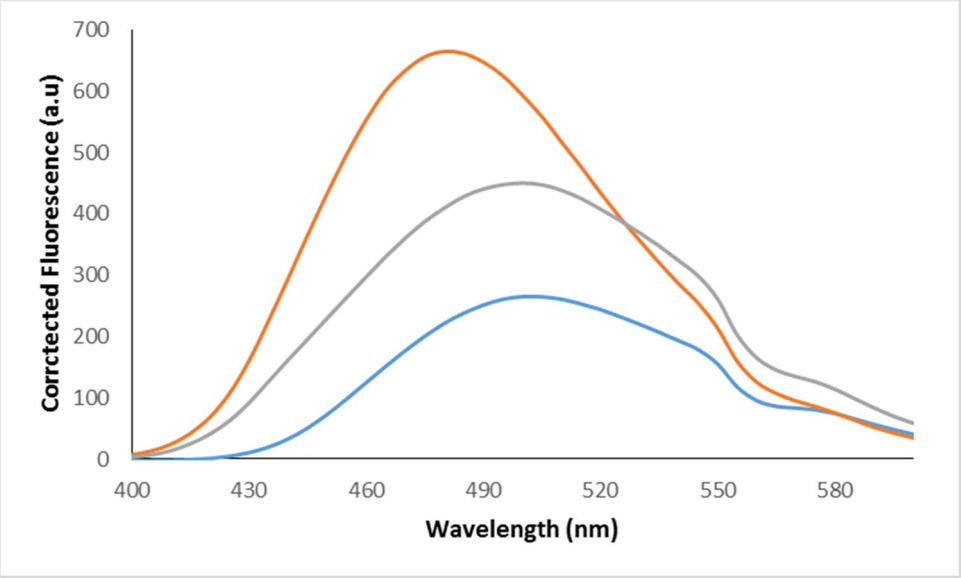


Figure 3.11: A spectrum showing extrinsic ANS fluorescence. Spectra of 200 µM free-ANS (blue) bound to 10 µm BC1 (orange) and 8 M Urea+ 10 µM BC1 (grey) All spectra were subtracted from the free-ANS spectrum.

3.3.8 Intrinsic Fluorescence: MANT-ATP Binding Assay

The ATP binding assay uses 2' (3')-N-methylanthraniloyl-ATP (MANT-ATP) where the ATP molecule displays fluorescence. The method is employed to reveal the degree to which a protein binds ATP (Shi *et al*., 2010). When excited at lower wavelengths, free MANT-ATP displays low levels of fluorescence, when bound to a target, such as BC1, a significant peak emerges (Shi *et al*., 2010). Herein, figure 3.12 shows that in the native form of BC1, while containing hydrophobic

binding pockets, also contained some hydrophilic regions which was able to bind the ATP nucleotide with the MANT moiety attached to it. Upon excitation at 360 nm, the fluorescence of MANT-ATP increased upon binding to BC1 in the native state and increased subsequently higher upon binding to BC1 in the denatured state without a significant shift in the emission maximum (400 nm). Indicating that BC1 has the ability to bind ATP, which binds to the hydrophilic exposed regions of the protein.

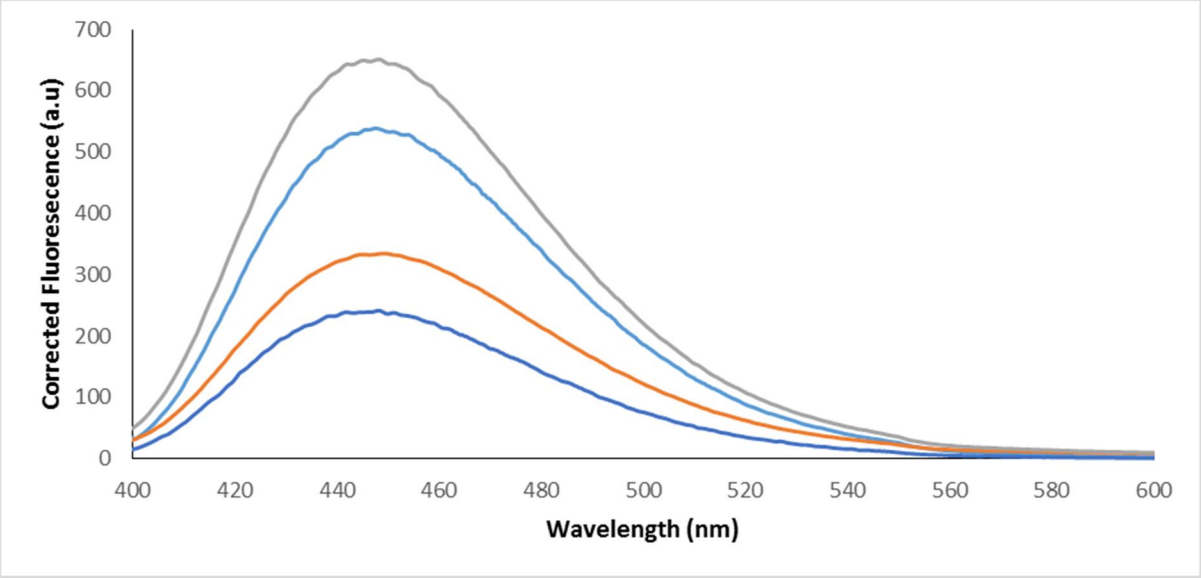


Figure 3.12: Intrinsic MANT- ATP fluorescence spectra. Spectra of 100 µm free MANT-ATP (orange) bound to: 10 µm BC1 + MANT-ATP (light blue), 8 M Urea + 100 µm MANT-ATP + BC1 (grey), and 8 M Urea + 100 µm MANT-ATP (dark blue) (all spectra subtracted from free MANT-ATP spectrum).

3.3.9. Intrinsic Fluorescence: Tryptophan Fluorescence

Proteins possess three intrinsic fluorophores or aromatic amino acids that intrinsically fluoresce namely: tyrosine, phenylalanine and tryptophan. Intrinsic tryptophan fluorescence maximum and intensity are highly influenced by the polarity of the micro-environment, hydrogen bonding and other non-covalent interactions, intrinsic protein fluorescence is due to aromatic amino acids, mainly tryptophan, which can be selectively measured by exciting at 295 nm. Changes in emission spectra of tryptophan are due to the protein conformational transitions, subunit association, ligand binding or denaturation, which affect the local environment surrounding the indole ring. Tryptophan fluorescence in this study was utilized for the purpose of monitoring the conformational changes in BC1. BC1 contains three tryptophan residues. Figure 3.13 (A) shows a blue shift (λmax) from 375 nm to 345 nm for native BC1 in the absence of ATP (orange) and native BC1 in the presence of ATP (grey) even though there is a significant drop in quantum yield as compared to the buffer control

(blue). Figure 3.13 (B) shows an increase in quantum yield for denatured BC1 in the absence (purple) and presence of ATP (red), while no significant blue shift can be seen when compared to the urea buffer alone (green). The changes in the emission spectra of tryptophan occur in response to conformational transitions, substrate binding, denaturation or subunit association (Lakowicz, 2006). The spectra results of BC1 indicates that when in the presence of ATP in the native form the protein may not undergo much conformational change, however in the denatured form in the presence of ATP the protein undergoes a significant change in conformation. Based on this analysis it can be proposed that while the protein is in its native form in the cell, binding of ligands such as ATP to BC1 causes no conformational change. It is possible that after the binding of a different ligand or complex in the cell BC1 changes conformation and this allows the protein to transfer the DNA/ DNA-protein complex to adjacent cells or to cells along the plasma membrane as indicated by the two models ('Couple-Skating and 'Relay-Race') describing the proposed active role of BC1 (Krentz *et al*., 2012; Fontes *et al*., 2004).

