खीरा हरा मटमैला मोज़ेक वायरस के विनिर्माणित जीनोम का कार्यात्मक विश्लेषण

Functional analysis of deconstructed genome of cucumber green mottle mosaic virus

Mr. Anirudha Chattopadhayay

DIVISION OF PLANT PATHOLOGY

ICAR-INDIAN AGRICULTURAL RESEARCH INSTITUTE

NEW DELHI - 110012

2020

"Functional analysis of deconstructed genome of cucumber green mottle mosaic virus"

A Thesis

By

Anirudha Chattopadhyay

Submitted to the Faculty of the Post-Graduate School,
ICAR-Indian Agricultural Research Institute, New Delhi
In partial fulfillment of the requirements for the award of the degree of

Doctor of Philosophy in Plant Pathology

2020

Approved by the Advisory committee:

(Dr.Bikash Mandal)
Chairperson

(Dr.Anirban Roy)
Co-Chairperson

(Dr. Kumar Aundy)
Member

(Dr.Rakesh Kumar Jain)
Member

(Dr. Sanjay Kumar Lal)
Member

(Dr. Amolkumar U. Solanke)
Member

Dr. Bikash Mandal
Principle Scientist
Division of Plant Pathology

Division of Plant Pathology
ICAR-Indian Agricultural Research Institute,
New Delhi-110012

CERTIFICATE

This is to certify that the thesis entitled, **"Functional analysis of deconstructed genome of cucumber green mottle mosaic virus"**, submitted to the Faculty of the Post-Graduate School, Indian Agricultural Research Institute, New Delhi, in partial fulfilment of the requirements for the degree of **Doctor of Philosophy** is a record of bona fide research work carried out by **Anirudha Chattopadhyay. (Roll NO: 10166)** under my guidance and supervision, and no part of this thesis has been submitted for any other degree or diploma. It is further certified that all the assistance and help availed during the course of investigation as well as all sources of information have been duly acknowledged by him.

Date: 02-11-2020
Place: New Delhi

(Dr. Bikash Mandal)
Chairperson
Advisory Committee

My Parents and family

ACKNOWLEDGEMENT

*As a prelude to my thanks giving, at first I wish to thank **MY FAMILY MEMBERS,** for giving me strength, courage and confidence.*

*It is my proud privilege to express my deepest sense of respect and gratitude, sincere appreciation and heartfelt thanks to my eminent guide and the Chairman of my Advisory Committee **Dr. Bikash Mandal**, Principal Scientist, Division of Plant Pathology, ICAR-IARI New Delhi, for his noble guidance, constant encouragement, unfailing help, constructive criticism and valuable suggestions during the entire course of present investigation. He is a man of clarity of thoughts, dynamism and has immense knowledge of his field. He has enabled me to complete this study with his untiring effort and personal interest. He has enlightened the dark corners of my personality and it feels tough to find appropriate words to say thanks to him.*

*With utmost degree of sincerity, I express my heartfelt gratitude to my Co-Chairman **Dr. Anirban Roy,** Principal scientist, Division of Plant Pathology, IARI, New Delhi and **Dr. Kumar Aundy**, Principal scientist, Division of Plant Pathology, IARI New Delhi for providing his valuable suggestions as a member of my advisory committee and kind help in many ways to me during my Doctor of Philosophy and without him this thesis, too, would not have been completed.*

*I am very much thankful to **Dr. Rakesh Kumar Jain,** Emiratis scientist, Division of Plant Pathology, for providing his valuable suggestions as a member of my advisory committee and kind help in many ways to me during my Doctor of Philosophy and without him this thesis, too, would not have been completed.*

*My sincere thanks also goes to **Dr. Sanjay Kumar Lal**, Principal Scientist, Division of Genetics, IARI, New Delhi, member of my advisory committee for his kind help during my research work.*

*My sincere thanks also goes to **Dr. Amolkumar U. Solanke**, Principal Scientist, Division of Molecular Biology & Biotechnology, IARI, New Delhi, member of my advisory committee for his kind help during my research work.*

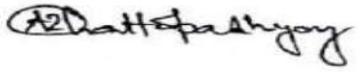
*I express my sincere sense of gratitude & heartfelt thanks to **Dr. Rashmi Agrawal** Head Division of Plant pathology, ICAR-IARI, New Delhi and **Dr. V.K. Baranwal,** Professor,*

Division of Plant pathology, ICAR-IARI, New Delhi for their laudable resplendence in providing guidance from time to time and affection during the whole course of my study.

I will ever remain indebted for the facilities and knowledge, ***IARI*** *provided to me. I express my sincere thanks to the* ***Indian Council of Agricultural Research*** *for providing me the financial assistance in the form of PhD scholarship.*

The kind of unending love, moral supports and the ever extending help provided to me by my dear lab mates in any case cannot be expressed by words. I am immensely obliged to ***Dr. Abdul, Dr. Dipinti, Mr. Vikas, Mr. Anand, Miss. Ankita, Dr.Nagamani, Mr. Pradeep, Mr. Oinam Washington, Mr. Naveen Nayaka, Mr. Suryakant, Mr. Sanjeet, Mr.Vijay and Dinesh ji.*** *I am also heartily thankful to all dear juniors;* ***Sajad, Mukesh, Kuleshwar, Vankatesh, Asha, Aditya, and all.***

Omission, if any in this brief document of acknowledgement does not imply ingratitude.



Dated: 02-11-2020 (**Anirudha Chattopadhayay)**

Place: IARI, New Delhi

ABBREVIATIONS AND SYMBOLS

+	Plus
-	Minus
+ve	Positive
-ve	Negative
%	Per cent
1^0	Primary
2^0	Secondary
3^0	Tertiary
°C	Degree centigrade
µg	Microgram
µl	Microliter
µM	Micromolar
A260	Absorbance at 260 nm
A280	Absorbance at 280 nm
AA/aa	Amino acid
Acetosyringone	3, 5-dimethoxy-4-hydroxy-acetophenone
ALSV	*Apple latent spherical virus*
APS	Ammonium per sulphate
ATP	Adenosine 5'triphosphate
BaMV	*Bamboo mosaic virus*
BCIP	5-Bromo-4-chloro-3-indolyl phosphate
BLAST	Basic Local Alignment Search Tool
BM	Bikash Mandal
BMV	Brome mosaic virus
bp/kb	Base pair/kilo base pair
BSA	Bovine serum albumin

$CaCl_2$	Calcium carbonate
cDNA	Complementary Deoxyribonucleic acid
cfu	Colony forming unit
CGMMV	Cucumber green mottle mosaic virus
cm	Centimeter
CMV	Cucumber Mosaic Virus
CP	Coat protein
CR	Coding region
C terminus	Carboxy terminus
CV3	Cucumber virus 3
CV4	Cucumber virus 4
CVP	Chimeric virus particles
DAS-ELISA	Double antibody sandwiched enzyme linked immunoassay
DEPC	Diethyl pyrocabonate
DG	Deconstructed genome
DMF	N, N-Dimethyl formamide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxyribonucleotide triphosphate
dRNA	Deconstructed Ribonucleic acid
dRT-V	dRNA-based TMV vector
Dpi/dpi	Days post inoculation
ds	Double stranded
DTT	Dithiothreitol
E. Coli	Escherichia coli
EDTA	Ethylene diamine tetra acetic acid
et al.	and co-workers/and others

G	Gram
GFP	Green fluorescent protein
gRNA	Genomic Ribonucleic acid
H	Hour
hGH	human growth hormone
HCl	Hydrochloric acid
Hlase	Helicase
RT-PCR	Reverse Transcription Polymerase Chain Reaction
IRES	Internal ribosome entry site
IPTG	Isopropyl β-D-thio galactopyranoside
K	Kelvin
kDa	Kilo Dalton
Kb	Kilo base
KGMMV	Kyrui green mottle mosaic virus
L	Litre
LA	Luria agar
LB	Luria broth
lic-LF	Lichenase-fused lethal factor
M	Molar
MES	2-(N-morpholino)ethanesulfonic acid
Mg	Milligram
$MgCl_2$	Magnesium chloride
min	Minute(s)
ml	Millilitre
mM	Millimolar
MP	Movement protein
MTase	Methyltransferase
MW	Molecular weight
N	Normality
N terminus	Amino terminus
NaOH	Sodium hydroxide
NBT	Nitroblue tetrazolium

NCBI	National Center for Biotechnology Information
ng	Nanogram
Nm	Nano meter
nt	Nucleotide
NTR	Non translated region
OAS	Origins of assembly
OD	Optical density
ORF	Open reading frame
PBST	Phosphate buffer saline tween20
PCR	Polymerase chain reaction
PDS	Phytoene desaturase
PEG	Polyethylene glycol
pk	Pseudoknot
pmol	Picomoles
PRRSV	Porcine reproductive and respiratory syndrome virus
PVP	Polyvinylpyrrolidone
PVX	Potato virus X
PTGS	Post transcriptional gene silencing
qPCR	Quantity polymerase chain reaction
RdRp	RNA dependent RNA polymerase
REC	Replicase enzyme complex
Reps	Replication protein
RNA	Ribonucleic acid
RNAi	RNA interference
RNase	Ribonuclease
rpm	Revolution per minute
RT	Room temperature
RT-PCR	Reverse transcription polymerase chain reaction
Rz	Ribozyme
Sec	Second(s)
SFV	Semliki Forest virus
SGP	Subgenomic promoter

SLC	Stem- loop C
sRNA	Small Ribonucleic acid
sgRNA	Subgenomic Ribonucleic acid
ss	Single stranded
SYCMV	*Soybean yellow common mosaic virus*
TAE	Tris-Acetate- EDTA
TE	Tris-EDTA
TLS	Transfer RNA–like structure
TMV	*Tobacco mosaic virus*
TORBO	TMV-based transient expression vectors
Tris	Tris (hydroxymethyl) aminomethane
TVCV	*Turnip vein clearing virus*
TYMV	*Turnip yellow mosaic virus*
U	Units
UTR	Untranslated region
UV	Ultra violet
V	Voltas
VIGS	Virus induced gene silencing
v/v	Volume/volume
w/v	Weight/volume
Xg	Time gravity
X-Gal	5-bromo-4-chloro-3-indoyl- β-D-galactoside
βME	Beta mercaptoethanol

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

For a long time, plant viruses are considered as a deadly pathogen causing huge economic loss. Their evil nature makes them a biological culprit. Beyond this, these plant viruses can also be explored and exploited as the biological toolkit especially for plant molecular farming and plant genome editing. The utilisation of plant viruses for the useful purposes is the emerging trend for the production of novel biologics like antibiotics, proteins, vaccines, etc. So far, various protein expression systems, *viz.*, bacteria, yeast, baculovirus mediated insect cell-line, and mammalian cell line, etc. are available. Of them, the plant-based system is gaining importance, over other protein expression systems, due to the less chance of contamination of human pathogen, low cost, and lesser time processing time. Further, easy delivery of edible vaccine through raw food/fruits, and long term storage and transportation without cold chain facility are the added advantages.

Initially, *in planta* expression of pharmaceuticals were made in the stably transformed, genetically modified plants. The development of transgenics is a highly time and huge money intensive programme. Moreover, enforcement of strict biosafety regulation with serious environmental and social concerns over the use of GM technology has become the major stumbling block in this regard; thus limits their commercial application. To overcome this issue, a novel transient expression system was developed through the use of replicable expression vectors. Utilizing plant viruses as the 'expression vector' is one such technology that circumvents many of these hurdles; this makes it easier to utilize plants as the biofactory for manufacturing pharmaceuticals. Furthermore, the risk of cross-contamination/recombination in virus and permanent change in the plant genome will be avoided which is a serious problem in case of stable transgenic expression. Hence, nowadays, the plant virus genome mediated transient expression system is utilized as the alternative to transgenic strategy. Its smaller size with rapid replicability and translational ability helps in the efficient expression of foreign proteins in lesser time and cost, even in the wider hosts. The virus-based transient gene expression system has got a lot of attention due to the high level of protein expression in a very short period. Initially, the virus vector-based foreign gene expression was tested in a bacterial system with bacteriophage as a viral

vector. Virus vectors have also been used in animal systems with double-strand DNA virus, single-strand RNA based retrovirus, etc. for transient expression of different foreign proteins. After the discovery of the cauliflower mosaic virus with a double-stranded DNA genome in 1960, the transient expression of foreign genes within the plant system was initiated (Shepherd, 1989). Later, RNA viruses were also used as virus-vector in many host plants after the development of reverse transcription technology which permited the cDNA preparation from RNA viruses and *in vitro* transcription systems (Ahlquist *et al.*, 1984; Boyer *et al.*, 1994). Easy *in planta* delivery of infectious clone viral DNA or cDNA through either mechanical or *Agrobacterium* based inoculation has the practical advantage (Grimsley *et al.*, 1987; Leuzinger *et al.*, 2013). These technologies are very rapid and robust; lack of the requirement of sophisticated equipment makes them amenable for commercialization. Hitherto, many plant viruses containing DNA or RNA genome has been drawn as the efficient expression vector. Among them, ssRNA viruses are most common. The major criteria for using a plant virus as an effective vector depends on its ability of infection, efficient replication, and stable expression of foreign genes; for this suitable delivery system should also be available for rapid infection, high levels of replication, and gene expression in a broad host range. All these virus functions need the absence of substantial interference/defence from the host plant. The successful expression of foreign genes in the plant system also depends on the biology and behaviour of virus used for vector construction.