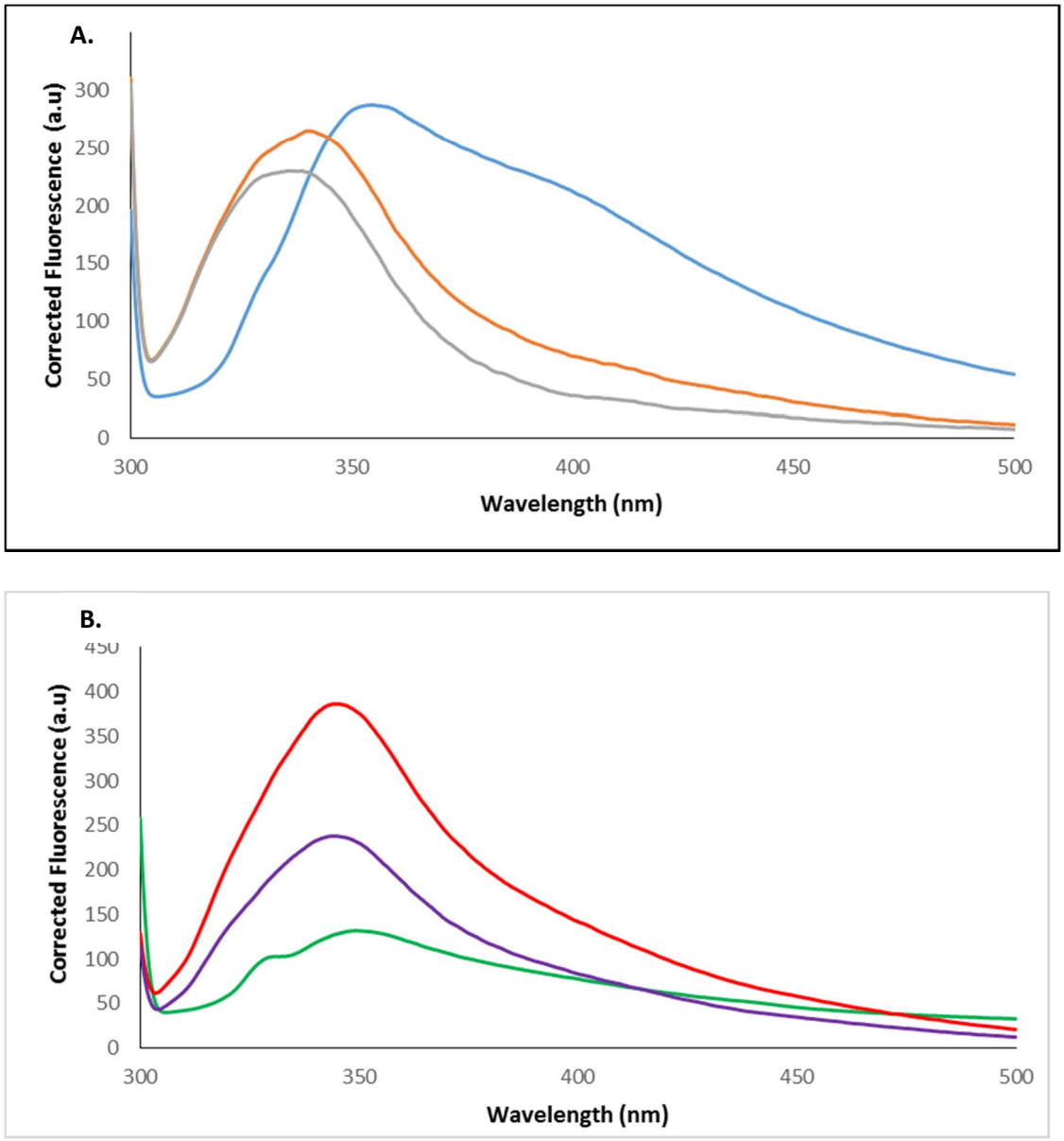


Figure 3.13: Intrinsic tryptophan fluorescence spectra. (A) Spectra of 10 µM BC1 protein in the absence of ATP (orange) and bound to 10 µM BC1 in the presence of ATP (grey). The sample buffer (50 mM Tris-HCl pH 7.5 containing 0.5 mM L-Arginine) spectrum signal is represented in blue. Each spectrum is the average of three accumulations of three replicate samples. (B) Spectra of 8 M urea (green), 10 µM BC1 (denatured in 8M Urea) in the absence of ATP (purple) and in the presence of ATP (red).

3.4. Conclusions

The results of this research have contributed to the first glimpse of the secondary structure of SACMV BC1. It is now established that BC1 has a secondary structure which predominantly consists of β-strands, and its predicted large unfolded domain presents some interesting speculations as to structure and function. BC1 has the ability to bind biomolecules due to its exposed hydrophobic pockets within its tertiary structure as shown by the corresponding ANS binding studies. BC1 changes in conformation in the presence of ATP, and adopts a different conformation when denatured as compared to when it is in its natured state. This may suggest that when in the cell (in the native form) the protein undergoes no conformational binding when bound to biomolecules such as ATP however, other ligands or complexes could cause a conformational change. Further structural characterization studies of this protein will be undertaken, and this study also serves as the foundation to unravelling the geminivirus BC1 functions *in vivo*.

Chapter 4: Preliminary studies of predictive interactions between BC1 and selected cassava proteins

4.1 Introduction

Proteins must interact with other molecules in order to fulfil their roles. Proteins do not exist isolated in a cell. Proteins which are involved in common pathways exist in a loose interaction. If an interaction can be identified between a protein of unknown function and a well-characterized protein it can be assumed that the unknown protein has a function somehow related to the protein of known function. Protein-protein interactions play critical roles for many physiological, pathological and developmental processes in most organisms (Braun *et al*., 2013; Wilson and Walker, 2012; Zhang *et al*., 2010).

Geminivirus proteins have been found to interact with a number of proteins within their respective hosts (Jeske, 2009; Hanley-Bowdoin *et al*, 2013), which enable them to establish infection. This occurrence is dependent on viral-host protein structures and interactions which leads to signalling and biological processes downstream (Whitham and Wang, 2004; Jeske, 2009; Guitierrez, 2004). Proteins of geminiviruses have been associated with host-responsive processes such as transcriptional regulation, DNA-replication, cell cycle control, cell proliferation and differentiation as well as molecular trafficking in whole plants (Ascencio- Ibáñez *et al*., 2008; Mariano *et al*., 2004; Fontes *et al*., 2004).To determine the complete function of BC1 in SACMV movement, it is important that host candidate protein interactors with BC1 be determined.

One such candidate host protein interactor of interest is Histone H3. Histone H3 often exists in a complex with Histones H2 and H4 (Allie *et al*., 2014). Histone H3 is a host nucleoprotein which has been identified in *Arabidopsis* and has been found to interact with the nuclear shuttle protein (NSP) and movement protein BC1 (MP) of BDMV (Zhou *et al*., 2011). Research done by Zhou *et al*., 2011 confirmed Histone H3 as a NSP and MP interacting protein in BDMV in *Arabidopsis*. The research was confirmed through co-immunoprecipitation assays *in vitro* and *in vivo*. Consequently, Histone H3 was selected as a potential interactor between the cassava host and the MP of SACMV.