For the expression of foreign proteins, we follow two strategies *viz.*, 'full' virus vector strategy, and 'deconstructed' virus strategy. In full virus genome-based strategy, the gene of interest is inserted within the coding frame of the full virus genome without eliminating any part. Thus foreign proteins can be expressed along with functional virus, as the fused proteins. But low stability and lower expression of foreign proteins with restricted cargo-capacity make this strategy difficult to apply on the industrial scale. To avoid such a problem, deconstructed virus genome-based-vectors are developed. Herein, the virus genome is redesigned in such a way that it will retain only the regulatory parts essential for replication and translation; and the rest of the other gene elements necessary for various viral functions, *viz.*, symptoms, transmission, movement, etc. are eliminated. Thus, it always requires the support of a helper virus for genome functions like replication, translation, etc. The stable and higher efficient

expression of large and multiple foreign proteins are the advantages. Furthermore, this shorter molecule is very easy to engineer, and no biosafety issues can be imposed due to the loss of virus function. This strategy is used to develop deconstructed virus vectors using the genome of some popular viruses. For instance, magnICON, dRT-V, TORBO vectors are successfully developed using the tobacco mosaic virus (TMV) genome. But the commercial production of recombinant proteins using dRT-V, TORBO vectors (TMV-based vector) is restricted within the solanaceous host. To increase the host range, magnICON is designed by combing the genomic elements of TMV and turnip vein clearing virus, a crucifer infecting tobamovirus.

To design for other plant host systems, cucumber green mottle mosaic virus (CGMMV), a member of genus *Tobamovirus* having single-stranded (+ sense) monopartite RNA genome has been exploited by different researchers. The beauty of this virus is that it infects a larger number cucurbits and produces very mild symptoms, and replicates in a large number; thus can be a suitable option for deploying it as the biological toolkit for the cucurbitaceous crops. The CGMMV has a single-stranded positive-sense, monopartite RNA genome (6.4 kb) with four overlapping open reading frames (ORFs) that encodes four proteins *viz*., two replication proteins (129Kd and 186Kd), one 30Kd movement protein (MP), and 17-18Kd coat protein (CP). The 5´termini of its genome plays a key role in the expression of replicase enzymes, whereas the two 3´ co-terminal subgenomic RNAs (sgRNAs) express MP and CP (Dombrovsky *et al.*, 2017). The sgRNAs are naturally produced in the infected cell due to erroneous replication mechanism and are shorter than their cognate genomic RNAs coterminal with the 3′genomic part, sometimes with the 5′ regulatory sequences. The production of sg-mRNAs is one of the strategies employed by eukaryotic RNA viruses for the quantitative and temporal expression of proteins. This strategy can also be useful for heterologous expression of other proteins using CGMMV based expression vector. As yet, limited attempts have been made to utilize CGMMV as the expression vector. Initially, the full-length genome of CGMMV was used for expressing smaller sized (~100 to 150 nt long) foreign genes. Later, Zheng *et al.* (2015) had attempted to express GFP (~800 nt) but got weak expression. In our laboratory (Jailani et al 2017), the full-length genome of CGMMV was successfully utilized for expressing GFP and obtained 45 fold expression upto14 days post-inoculation (DPI), thereafter declined significantly when GFP was inserted at the end of the CP gene. But, once it was inserted at 105 nt

of CP, its expression reached up to 243 fold compare to host actin; but remain stable only for a 5-10 DPI period. The difficulty in the insertion of large foreign gene and their unstable expression is the major stumbling block for utilizing full-length viral genome based vector. Furthermore, the lack of suitable restriction sites with undefined regulatory elements within the CGMMV genome and difficulty in loading multiple foreign genes limits its wider utilization.

Thus, to avoid such limitation of unstable expression and insertion of large and multiple foreign genes, we are here aiming to design a new expression system using CGMMV. The hypothesis is that (1) the deconstruction of the CGMMV genome will lead to the development of shorter replicons that will be replicated with the help of the main virus; (2) This short replicon can be used as the two components vector system. For this, a short deconstructed genomic RNA component of CGMMV would be created by keeping its minimal regulatory sequences and its biology and behaviour, i.e., *in planta* replication, movement, persistence, etc. should be studied before using them as the potential expression vector. To generate such information, the research programme was formulated as the **'Functional analysis of deconstructed genome of cucumber green mottle mosaic virus'** with the following objectives:

1. To create a deconstructed genome of CGMMV with a *cis*-regulatory sequence.

2. To evaluate the trans-replicability and translation ability of the deconstructed genome in the presence of wild type CGMMV

CHAPTER 2 REVIEW OF LITERATURE

2.1. Cucumber green mottle mosaic virus

Cucumber green mottle mosaic virus (CGMMV) is considered as a predominant representative of the genus *Tobamovirus*, appertaining to family *Virgaviridae*, which also includes *Tobacco mosaic virus* (TMV) as the type species. Among six different cucurbit-infecting tobamoviruses, CGMMV is most detrimental to many cucurbits, *viz.*, cucumber (*Cucumis sativus*), bottle gourd (*Lagenaria siceraria*), melon (*Cucumis melo*), watermelon (*Citrullus lanatus*), etc. It affects the foliage and fruit, and typically produces green mild mottle, mosaic symptoms on young leaves and blistering, malformation of fruits with rotten, spongy, discoloured internal tissue (Komuro *et al.*, 1971). The higher percentage of indigent quality fruits from the infected plants become unmarketable, resulting in gigantic crop losses. Further, long persistence on seed, soil, and plant debris as the highly stable particle makes it difficult to manage and expedites their long-distance movement (within and between countries) through contaminated seeds. Since its first symptomatic description on cucumber in 1935 from England, is reported to infect a larger number of plant hosts including many cucurbits (bottle gourd, watermelon, cucumber, zucchini, pumpkin, squash, gourds, bitter gourd, melons, etc.) along with other diverse plant families, *viz.*, *Amaranthaceae*, *Chenopodiaceae*, *Euphorbiaceae*, *Iridaceae*, *Polygonaceae*, *Solanaceae*, etc., as the asymptomatic carrier. For a long time, CGMMV is known from Europe (Celix *et al.*, 1996) and Asia (Inouye *et al.*, 1967); then it is distributed globally. Up to 1985, CGMMV spreads slowly but gradually accelerates in between 1986-2006, and rapidly escalates in the last decade (after 2007), with the world trade liberation and agro-ecological intensification (Dombrovsky *et al.*, 2017). The wide-scale distribution of this disease within countries where it was already known and the concurrent report from other countries on the continental scale is the clinching evidence of that.

Afterward, CGMMV became economically far-reaching, a considerable threat to cucurbit industries, and the sudden epidemic emergence in Israel during 2007 (Reingold *et al.*, 2013), and in Australia (Kehoe *et al.*, 2017) and North America (Li *et al.*, 2015) during 2013–2016 implies this. Many agro-based industries including seed companies, seedling nurseries, and commercial cucurbit producers, retailers are highly affected with colossal financial loss (Reingold *et al.*, 2016; Dombrovsky *et al.*, 2017).

Contaminated seeds are the primary source of inoculum that initiate the disease in the field and also act as the leading agents for long-distance trans-boundary movement of CGMMV. With infected seeds, the virus can be moved to countries that are previously free from the disease, and sometimes appear in the deadly form to cucurbit industries. After the recent introduction to Australia, CGMMV becomes havoc and has destroyed the entire water-melon industry in the Northern Territory, of about AUD60 million worth (Keane, 2015), with imposing huge biosecurity risk to the $90 million melon industry of the neighbouring Queensland state as well as Western Australia (Coates and McCarthy, 2015). To avoid further spread, regular monitoring and the accurate diagnostic system should be developed to choose the healthy seedlings for plantation, and further virus-free seed production. The virus-free seeds will prevent the unsought introduction to new territory. Moreover, the phytosanitary measures should also be reinforced to prevent dispersal to new countries *via* international seed trade.

2.2. Genome organization of CGMMV in comparison with other Tobamoviruses

Resembling distant tobamoviruses, CGMMV owns single-stranded, positive (+ve)-sense (6.4 kb) RNA genome, encapsidated with coat protein (CP) to form a rigid rod-shaped particle of approximately 300 nm ×18 nm size. Its genome contains four overlapping open reading frames (ORFs) that encodes four different proteins (**Fig. 2.1)**. The first two ORFs are located at the 5′ end of genomic RNA (gRNA) that directly encode two functional proteins, *viz.*, 129 and 186 kDa protein, respectively (Ugaki *et al.*, 1991). Of them, N-terminus and C- terminus possess methyltransferase (MTase) and helicase (Hlase) domain at smaller 129 kDa protein (P129), where the larger 186 kDa protein (P186) expressed occasionally *via* the read-through of leaky termination codon (UAG) of P129, contains an RNA dependent RNA polymerase (RdRp) domain (Li *et al*., 2015). Both the proteins oligomerize to form the complete replicase enzyme. Although, the last two ORFs are expressed through the production of two sub-genomic RNAs (sgRNAs) that are translated into two proteins, i.e., movement protein (MP) and coat protein (CP). Other than the coding regions (CRs), the virus genome also contains two specific untranslated regions (UTRs) at the 5′ and 3′ terminal end (Dombrovsky *et al.*, 2017). The significance of each component is discussed below:

2.2.1. Untranslated regions (UTRs)

The 5′ UTR is ∼60 nucleotides (nts) long and prolific in residues (i.e., A and C). The 3′ UTR is ∼176 nts long and carries transfer RNA-like structure (TLS) as well

as three upstream pseudoknot like (pk) structures (Ishibashi and Ishikawa, 2016). The tobamovirus genome also contains 5′-m7G-cap at its 5′-terminal. Besides that, for translation as well as complementary negative (-ve) strand RNA synthesis gRNA serve as a template. Therefore, enhanced translation can be seen due to the presence of some elements at the 5′- and 3′- UTRs of the gRNA. Likewise, the 3′-UTR too accommodates the elements essential for the synthesis of (-ve) strand RNA. Later on, investigations are carried out by utilizing a cell-free system / *in vitro* system for virus replication and translation (Osman *et al.*, 2000; Dreher and Miller, 2006). These experimentations concede that only 70 nt stretch the 5′-terminal end of the genome is the cotranslational binding site for replicase enzyme complex (REC) of TMV which is translated from gRNA and further its binding drives the gRNA to RNA replication pathway. The machinery shows that replication proteins are the elite component for *cis*-preferential replication of TMV. It is also reported that further translation can be inhibited through this binding to refrain a lethal ribosome-RNA polymerase encounter, which can only proceed when replication and translation ensue synchronically over sole gRNA molecule (Chujo *et al.*, 2015). Thus, manifold substantial roles played by 5′- and 3′-UTRs in the life cycle of tobamovirus.

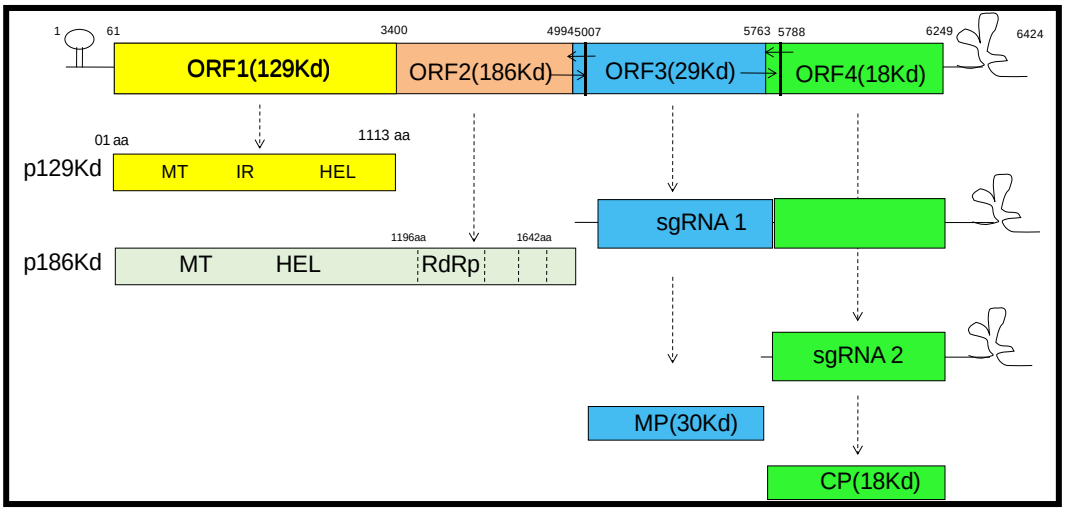


Fig. 2.1. The genome organization of CGMMV showing four overlapping open reading frames (ORFs).The ORF1 and ORF2 are located at the 5′ end of genomic RNA (gRNA) encodes two functional proteins, *viz.*, 129 and 186 kDa protein, respectively. The expression of 129 kDa protein is much higher than 186 kDa which is expressed occasionally *via* the read-through of leaky termination codon (UAG) of 129kDa. Whereas, the last two ORFs (ORF3 and ORF4) are expressed through the production of two 3'-co-terminal sub-genomic RNAs (sgRNAs) that are translated into movement protein (MP) and coat protein (CP), respectively.