Host proteins (such as cellular chaperones) are required for viral replication as well as cell-to-cell movement in RNA viruses (Verchot, 2012). Plant host proteins such as heat shock proteins (HSPs) 40, 70 and 90 have been found to interact with viral proteins in a number of RNA and DNA viruses (Chen *et al*., 2008; Mayer, 2010). Particularly, HSP 70 has been suggested in the involvement of the cell-to-cell transport of movement proteins of various RNA viruses such as closterovirus (Krenz *et al*., 2010; Lu *et al*., 2009; Shimizu *et al*., 2009; Soellick *et al*., 2000). Demonstration that HSPs interact with MPs in a number of viruses (DNA and RNA), and considering that HSP70 is involved with the MPs of RNA, a cassava HSP70 was selected in this study as a possible interactor with the MP of SACMV.

One of the most successful and commonly used methods for the investigation of protein-protein interactions is the yeast-two-hybrid (Y2H) system (Zhang *et al*., 2010). The system exploits and manipulates transcription factors, such as GAL4, that are separated into coding regions for two domains, one being a DNA-binding domain the other being a trans-activation domain. Each of these domains is linked to a different protein one of which is the potential binding target and the other which is the known protein (in separate yeast cells). When the two yeast cells are mated, if the bait (known protein) and prey (potential binding target) bind to each other, the two domains of the transcription factor will be brought together and will then be active and will further bind to the promoter of the reporter gene inducing its expression. The identification of cells which express the reporter gene product is proof that the bait and prey proteins interact.

This chapter reports on preliminary studies of predictive interactions between the cell-to-cell movement protein BC1 of SACMV and H3 and HSP70 from cassava, using *in silico* studies. Post identification of suitable candidates, cloning of the Histone H3 and HSP70 genes into the pDEST22 expression vectors was also achieved for futureY2H assays.

4.2. Materials and Methods

4.2.1. Selection of cassava host proteins (bait proteins)

The selection of bait proteins was done by a review of literature indicating complexes or interactions with geminivirus host proteins and geminivirus MPs:

- Histone H3 was selected based on a previous study (Zhou *et al*., 2011) and based on its up-regulation (within the Histone superfamily protein group) at 32 and 67 days post inoculation in the cassava T200 landrace (Allie *et al*., 2014).
- HSP70 was selected based on previous studies (Chen *et al*., 2008; Krenz *et al*., 2010; Lu *et al*., 2009; Mayer, 2010; Shimizu *et al*., 2009; Soellick *et al*., 2000) and based on its down-regulation at 32 and 67 days post inoculation in the T200 landrace (Allie *et al*., 2014).

4.2.2. *In silico* studies of Histone H3 and HSP70

The protein and gene sequences (in Appendix B) for both the proteins in cassava were accessed using Phytozome v12.1 and the accession numbers for both proteins (Histone H3: cassava4.1_017700m.g; HSP70: cassava4.1_002955m.g). These sequences were confirmed with a BLAST search. The protein sequences were uploaded into the Predict Protein webserver where the predicted protein binding regions were compared with the BC1 predicted protein binding regions.

4.2.3. Cloning of the Histone H3 and HSP70 genes into the pDEST22 expression vectors

The Histone H3 and HSP70 template sequences were sent to GenScript USA for synthesis, and cloned into the expression vector system pDEST22, Using *HindIII* cut sites for Histone H3 and *BamHI* cut sites for HSP70.

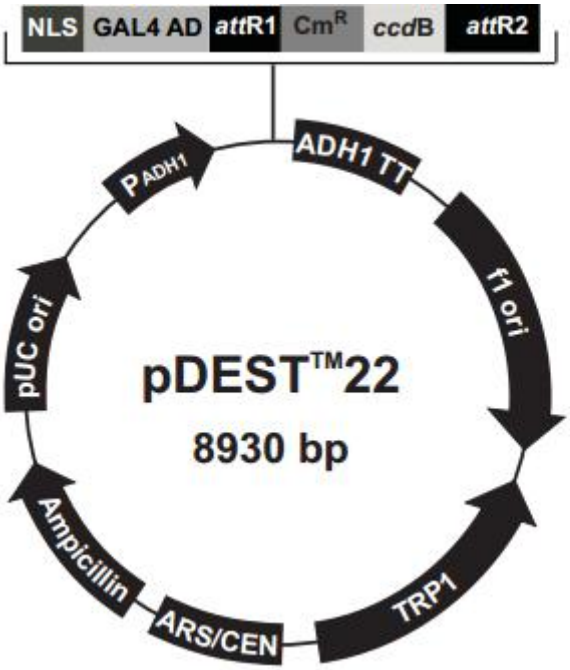


Figure 4.1: Vector map of expression vector pDEST™22 containing ampicillin resistance and a GAL4 DNA activation domain.

4.2.4. Cloning of BC1 gene into pDONR™221 vector

BC1-attB site specific primers, 5' BC1-attB (GGGG ACA AGT TTG TAC AAA AAA GCA GGC TAT GGA CGC CCA ATT TACC) and 3'BC1-attB (GGGG ACC ACT TTG TAC AAG AAA GCT GGG TTT ACA ATT GCC TTG GTGC) were designed to bind to the BC1 gene and amplify the gene. Polymerase chain reaction (PCR) was carried out using 200 ng of BC1 plasmid using the T100™ Thermal Cycler (Bio Rad). The final reaction mixture contained 1X Phusion Green HF Buffer, 200 µM of each of the dNTPs, 0.5 µM each of the forward (5' BC1-attB) and reverse (3' BC1-attB) primers, 0.02 U/ µL of Phusion DNA polymerase, in a total volume of 50 µl. PCR was run for 34 cycles; initial denaturation at 98 ºC for 30 sec, denaturation at 98 ºc for 10 sec, annealing at 65 ºC for 10 sec, extension at 72 ºC for 30 sec and a final extension at 72 ºC for 7 min.

Following amplification, the BC1-attB PCR product was purified in order to remove any primer dimers prior to cloning into the entry vector (pDONR221). TE buffer (150 µl) at a pH of 8.0 was added to 50 µl of the amplification reaction containing the attB-PCR product. This was followed by the addition of 100 µl of 30 % PEG, 30 mM $MgCl_2$. The mixture was vortexed and immediately centrifuged at 10 000 xg for 15 minutes at room temperature. The pellet was dissolved in 50 µl of

TE, pH 8.0 to a final concentration of (>10 ng/ µl). The quality of the attB-PCR product was assessed by agarose gel electrophoresis.

The following recombination reactions were set up in order to ensure the cloning of the attB-PCR product into the pDONR™221 entry vector.

Table 4.1: Table of reagents for set up of recombination reactions for cloning into pDONR™221.

Components	Sample	Positive control	Negative Control
attB-PCR product (33 ng)	7 µl	-	7 µl
pDONR vector (150 ng/ µl)	1 µl	1 µl	1 µl
pEXP7- tet positive control (50 ng/ µl)	-	2 µl	-
TE Buffer, pH 8.0	2 µl	5 µl	2 µl

Two microliters of the Gateway® BP Clonase II enzyme mix was added to the sample and positive control tubes. The tubes were mixed well by vortexing, and the reactions were then incubated at 25 ºC for 1 h. This was followed by the addition of 1 µL of Proteinase K solution to all three reactions which were incubated for 10 min at 37 ºC.

The reactions were then transformed into competent DH5α *E.coli* using the One shot ® Chemical Transformation Protocol (Gateway® Technology with Clonase II, Invitrogen by Life Technologies) and selected on LB media + 50 µg/ ml kanamycin. Colonies were screened by colony PCR and successful colonies were cultured and stored for future transformation into competent yeast cells.