2.2.2. Replicase enzyme

The replicase enzyme of tobamoviruses is a complex protein, containing 129kDa and 186 kDa proteins, which are oligomerized at 1:1 to 5:1 ratio to shape a replicase complex. Although both these proteins (129kDa & 186 kDa) are essential for the potent replication, and increasing concentration of 129 kDa protein in the replicase complex is necessary for the efficient replication. Thus, the expression level of 129kDa is much higher (>10 times) than that of 186 kDa protein (Ugaki *et al.*, 1991). This 126 kDa protein contains N terminus-MTase and a C-terminus Hlase domain that are separated by an intervening region (IR). The MTase domain possesses two indispensable functions, i.e. guanine-7-methyltransferase and guanylyl-transferase, that are necessary for RNA capping which in turn confers the stability of RNA replicons. However, Lecoq and Desibiez (2012) reported that P129 is essential for replication of CGMMV genome, and also as an inhibitor of post-transcriptional gene silencing (PTGS).

The formation of the 5′ cap in RNA replicon of tobamoviruses requires the sequential actions of these two enzymes. The MTase catalyses the transfer of methyl group to N^7 position of GTP from S-adenosyl methionine (AdoMet), a methyl donor, and yields 7-methyl-GTP (m7GTP) and S-adenosyl homocysteine (AdoHcy) (Dunigan and Zaitlin, 1990). Interaction of P126 with m7GTP forms a complex 7-methyl-GMP-p126 in presence of AdoMet, from which 7-methyl-GMP is transferred to the 5-diphosphate terminus of the viral RNA to generate the cap-like structure (Merits *et al*., 1999). Thus, the cap formation of viral RNA (gRNA & sgRNA) in tobamoviruses is distinct from that of eukaryotic cellular systems, where GMP has methylated later the trans-guanylation reaction and cap formation.

In spite of eloquent evolutionary divergence, the members of the alphavirus-like superfamily, including many animal viruses family (*Togaviridae*, *Hepeviridae*), and plant viruses family (*Bromoviridae*, *Virgaviridae*, *Alphaflexiviridae* and *Tymoviridae*) follow alike virus-specific capping mechanism (Merits *et al.*, 1999). In alphavirus-like super-family, the MTase is a membrane-bound enzyme (Laakkonen *et al.*, 1994) that can directly methylate GTP and dGTP but not capped RNA molecules (Merits *et al.*, 1999). The events of capping largely depend upon two factors, *viz*., (1) conservation of some critical residues that are indispensable for both MTase and guanylyl-transferase reactions in the alphavirus-like superfamily, and (2) structural configuration of the MTase protein. There are certain amino acids (aa), *viz*., histidine

(H), aspartate (D), arginine (R) and tyrosine (Y) are found to be highly conserved among alphavirus-like MTases. Through specific mutation, their significance was deciphered. For example, a single amino acid Histidine located at the 38th position (H^{38}) in nsP1 protein of Semliki Forest virus (SFV), 68th position (H^{68}) in 155-kDa protein of *Bamboo mosaic virus* (BaMV), of 81st position (H^{81}) in p126 protein of TMV, are specific for the guanylyltransferase, and MTase activity (Huang *et al*., 2004). Certain other critical residues, *viz*., D64 and D90 of nsP1 protein of SFV were identified to be necessary for AdoMet binding (Ahola *et al.*, 1997), whereas D122, R125, and Y213 in the155-kDa protein of BaMV are substantial for the catalytic activity of capping enzyme (Huang *et al*., 2004). Further, the putative secondary structure-based comparison of viral and cellular MTases also depicts the position of critical aa residues at the active sites of the functional domain. The critical residues in MTase domain of nsP1 protein of SFV are analogous to AdoMet-binding sites of cellular MTases, suggesting their structural and functional relatedness (Ahola *et al.*, 1997). It also indicates corresponding RNA capping mechanism acquired by the different members (including animal and plant infecting RNA viruses) of Alphavirus-like superfamily.

Besides MTase, the P129 of tobamoviruses contains superfamily 1 helicases that carry conserved RecA-like α/β domains as found in another superfamily 1 helicases (Nishikiori *et al.*, 2012). Hlase enzyme acts like a motor protein that can separate the duplex strands of nucleic acid through hydrolysis of energy (ATP). The viruses with single-stranded genomes including tobamoviruses harness its Hlase for the ATP dependent unwinding of intra-molecular secondary structures of gRNA (Lazareva *et al.*, 2015) or to separate inter-molecular RNA duplexes/RNA intermediates (containing viral plus and minus strands) during genome replication (Frick and Lam, 2006). Besides that, tobamovirus Hlase contains five conserved motifs (I, II, III, V, VI), *viz*., the motif for nucleotide triphosphatase activity (for NTP binding), the motif for oligomerization of 126kDa and 186kDa protein to form replicase complex, etc. Furthermore, several other enzymatic activities were also reported for RNA Hlases, such as non-energy-dependent RNA chaperone, and viral RNA stripping proteins (RNPase) (Lazareva *et al*., 2015). Besides its pivotal role in RNA replication, the Hlase have suppressor activity on host-mediated RNAi defenses and functions in the symptom development (Kubota *et al.*, 2003, Ding *et al.*, 2004, Padmanabhan *et al.*, 2005; Wang *et al.*, 2010). Furthermore, RNA binding activity of replicase also lies in the Hlase domain and its associated intervening region. The functional disruption of

motif V in the tobamovirus Hlase affects their ATPase hydrolysis and RNA binding activity (Wang *et al.*, 2010). Whereas, some other reports indicate the localization of RNA binding site in the intervening region (located within 314-423aa of TMV) of 126kDa protein. Thus different domains of 126 kDa protein are indispensable for the RNA binding, unwinding of RNA duplex, and further continuation of RNA replication in tobamoviruses (Kawamura-Nagaya *et al*., 2014). Similarly, the RdRp (186 kDa) of tobamoviruses replicase enzyme can recognize the virus templates in *trans*, but its replication requires some minimal regulatory sequences that act as the *cis*-preferential domain. Of them, some information on the regulatory sequences located at 5′ and 3′ UTR are already available. But the 'internal' *cis*-regulatory element(s) located within 126 kDa protein (essential for efficient replication) or other ORFs (essential for efficient replication/translation) are yet to be explored.

2.2.3. Coat protein and movement protein

The sgRNAs helps in translation of ORF's 3 and 4 which encodes MP and CP of 29kDa and 17-18kDa, respectively (Lecoq and Desibiez, 2012). Primarily MP involves in activities like short distance cell-to-cell movement *via* plasmodesmata and active transfer of viral nucleic acid to the nearby cell for localised systemic spread of viruses in a single leaf of the plants (Simon-Buela and Garcia-Arenal, 1999). Whereas, CP plays multifunctional roles including the long-distance spread of viruses to the entire plant system but still very less information retrieved regarding its long-distance movement through phloem cells.