4.3. Results and Discussion

4.3.1. Selection of cassava bait proteins

Several studies have shown the involvement of histones in regulation of geminivirus replication and transcription (Hanley-Bowdoin *et al*., 2013). Transcriptome data from Allie *et al*., 2014 (table 4.2) showed a significant up-regulation in the Histone superfamily protein in SACMV-susceptible cassavaT200 landrace (at 32 and 67 dpi post SACMV infection), which is thought to exist in a complex with histones H2 and H4. Histone H3 as suggested by Kong and Hanley-Bowdoin (2002) has also been implicated in protein-protein interactions with the Rep protein. Geminiviruses replicate forming mini-chromosomes in the nucleus (Hanley-Bowdoin *et al*., 2013). The binding of Histone H3 to Rep displaces nucleosomes from viral DNA in order to allow access to the replication machinery and/or prevent the methylation of histone H3 (methylation impairs viral replication) this results in the removal of histones and facilitates replication of the geminiviruses (Hanely-Bowdoin *et al*., 2013). This would explain high virus replication in T200 resulting in disease (Allie *et al*., 2014). In geminiviruses, histone H3 binds to the MP and has been identified in the plasmodesmata of infected cells. This suggests that the viral DNA moves between cells together with the nucleosome (Hanley-Bowdoin *et al*., 2013). This suggests that DNA requires a complex with histone H3 which allows interaction with the MP which facilitates movement through the plasmodesmata into an adjacent cell, indicating that the DNA, histone H3 and MP travel in a complex from cell-to-cell. Studies establishing the binding of histone H3 to geminivirus MP were confirmed by Zhou *et al*., 2011. The study indicated binding of the MP of Bean dwarf mosaic virus (BDMV) to histone H3 of the host plant using gel overlay and co-immunoprecipitation experiments. The significant up-regulation of expression in members of the Histone superfamily in SACMV infected T200 (Allie *et al*., 2014) would favour replication of the viral genome. These and other studies suggest a high likelihood for a protein-protein interaction between cassava Histone H3 and SACMV BC1 movement protein in cassava, a non-model host.

While the transcriptome data (table 4.2) Allie *et al*., 2014 showed a significant up-regulation for the Histone superfamily protein, the data showed a significant down-regulation (both 32 and 67 dpi) for HSP70. HSP70 is a chaperone protein and assists in folding, solubilizing protein aggregates, the disassembly of protein complexes, membrane translocation and in the cell-to-cell transport of proteins. (Krenz *et al*., 2010). In geminiviruses HSP70 and a synaptotagmin protein (SYTA) bind to the MP. The downregulation of both these proteins restricts or delays infection. This indicates that geminiviruses employ host transport systems for their movement (Hanley-Bowdoin *et al*., 2013).

Various studies have shown indications of heat shock proteins, in particular HSP70, binding to geminivirus MP (Chen *et al*., 2008; Krenz *et al*., 2010; Lu *et al*., 2009; Mayer, 2010; Shimizu *et al*., 2009; Soellick *et al*., 2000). Y2H studies (Krenz *et al*. 2010) indicate that the C-terminus of HSP70 in Arabidopsis interacts with the N-terminus of the Abutilon mosaic virus (AbMV) BC1 MP. HSPs as chaperones or accessory folding proteins are widely reported in other plant-virus interactions, for example in RNA closteroviruses that code for a HSP70-like protein involved in movement. The role of HSP70 in cassava forming a complex with BC1 therefore appears highly probable, and therefore was selected as a candidate bait protein.

Table 4.2: Transcriptome data used for selection of bait proteins for protein-protein interaction studies for T200 (Allie *et al*., 2014).

Cassava ID	AGI Arabidopsis	Functional Annotation	32dpi Log2 fold change	32dpi p-value	67dpi Log2 Fold Change	67dpi p-value
cassava4.1_030637m.g	AT3G27360.1	Histone superfamily protein	2.331654292	0.010431719	1.865923186	0.040039458
cassava4.1_024615m.g	AT3G27360.1	Histone superfamily protein	2.428634261	0.040030394	3.258934379	0.017668884
cassava4.1_008936m.g	AT5G02500.1	Heat shock cognate protein 70-1	-1.634859014	0.034972279	-2.323668849	0.002855427
cassava4.1_003340m.g	AT3G12580.1	Heat shock protein 70	-1.54099209	0.035596361	-1.66926427	0.02935777

4.3.2. *In silico* studies of Histone H3 and HSP70

It was necessary to confirm predicted protein binding sites of the host proteins (Histone H3 and HSP70) for purposes of gene synthesis (to confirm if the entire gene should be cloned into pDEST22 or just the N terminus or C terminus). Confirming the possible binding sites using an in silico approach lends further support (apart from literature) that these host proteins may be suitable bait candidates to use downstream for Y2H protein-protein binding analyses.

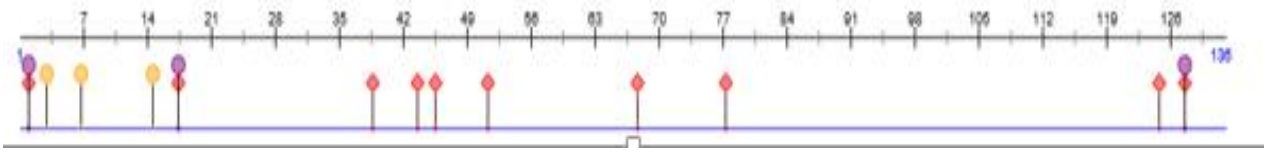


Figure 4.2: 'Predict Protein' image showing the predicted cassava Histone H3 protein binding sites (indicated by the red diamonds) and nucleotide binding sites (not relevant for this study) (of the transcript) (DNA indicated by yellow circles, RNA indicated by purple circles) (Source: Predict Protein, 2017).

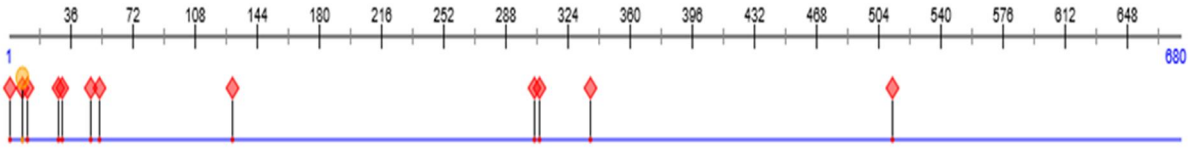


Figure 4.3: 'Predict Protein' image showing cassava HSP70 predicted protein binding sites (indicated by the red diamonds) and nucleotide binding sites (not relevant for this study) (of the transcript) (DNA indicated by yellow circles) (Source: Predict Protein, 2017).

4.3.3. Cloning of BC1 gene into pDONR™221 vector

The Y2H system is based on the ProQuest™ Two-Hybrid System (Invitrogen). This system requires Gateway® cloning technology and recommends recombination of an entry clone with the prey protein (protein of interest) prior to recombination into a pDEST32 vector (enables transformation into competent yeast cells). The donor vector pDONR™221 was used to construct an entry clone with the prey protein BC1. Figure 4.4 shows successful amplification of the BC1 gene flanked with the attB (also known as attL) specific primers. This clone (subsequent to purification) was used to construct the BC1-attB clone into pDONR™221 where the attL flanked clone is recombined with

attR sequences within the vector (figure 4.5). Figure 4.6 indicates the colonies (colonies grow if successful transformation into the *E.coli* takes place) which were cultured and subsequently stored after transformation of the entry BC1 clone into the entry vector. Y2H assays will be performed in studies in the future.

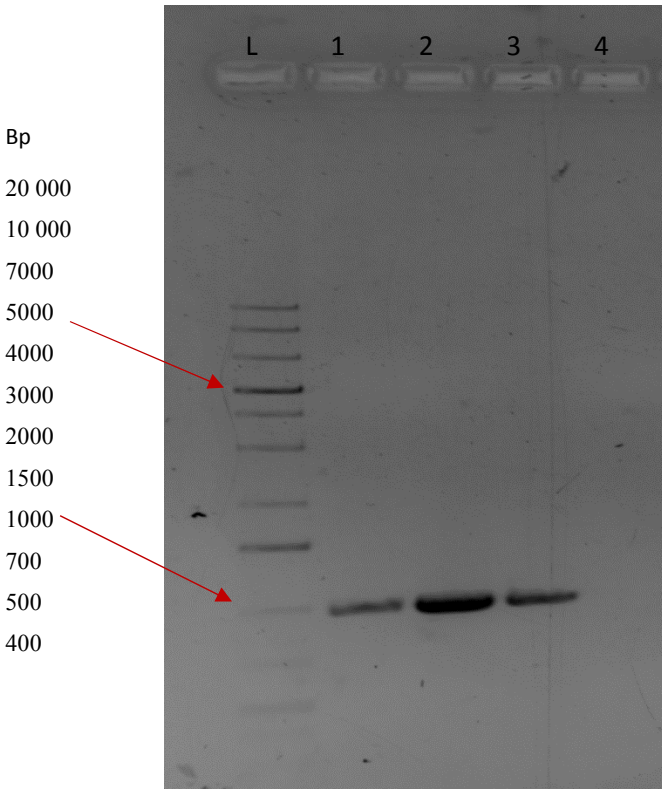


Figure 4.4: A 1% agarose gel showing PCR amplicons (BC1 gene amplified with BC1-attB primers at 970 bp) from PCR with BC1-attB primers. L: Ladder Generuler 1 KB plus (Thermo Scientific), Lanes 1, 2 and 3: PCR product with BC1-attB primers annealed at 65 ºC; 4, non-template control.

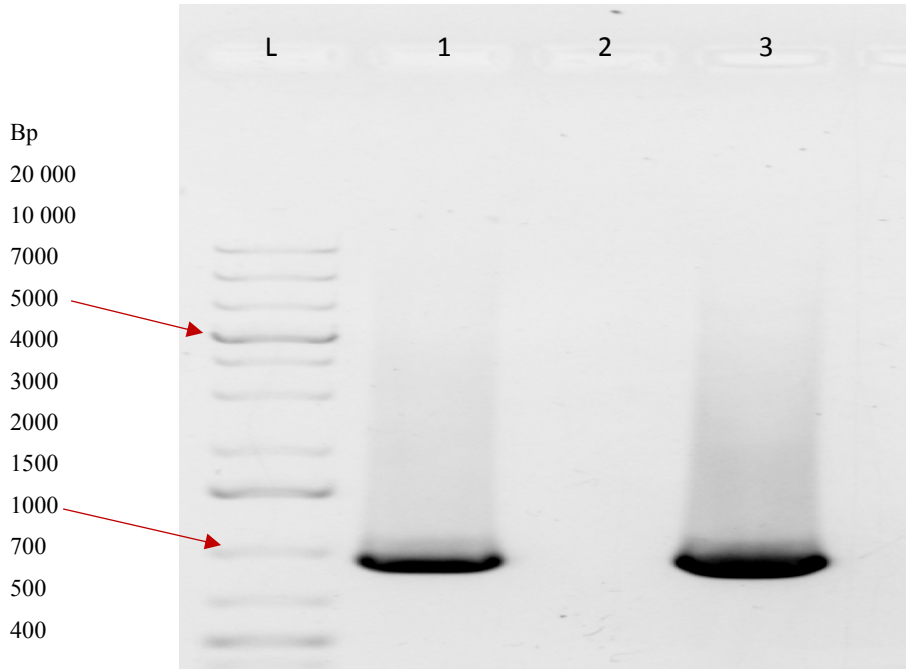


Figure 4.5: A 1% agarose gel showing the quality (single, visible band at ~970 bp) of the purified BC1-attB PCR products. L: Ladder Generuler 1 KB plus (Thermo Scientific), Lanes 1: Purified PCR product 1; 2: Empty; 3: Purified PCR product 2.

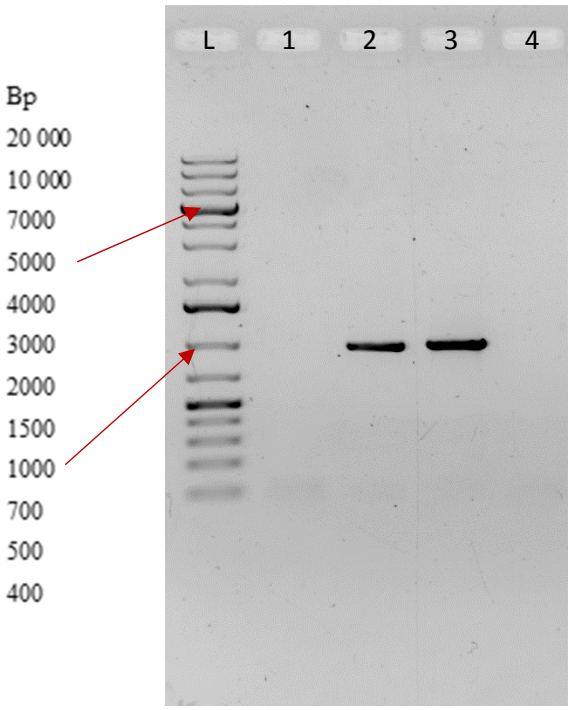


Figure 4.7: A 1% agarose gel showing the products of a colony PCR following transformation into competent DH5α cells. L: Ladder Generuler 1KB plus (Thermo Scientific), Lanes 1: Colony 1; 2: Colony 2; 3: Colony 3; 4: Negative Control.

4.4. Conclusions

While the goal is to experimentally determine protein-protein interactions between cassava host proteins and the cell-to-cell movement BC1 protein of SACMV employing the Y2H system, it is useful to select host proteins with the most likelihood of interaction with BC1. With the aid of both bioinformatics and experimental data from the literature, cassava host proteins Histone H3 and HSP70 were identified as the most suitable candidates for future Y2H protein-protein interactions or co-immunoprecipitation with SACMV BC1. To date, preparation has been done in selection and synthesis of the host candidates, and cloning into the pDEST22 vector for transformation into yeast cells for creation of the bait expression vectors. Expression of the prey (BC1) into yeast vector was achieved. Future work includes transforming both bait and prey plasmids into competent *S. cerevisiae* (MaV203) cells, and completion of the test controls and X-gal assays to determine the protein-protein interactions. In order to determine the validity of the results attained from the Y2H assay, a co-immunoprecipitation assay will be carried out with BC1 antibodies which will be generated through peptide sequencing and identification of an antigenic peptide signal. The completion of the Y2H assays will confirm or refute the hypothesis (based on literature) that cassava host proteins Histone H3 and HSP70 interact with the cell-to-cell movement protein of SACMV.

Concluding Remarks

A small number of structural studies of the N-terminus of begomovirus Rep have been undertaken, for example in tomato yellow curl Sardinia virus (Campos-Oliva's *et al*., 2002), however full-length geminivirus proteins remain recalcitrant to expression and purification due to size and hydrophobicity. While many functional studies have been performed on the nuclear shuttle protein (BV1) (Fontes *et al*., 2004) and cell-to-cell movement (BC1) proteins (Rojas *et al*., 1998), that can infer or predict secondary structure, little is known about the structure of these elusive proteins. BC1 in phloem-limited geminiviruses, such as Squash leaf curl virus, has 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 would possibly point to the highly insoluble nature of BC1. The purification of BC1 in the study proved difficult due to its insolubility, and likely due to its predicted unusual intrinsically unfolded regions on the C-terminus. This, in addition to other reasons, may explain why to date no in vitro expression or purification studies have been reported. It is interesting to note that that no proteins in any database could be found with any appreciable homology to BC1. Also noteworthy was the identity of amino acid sequence homology between SACMV BC1 and BC1 from EACMV.