2.2.4. Comparative analysis of tobamovirus genome organization

CGMMV has been first reported by Ainsworth (1935) in Great Britain, *viz*., Cucumber virus 3 (CV3) and also Cucumber virus 4 (CV4) (Ugaki *et al.*, 1991). It belongs to the well-known genus *Tobamovirus*, which includes TMV (Mandal *et al.*, 2008). Firstly, the small RNA (sRNA) deep sequencing as well as, rapid amplification of cDNA ends (RACE) techniques, were employed in full genome sequencing of CGMMV isolate (USA) on cucumber and its phylogenetic analysis revealed 98-99% similarity towards Asian isolates i.e., China, Israel, Japan, Korea, India, and Taiwan whereas it shares 90% identity with the European genotypes (Li *et al.*, 2015). Likewise, Indian strain of CGMMV (Accession Number: DQ767631.2) isolated from bottle gourd which seems to be monopartite, linear, ss RNA (+ve) genome of 6.4 kb. The (+ve) strand contains four ORFs encoding presumptive proteins of 186 kDa, 130 kDa, 30 kDa and 17.3 kDa (Kollipara, 2016). Besides this, the 3' terminal end carries the

TLS and the 5' terminal have the methylated nucleotide cap (m7G5'pppG) (Reingold *et al.*, 2016). It causes light green mottle mosaic and necrotic spots on leaves of greenhouse cucumber resulting in yield losses to the tune of 10-15% based on grower's estimation.

The sequencing of CGMMV (complete genome) watermelon strain (CGMMV-SH) revealed that it's four ORFs corresponding to proteins, *viz.*, 186 kDa, 129 kDa, 29 kDa and 17.3 kDa. It was also found that CGMMVSH watermelon strain is in close proximity to other strain of watermelon (CGMMV-W) with substitution of three AA in CP (Ugaki *et al.*, 1991). Chen *et al.*, (2009) determined the sequence of watermelon isolate (complete genome) from Liaoning province of China *i.e.*, CGMMV-LN. The gRNA of CGMMV-LN comprised 6422 nts encoding two replicase proteins (186 kDa and 129 kDa), MP (29 kDa) and CP (17.4 kDa).

Further, tobamoviruses causing infection to cucurbits were divided into two subcategories based on biological information, phylogenetic analysis and sequence data. CV3, CV4, CGMMV-W, CGMMV-SH and CGMMV-Is strain form the subcategory I and KGMMV, CGMMV-Y, CFMMV comprise subcategory II. However, it was found that KGMMV and CGMMV-Y have close relationships based on molecular and biological characterization (Antignus *et al.*, 2001). Varveri *et al.*, (2002) reported the CGMMV host range of Greek isolates and suggested that the virus retains its infectivity for 10 months when stored at 4°C. The sensitivity of ICRT-PCR based detection was 10^5 times higher than DAS-ELISA and 10^2 times greater in comparison to F(ab')2-ELISA.

2.3. Identification of *cis*-acting and *trans*-acting signals in the tobamovirus genome

Deconstruction of virus genome for developing virus-based vectors is a challenging attempt to avoid the inherent limitations of the full-length virus vector. The key approach for designing of deconstructed virus genome is that it should retain only some essential viral elements which are vital for the replication and translation (Gleba *et al.*, 2004, Gleba *et al.*, 2005; Gleba *et al.*, 2007; Mortimer *et al.*, 2015). A virus genome is composed of some *cis* and *trans*-regulatory sequences. Of them, *trans*-regulatory sequences contain the part of viral replicase and MP which is responsible for replication as well as the cell to cell movement but not obligatory for virus replication; thus can be deleted in the deconstructed virus genome. Similarly, it has been reported that in viral genomes, *cis*-acting RNA elements are also located and show

cardinal functions in most of the bottom line viral infection cycles, such as translation and replication (Hyodo *et al.*, 2017). It contains the primary (1^0) sequences, secondary (2^0) and tertiary (3^0) structures that are required for the binding of replicase, long-distance interactions between RNA elements, and interaction of viral /host proteins within the genome. Therefore, identification of *cis*-regulatory part is essential for its assiduous replication of deconstructed genome of the virus and efficacious transitory expressions of the targetted genes in plants.

Generally, these *cis*-elements sequences are located at 5′ and 3′ UTRs and internal subgenomic promoter (SGP) region that are recognized by the polymerase during minus-strand synthesis. In tobamovirus, about 70 nt long G-deficient CAA rich repeats sequence at 5′- UTR and about 200 nt long 3′-UTR containing ∼100nt TLS followed by one after the other three 70 nt long pk structures are highly momentous for replication and translation (Zenko *et al.*, 2002; Agalarov *et al.*, 2011; Chujo *et al.*, 2015). The replicase enzyme of TMV binds to 70 nt region located at 3-94 nt of the 5′ end during cotranslational replication of gRNA and this 5'UTR act as a *cis*-regulatory element for the synthesis of -ve strand RNA (Kawamura-Nagaya *et al.*, 2014). In addition to that, the N-terminal part of replicase enzyme-containing methyltransferase domain (MTase of TMV 130 kDa protein) might tether to the 5′-end in posttranslational fashion, and become essential for the replication of the genome. Later on, it is investigated that such *cis*-preferential viral protein obtained from many plants as well as animal +ve sense RNA viruses and also from their defective RNA (dRNA)/sgRNA particles. But the exact mechanism of these *cis*-preferential proteins is not explored yet.

2.3.1. *Cis*-acting and *trans*-acting for replication of the genome

The genomic replication of plus (+) strand RNA plant viruses is carried out by RdRp of replicase enzyme complex (REC) which synthesize minus (-) strand RNA intermediates. The genome replication and its regulation are performed by *cis*-acting elements which usually situated at 5′UTR, internal regions (IRs), 3′UTR in the viral genome. Among them, Brome mosaic virus (BMV) was considered as extensively investigated virus having tripartite genome (RNA 1-3) and its 3′-UTR folds into TLS in which a stem-loop C (SLC) have the major element for binding of virus replication complex (Sivakumaran *et al.*, 2003; Baumstark and Ahlquist, 2001; He *et al.*, 2020). The SLC consist of the flexible internal loop at lower stem region and AUA terminal loop containing a rigid upper region (Kim *et al.*, 2000). However, for the synthesis of

-ve sense RNA strand AUA loop specific conformation is required and that is recognized as clamped adenine motif (Sivakumaran *et al.*, 2000; Yi *et al.*, 2009). The inter-cistronic region of BMV RNA3 regulatory sequence helps in self refolding to an extended structure which matches RNA polymerase III promoter box-B sequence of the and also interacts with the BMV protein 1a (Gopinath *et al.*, 2005; French and Alquist, 1987). The B-box motif in BMV RNA1 and RNA2 5′-UTR plays important role in the replication of viral genome in 1a-dependent manner and –ve sense RNA2 SL at 3′-end act as core promoter. (Sivakumaran *et al.*, 1999; Chen *et al.*, 2001, Yi and Kao, 2008; Yi *et al.*, 2009). Similarly, RdRp (126-kDa) of TMV binds to 5′UTR and serve as a template for replication (Kawamura-Nagaya *et al.*, 2014). Whereas a stretch of pk like structure at the 3′ terminal end of TMV is the supreme regulatory region for (+) strand synthesis. However, in TMV RNA, 3′-terminal region (3′-TR) can be grooved into TLS with adjacent upstream pks (Felden *et al.*, 1996, Dreher 2009) which are essential for replication of TMV RNA (Osman *et al.*, 2000). Thus, the 5′ and 3′ terminal-ends of deconstructed virus genomes show similarity to the gRNAs from where they eventually derivated (Maia *et al.*, 1996); but may have different replication signals or sequences (Chandrika *et al.*, 2000). It has been observed that these replication signals/sequences may generate different 1^0 and 2^0 structures (Boccard and Baulcombe, 1993; van der Kuyl *et al.*, 1990; Zavriev *et al.*, 1996). Thus, the similarity of 1^0 and 2^0 structures in the 5′ and 3′ terminal end of the genomic and deconstructed genome may advise a valuable role in replication and its regulation. Sometimes, the *cis*-regulatory region may localize not only in genome's IRs but also in CRs of upstream ORFs from 5′ terminal end. In such cases, the mapping of these *cis*-regulatory regions is important for the replication of the deconstructed genome.

Additionally, for replication of TMV-based dRNA and its helper virus, *cis* and *trans* elements provide different requirements accord to TMV RNAs. It is reported that in absence of pk 1 and/or 2, replication of full-length RNA of TMV declined to a great extent, but the presence of pk3 is essential for full-length RNA replication as well as for dRNAs, and replication did not alter by identical deletions into dRNAs. Although comparability of replication between dRNA and full-length TMV RNAs shows that more homologous sequences are needed in case of dRNA replication whereas in presence of heterologous 3' nontranslated regions (NTRs) dRNAs unable to replicate/replicate nominally, the replication of full-length RNAs which mutated similarly seems to be little affected. The undermine interlinkage in between replicase

and 3' NTR can be compensated by an increase in the number of adjacent heterologous sequences (Chandrika *et al.*, 2000). Thus, pk revealed 28 nts termini involvement towards TLS aminoacyl acceptor arm, and that is noticeable in the replication of dRNAs and also full-length RNAs of TMV.

2.3.2. *Cis*-acting and *trans*-acting for genome translation

Extensively plant viruses have +ve sense RNA genome, hence to build up a stronghold in the plant cell the initial step is translation initiation was reported in most of the viruses. After entering the host cell, disassembly of a (+) strand RNA virion takes place and translation becomes the first event, while synthesis replication proteins of viruses. The (+) strand RNA genome must transform its guise from messenger RNA to replicase template once the accumulation of viral replication proteins. These events are inconsistent on the similar molecule so that *cis*-acting signals highly regulate the translation and replication (Dreher and Miller, 2006). The translation initiation elements are situated in the UTRs of viral genomes, while those control recoding events functioning as ribosomal frameshifting and in-frame read-through of stop codons are primarily in the coding regions adjacent to the site of coding event.

The appropriate and timely production of viral proteins inflates virus replication which mainly regulated by *cis*-acting signals *via* translational control events of viral RNAs *viz*., read through leaky termination codon, polyprotein synthesis, sgRNA production, etc. (Miller and Koev, 2000; Newburn and White, 2015). However, these events are usually controlled by *cis*-acting RNA (2^0structures) elements engaged in translational strategies.

Likewise, another mechanism shows that initiation of translation takes place *via* internal ribosome entry site (IRES), with the assistance of trans-acting factors to an internal initiation codon in the mRNA (Jackson, 2000). Whereas protein synthesis inhibited due to the shift in phosphorylation of eIF2 with virus infections/stress conditions, that lowers the ternary complex (eIF2-GTP-tRNA-Meti) formation. Contrarily, modification in the functionality of eIF4F changes specific shut-off of cap-dependent initiation (Martinez-Salas *et al.*, 2001).

The presence of TLSs, 5′-cap and absence of 3′- poly (A) tail are evident in TMV and turnip yellow mosaic virus (TYMV) (Gallie and Kobayashi 1994; Matsuda and Dreher, 2004). The crystalline nature of TLS of TYMV has aminoacylation activity loaded with eEF1A which promotes ribosomal binding (Dreher, 2009). However,

TMV TLS utilises upstream pks of RNA as translation enhancer where omega sequence at 5′UTR also promotes translation with the help of eukaryotic initiation factor (eIF)4F (Gallie, 2002). Hence, it can be concluded at TMV 5′ and 3′ end unconventional RNA elements facilitate cap-dependent translation.

2.3.3. *Cis*-acting and *trans*-acting for virion assembly

Virion assembly formation is the final stage of virus reproduction among RNA plant viruses either helical or icosahedral has been studied for a long time (Rossmann, 2013). The origin of assembly (OAS) site of TMV was the first known, located at 3′-end nearly 900-1300 nts and composed of three SL structures (Zimmern, 1983). Of them SL1 is sole requisite for bidirectional assembly formation and which binds the first double-disk CP subunit complex (Steckert and Schuster, 1982; Turner *et al.*, 1988; Newburn and White, 2015). On the other hand, the OAS of PVX (rod-shaped) is situated at 100 nt 5′-terminal and composed of large SL1, that binds CPs to the genome and also 5′-terminal found determinant for its assembly formation (Lough *et al.*, 2006; Petrova *et al.*, 2013). Hence, during infection OAS location could be beneficial for assembly formation as it prevents encapsidation of 3′-coterminal sg mRNAs. Similarly, in the case of BMV (icosahedral), TLSes have the main role in the packaging *in vitro* and hypothesized to the representation of OASes (Newburn and White, 2015). Moreover, TLS of RNA3 plays a significant role *in vivo* for non-replicating assembly assays.

Of four ORFs in the genome, only MP protein (29 kDa) is involved in the cell to cell movement and CP (17.3 kDa) protein play a crucial role in genome encapsidation (Ugaki *et al.*, 1991). Other than this, the CP has multiple other functions, for example, cell-to-cell movement, replication, virion formation, etc. (Callaway *et al.*, 2001). It is reported by Fukuda *et al.*, (1978) that the CP not only functions for the long cell to cell movement in tobamoviruses but also identifies the gRNA assembly initiation site for virion formation as well as 3′ and 5′ direction rod elongation of the virus. The CR of the CP gene is the main assembly initiation site of CGMMV (Zhang *et al.*, 2017). Thus, it is assumed that for virus infectivity, there are some interconnection works between the origin of virion assembly and adjacent region.