Using Far-UV CD, a predominant secondary structure for BC1 has been established. While the secondary structure elucidated does not correspond to the *in silico* predicted secondary structure, the largely unfolded region of the protein presents speculation on the C-term of the protein. It was concluded that molecules such as ATP are able to bind to BC1 owing to the presence of hydrophobic pockets within the tertiary structure of the protein. The presence of a denaturing agent such as 8 M urea on native BC1 indicates conformational change upon ligand binding. While the molecular weight of BC1 was elucidated by SDS-PAGE, further confirmation of the size of the protein could be confirmed by size exclusion high pressure liquid chromatography (SE-HPLC), particularly of the native protein in comparison to the denatured protein. *In silico* and *in vitro* studies undertaken have suggested that a large region of BC1 is disordered. This disordered region is speculated to be on the C-term of BC1. Based on this speculation it would be significant to structurally characterize each of the three predicted domains (Fondong, 2013) of BC1 beginning with the N-terminus (pilot domain).

Movement proteins have been suggested to bind to host proteins in plant cells. While many protein-protein interaction studies have been undertaken two stand out. The suggested interaction between

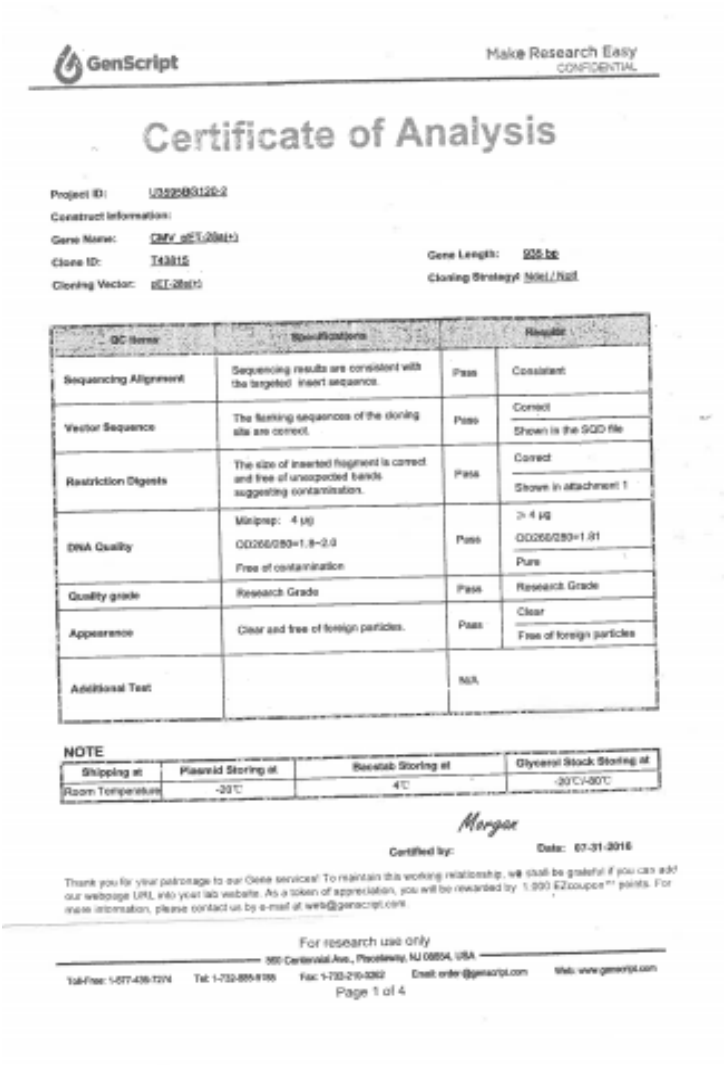
histone H3 and BC1 (Hanley-Bowdoin *et al*., 2013) and HSP70 and BC1 (Hanley-Bowdoin *et al*., 2013) indicate that these host proteins are of interest for protein-protein interaction studies. Completing the yeast two hybrid (Y2H) assays of which preliminary studies have been completed would be beneficial in confirming the binding of BC1 to host proteins histone H3 and HSP70. The confirmation of these protein-protein interactions would support previous studies which indicate binding of histone H3 (Zhou *et al*., 2011) and HSP70 (Krenz *et al*., 2010) to the geminivirus MP. Confirming protein-protein interactions between the viral and host proteins will corroborate functional studies on BC1 and the protein complexes it forms during viral replication (Hanley-Bowdoin *et al*., 2013).

Movement proteins (MPs) also need to bind to geminiviral DNA as most bipartite geminiviruses are transported without their coat protein, unlike plant RNA viruses (Hanley-Bowdoin *et al*., 2013), and therefore are predicted to have a DNA binding domain (reviewed in Fondong , 2013). Based on these findings it would be significant to carry out DNA-protein binding studies in the future.

In conclusion, the results of this research have contributed to the first glimpse of the secondary and tertiary structure of SACMV BC1 using *in silico* and *in vitro* approaches. This research has also identified possible host proteins which may interact with BC1. Further research is required to fully structurally characterize BC1. Further studies completing the Y2H assay will confirm or refute protein-protein interactions between BC1 and host proteins histone H3 and HSP70.

Appendix A

BC1- pET28a(+) :GenScript certificate



GenScript

Make Research Easy
CONFIDENTIAL

Certificate of Analysis

Project ID: US22060120-2

Construct Information:

Gene Name: CMV_pET-28a(+)

Clone ID: T43815

Cloning Vector: pET-28a(+)

Gene Length: 935 bp

Cloning Strategy: NdeI / NotI

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 SGD file
Restriction Digests	The size of inserted fragment is correct and free of unexpected bands suggesting contamination.	Pass	Correct Shown in attachment 1
DNA Quality	Miniprep: 4 μg OD260/280=1.8~2.0 Free of contamination	Pass	> 4 μg OD260/280=1.91 Pure
Quality grade	Research Grade	Pass	Research Grade
Appearance	Clear and free of foreign particles.	Pass	Clear Free of foreign particles
Additional Test		N/A	

NOTE

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

Certified by: Morgan Date: 07-31-2015

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 EZcoupce™ points. For more information, please contact us by e-mail at web@genscript.com

For research use only

860 Centennial Ave., Piscataway, NJ 08854, USA

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

Page 1 of 4

Histone H3 in pDEST22: GenScript Certificate

GenScript

Make Research Easy

CONFIDENTIAL

Certificate of Analysis

Project ID: U8141BJ070-6

Construct Information:

Gene Name: Histone H3_pDEST 22

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

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

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

NOTE

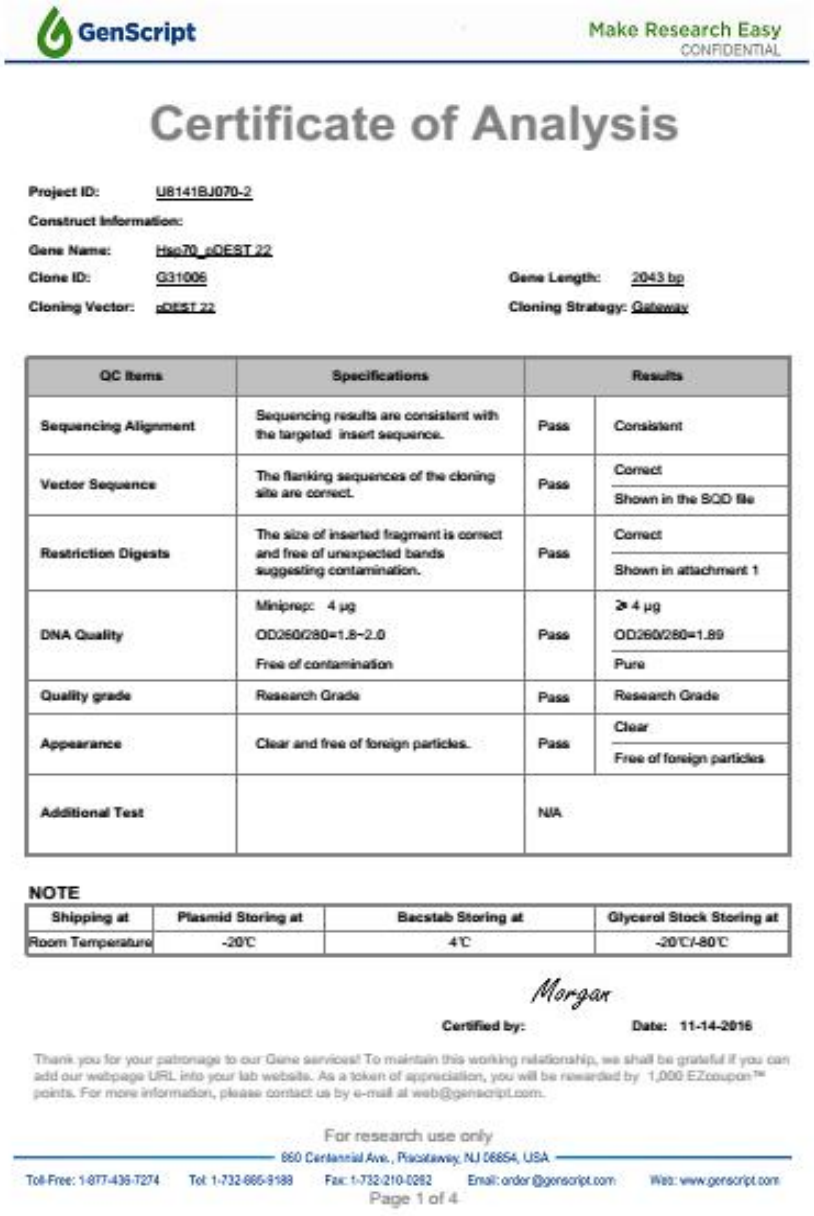
Shipping at	Plasmid Storing at	Bacstab Storing at	Glycerol Stock Storing at
Room Temperature	-20℃	4℃	-20℃/-80℃

Morgan

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

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

HSP 70 in pDEST22: GenScript Certificate



GenScript

Make Research Easy
CONFIDENTIAL

Certificate of Analysis

Project ID: U8141BJ070-2

Construct Information:

Gene Name: Hsp70_pDEST 22

Clone ID: G31006 **Gene Length:** 2043 bp

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

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

NOTE

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

Morgan

Certified by: **Date:** 11-14-2016

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

For research use only

860 Centennial Ave., Piscataway, NJ 08854, USA

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

Page 1 of 4

Appendix B

Template sequence for Histone H3:

ATGGCTCGTACAAAGCAAACGGCTAGGAAATCTACTGGAGGTAAGGCTCCGAGGAAGC
AGCTGGCACCAAGGCAGCGAGAAAATCTGCTCCGGCAACCGGTGGAGTAAAGAAACCC
CACAGGTTCAGACCTGGTACTGTGGCTCTAAGAGAAATCAGGAAGTATCAAAAGAGCA
CGGAGCTTTTGATCAGGAAACTTCCATTCCAGAGATTGGTGAGAGAAATCGCTCAAGAT
TTCAAGACTGACCTGAGATTCCAAAGCAGTGCTGTTGCAGCCTTGCAGGAGGCTGCTGA
GGCCTACTTGGTTGGGTTGTTTGAGGATACAAACCTCTGTGCTATTCACGCAAAAAGAG
TTACCATCATGCCTAAGGATATTCAGCTGGCACGAAGGATTAGGGGCGAGAGGGCTTAG

Protein sequence for Histone H3:

MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRFRPGTVALREIRKYQKSTE
LLIRKLPFQRLVREIAQDFKTDLRFQSSAVAALQEAAEAYLVGLFEDTNLCAIHAKRVTIMP
KDIQLARRIRGERA

Template sequence for HSP70:

ATGTCCACCGCAGCTCTCATCCGCTCGTTACGACGTCGCGACGTCGCCTTAGGCCCTGTT
TATGCTTACAAGTGTTTAAGTAGCAATTTGAAACCATCATGGTCCCCTTCTAACCTTGGT
TATAATTGGGCAAGTTTGTCAAGAGCATTCAGTGCAAAGCCGGCTGGCAATGATGTTAT
TGGTATTGACTTGGGTACCACCAATTCATGTGTTGCTGTCATGGAGGGGAAGAATCCCA
AAGTCATTGAAAATTCTGAAGGATCTAGGACCACCCCTTCAGTGGTTGCCTTCAACCAG
AAGGGAGAGCTACTTGTGGGTACTCCAGCAAAGCGGCAGGCTGTGACCAATCCAACCA
ACACTGTTTTTGGAACTAAGCGTTTGATTGGTAGGAAATTTGATGATCCTCAGACCCAG
AAAGAGATGAAAATGGTTCCATACAAGATAATAAGGGCACCAAATGGAGATGCATGGG
TTGAGGCCAATGGGCAGCAGTATTCACCTAGCCAGATTGGTGCCTTCATTTTGACAAAG
ATGAAAGAAACTGCAGAGTCTTACCTTGGAAAAACAGTTACAAAAGCTGTTATAACTGT
ACCGGCATATTTCAATGATGCTCAAAGACAAGCAACAAAGGATGCTGGGAGAATTTCTG
GCCTTGATGTGCAGAGAATCATTAATGAGCCTACGGCAGCTGCACTGTCCTATGGTATG
AACAACAAGGAGGGTCTAATAGCAGTTTTTGACCTTGGTGGTGGGACATTTGATGTTTC
AATTTTAGAGATCTCTAATGGTGTTTTTGAGGTGAAAGCCACAAATGGTGACACATTTTT
GGGAGGAGAGGACTTTGACAATGCACTGCTGGAGTTCTTAGTGAGTGAATTCAAGAGA
ACTGAGGGAATTGATCTAACAAAAGACAGGCTGGCTCTGCAGAGGCTTCGTGAAGCAG
CAGAAAAAGCTAAGATAGAGCTCTCATCAACAGCCCAGACTGAAATCAACTTGCCATTT

ATTACAGCTGATGCATCTGGTGCAAAACATCTAAACATTACTCTAACCAGATCCAAATT
TGAAAGTTTAGTAAATCACTTAATTGAAAGGACTAGGGATCCATGCAAGAATTGTTTGA
AAGATGCTGGCATATCCACCAAAGAGGTTGATGAGGTCCTCCTTGTTGGAGGAATGACC
CGTGTGCCCAAGGTTCAGGAAATTGTAGCAGAGATCTTTGGAAAAAGCCCGAGCAAAG
GAGTAAATCCTGATGAAGCAGTTGCTATGGGAGCTGCAATTCAGGGTGGTATTCTTCGT
GGGGATGTCAAAGAGCTGCTCCTTCTGGATGTCACTCCTCTATCCCTTGGTATCGAGACT
CTTGGTGGGATATTCACAAGGTTGATCAACAGGAACACAACAATTCCTACTAAGAAGAG
TCAGGTGTTCTCTACAGCAGCTGATAACCAAACACAGGTGGGTATCAAGGTGCTCCAAG
GTGAGCGTGAAATGGCATCTGACAACAAACTTCTGGGCGAGTTTGAACTCGTGGGTATT
CCACCAGCACCAAGGGGTTTGCCTCAGATTGAGGTGACATTTGATATTGATGCCAATGG
TATTGTCACTGTTTCTGCCAAGGACAAGTCCACTGGAAAAGAACAACAGATCACCATCC
GATCATCTGGTGGTCTTTCAGAAGATGAGATTGAGAAGATGGTTAAGGAGGCTGAGCAA
TTTGCCCAGAAAGATCAGGAAAGAAAGGCTTTGATTGACATAAAAAATAGTGCAGACA
CAACTATCTACAGCATTGAGAAGAGTTTGGCGGAATATAGAGACAAGATTCCTGCTGAA
GTTGCCAAGGAGATTGAGGATGCTGTTGCAGACTTGAGGAAGGCAATGGCGGGAGATA
ATATTGATGAGATCA
AATCCAAACTCGATGCTGCTAACAAAGCTGTGTCAAAGATTGGAGAGCACATGTCCAAG
GGCAGTGGTGGGGATGGTGCATCTGGTGGTTCCCAGGGTGGCGACCAGACTCCAGAAG
CGGAGTATGAGGAGGTCAAAAAGTGA

Protein sequence for HSP70:

MSTAALIRSLRRRDVALGPVYAYKCLSSNLKPSWSPSNLGYNWASLSRAFSAKPAGNDVIGIDLGT
TNSCVAVMEGKNPKVIENSEGSRTTPSVVAFNQKGELLVGTPAKRQAVTNPTNTVFGTKRLIGRKF
DDPQTQKEMKMVPYKIIRAPNGDAWVEANGQQYSPSQIGAFILTKMKETAESYLGKTVTKAVITV
PAYFNDAQRQATKDAGRISGLDVQRIINEPTAAALSYGMNNKEGLIAVFDLGGGTFDVSILEISNGV
FEVKATNGDTFLGGEDFDNALLEFLVSEFKRTEGIDLTKDRLALQRLREAAEKAKIELSSTAQTEIN
LPFITADASGAKHLNITLTRSKFESLVNHLIERTRDPCKNCLKDAGISTKEVDEVLLVGGMTRVPKV
QEIVAEIFGKSPSKGVNPDEAVAMGAAIQGGILRGDVKELLLLDVTPLSLGIETLGGIFTRLINRNTTI
PTKKSQVFSTAADNQTQVGIKVLQGEREMASDNKLLGEFELVGIPPAPRGLPQIEVTFDIDANGIVT
VSAKDKSTGKEQQITIRSSGGLSEDEIEKMVKEAEQFAQKDQERKALIDIKNSADTTIYSIEKSLAEY
RDKIPAEVAKEIEDAVADLRKAMAGDNIDEIKSKLDAANKAVSKIGEHMSKGSGGDGASGGSQGG
DQTPEAEYEEVKK

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. APS*net* 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.

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.

Chayen, N.E and Saridakis, E. (2008) Protein crystallization : from purified protein to diffraction-quality crystal. *Natural Methods,* 5: 147–153.

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).

Chen, Z., Zhou, T., Wu, X., Hong, Y., Fan, Z and Li, H. (2008) Influence of cytoplasmic heat shock protein 70 on viral infection of *Nicotiana benthamiana*. *Molecular Plant Pathology*, 9:809-817.

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 Lozano-Durán, 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 Virus-based Post Transcriptional Gene Silencing. *Viruses*, 5:998-1022.

Deepak, S.A., Ishii, H and Park, P. (2006) Acibenzolar-S-Methyl primes cell wall strengthening genes and reactive oxygen species forming/ scavenging enzymes in cucumber after fungal pathogen attack. *Physiological and Molecular Plant Pathology*, 69: 52-61.

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

Estrada, E and Peña, A. (2000) In silico studies for the rational discovery of anticonvulsant compounds. *Bioorganic & Medicinal Chemistry*, 8(12): 2755–2770.

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.

Feng, Y., Wang, K., Ma, C., Zhao. Y and Yin, J. (2015) Virus-induced gene silencing-based functional verification of six genes associated with vernalization in wheat. *Biochemical and Biophysical Research Communicators*, 458: 928-933.

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

Finn, R.D., Attwood, T. K., Babbitt, P.C., Bateman, A., Bork, A., Bridge, J., Chang, H., Daszt, Z., El-gebali, S., Fraser, et al (2017) Interpro in 2017- beyond protein family and domain annotation. *Nucleic Acids*, 45: 190-199.

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.

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.

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, E.-M. de Villiers, and H. zur Hausen, eds. (Heidelberg: Springer Verlag), pp. 185–226.

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.

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.

Kozlowski, L.P. (2016) IPC-Isoelectric point calculator. *Biology direct*, 11: 1-16.

Krentz, B., Jeske, H and Kleinow, T. (2012) The induction of stromule formation by a plant DNA-virus 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.

Lorber, B., Fischer, F., Bailly, M., Roy, H and Kern, D. (2012) Protein analysis by dynamic light scattering: Methods and techniques for students. *Biochem. Mol. Biol. Educ,* 40: 372–382.

Lu, L., Du, Z., Qin, M., Wang, P., Lan, H., Niu, X., Jia, D., Xie, L., Lin, Q and Wu, Z. (2009) Pc4, a putative movement protein of rice stripe virus interacts with a type I DnaJ protein and small Hsp of rice. *Virus Genes*, 38:320-327.

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.

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.

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.

Odronitz, F and Kollmar, M. (2007) Drawing the three of eukaryotic life based on the analysis of 2269 manually annotated myosins from 328 species. *Genome Biology*, 8: 196.

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.

Prajapat. R, Gaur. R. K and Raizada. R.G.V.K. (2010) In Silico Analysis of Genetic Diversity Of Begomoviruses Using Homology Modelling. *J. Biol. Sci*, 10: 1-7.

Radivojac, P., Clark, W., Oron, T and Schnoes, A. (2013) A large-scale evaluation of computational protein function prediction. *Nature*,10: 221–227.

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.

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

Roche, D.B., Buenavista, M.T., Tetchner, S.J and McGuffin, L.J. (2011) The IntFOLD server: An integrated web resource for protein fold recognition, 3D model quality assessment, intrinsic disorder prediction, domain prediction and ligand binding site prediction. *Nucleic Acids Res,* 39: 171–176.

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

Rost, B and Sander, C. (1994) Conservation and prediction of solvent accessibility in protein families. *Proteins:Structure, Function and Genetics*, 20: 216-226.

Rost, B and Liu, J. (2003) The PredictProtein server. *Nucleic Acids Res*, 31: 3300–3304.

Schlessinger, A., Liu, J., and Rost, B. (2007). Natively Unstructured Loops Differ from Other Loops. *PLoS Computational biology*, 3: 1335-1346.

Scholtz, J.M., Barrick, D., York, E.J., Stewart, J.M and Baldwin, R.L. (1995) Urea unfolding of peptide helices as a model for interpreting protein unfolding. *Proclaimed National Academy of Science. U. S. A,* 92: 185–189.

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.

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.

Shiba, K., Niidome, T., Katoh, E., Xiang, H., Han, L., Mori, T and Katayama, Y. (2010) Polydispersity as a parameter for indicating the thermal stability of proteins by dynamic light scattering. *Anal. Sci*, 26: 659–663.

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) *Germiniviridae*, 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.

Vanderschuren, H., Stupak, M., Fötterer, 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.

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.

Whelan, S and Goldman, N. (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Bol. Evol*, 18: 691–699.

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.

Whitmore, L and Wallace, B.A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res*, 32: 668–673.

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

Xu, X.M and Jackson, D. (2010) Lights at the end of the tunnel: new views of plasmodesmal structure and function. *Current opinion in Plant Biology*, 13: 1-9.

Malvern instruments (2004) *Zetasizer Nano Series User Manual*. Dep. Biochem. Biophys. Facil. , Univ. Cambridge 207.

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.