To delineate the infectivity of CGMMV, the infectious clone was created *via* deletion of GTT tri-nts at the origin of virion assembly (GAGGTTG), resulting abolition of valine^{97} (deletion mutation at residue 97 of CP). This is followed by the

creation of other mutant constructs with substitution mutation of valine^{94} to alanine and threonine^{104} to alanine. All three mutations are showing the delay in the production of the symptoms on *Nicotiana benthamiana*, as compared to wild type. In CP, the substitution of glutamic acid^{96} to lysine was very found common. These results indicated that infectivity of virus is largely affected by virus origin of virion assembly and adjoining region, and in this process glutamic acid^{96} might play a key role. Similarly, other CGMMV cDNA constructs with two single point-mutation, (i.e., glutamic acid^{96} to alanine, glutamic acid^{96} to lysine) and three double mutations (i.e., valine^{94} to alanine - glutamic acid^{96} to lysine, and valine ^{97} to glutamic acid, threonine^{104} to alanine, and glutamic acid^{96} to lysine) used to acknowledge the importance of glutamic acid^{96} in diseases development. On *N. benthamiana* inoculation, glutamic acid ^{96} to alanine have shown systemic but delayed symptoms whereas three double mutants and glutamic acid^{96} lysine displayed characteristic mosaic symptoms on the seventh day after inoculation. Then, watermelon plants were mechanically inoculated with the sap of *N. benthamiana* leaves already infected by CGMMV and confirmation of infection was done through double antibody sandwich -enzyme-linked immunosorbent assay ((DAS-ELISA), reverse transcription-polymerase chain reaction (RT-PCR), and sequencing (Zhang *et al.*, 2017). It was observed from the finding that infection of mutants can be compensated *via* E96K on virus infectivity. Additionally, the viral RNA accumulation confirmed the severity of the symptoms which revealed through northern blotting. In this way, the importance of single amino acids in the CP was addressed.

2.3.4. *Cis*-acting elements for subgenomic RNA production

Similarly, for translation, synthesis of sgRNAs is important which depends on some regulatory sequences. The sgRNAs are synthesized *via* a different mechanism. Some sgRNAs that are co-terminal at either 3′ or 5′ end; can be synthesized through internal initiation from SGP or premature termination at specific regulatory (stop) signal (Miller and Koev, 2000). In construct some sgRNAs having 5′ ladder with 3′ terminal sequences and synthesize gRNA at 3′ end -ve strand copy followed by a discontinuous step and switching of 5′ leader sequence. This 5′ leader sequence becomes the anti-leader in its -ve strand form and serves as the replicase binding site for sgRNA replication (Hofmann *et al.*, 1990, Sethna *et al.*, 1989). Besides this, a nested SGP located internally can be utilized for the expression of a foreign protein. This replication ability and subsequent expression of any foreign protein can be utilized

for the conversion of a sgRNA into an expression vector. Before that, the identification of *cis*-acting regulatory elements is necessary for the expression production of sgRNAs (Koev *et al.*, 1999; Miller and Koev, 2000).

Furthermore, promoter mapping is also important for the expression study of the sgRNAs. These SGP sequences can be placed anywhere within the genome and their characterization is important and also serve as replication sites during recombination (Miller *et al.*, 1995; Roossinck, 1997). The promoters differently function while *in vitro/in vivo* analysis and generally *in vivo* there is the requirement of additional sequence for its full function. *Bromoviruses* have core promoter region, that supports elementary transcription, and enhancer regions. The overall sgRNAs synthesis boosted up by these regions in viruses, which usually localized at promoter transcription start site upstream or downstream and overlapping it sometimes. In BMV they are located (+1 to +16 nts), upstream poly(A) stretch (-20 to -37 nts), and *UUA* repeats (3X) (-38 to -48 nts) (Marsh *et al.*, 1988). The first nt is identical for gRNA and sgRNA in many viruses but in some viruses, it appears as conserved due to mutations there is a great reduction or abolish sgRNA synthesis (van der Vossen *et al.*, 1995). Hence, sequence homologies revealed transcription initiation site upstream possess AU- or U-rich regions, of the and can be extended from 5′ ends up to 20 nt of these RNAs.

Such internal SGPs are mapped in case of BMV, TMV, etc. In case of BMV, the SGP for sgRNA4 is located within the sequence from -95 to +16 nt of CP start site with an enhancer (located between nts -95 to -20) (Wierzchoslawski and Bujarski, 2006). This region is enriched in AU sequence followed by a poly (U) tract, and a core region (-19 to-1 nts) containing an SL structure responsible for the binding of RdRp. The first promoter of TMV situated in between 183-kDa RdRp protein and MP ORFs, similarly in between MP and CP ORFs second promoter is located. The full SGP region of MP located at -95 to +40nt, with the -35 to +10nt as the core promoter (Grdzelishvili *et al.*, 2000). Whereas the minimal SGP for CP was mapped between -69 and +12 nt, and the fully active promoter was between -157 and +54 nt (Grdzelishvili *et al.*, 2000). These promoters as well as other regulatory sequence are located closer to the 5′, 3′ termini as well as within internal ORFs of the virus genome. Understanding about these regulatory sequences is key to synthesize sgRNAs with improved expression systems. Here, similar to TMV, our CGMMV has a unique genome organization with overlapping ORFs. It has two SGPs, posing 1^0 and 2^0 structural difference and to

control the expression of its gene, transcription and translation activities are utilized. This subgenomic transcript characterization and promoter identification have not been addressed in CGMMV but the outcome knowledge of study may contribute significance of promoter function in the plant.

2.4. Development of deconstructed genome & its biological assay

The replication of deconstructed genome occurs *in-trans* with help of wild type virus. The deconstructed genomic parts are mostly lack of replicase enzyme, and can trans-replicate when replicase of helper virus binds to the *cis*-regulatory element of the deconstructed genome. This type of deconstructed gRNA may establish either parasitic or symbiotic relationship with helper virus. The replication model of deconstructed gRNA is similar to their helper virus, and seems to be consistent with the facts, (i) the rates of appearance of deconstructed gRNA (+) and (-) strands are length-dependent phenomenon (ii) deconstructed gRNA may also be formed during + strand synthesis (Sethna *et al.*, 1989; Sawicki and Sawicki, 1990); and (iii) the 2^0 structure at 5′ and 3′ terminal sequence of deconstructed gRNA, is essential for its replication (Li and Aaskov, 2014; Lin *et al.*, 1994). During replication, this deconstructed gRNAs must be competing with helper virus genomes for viral proteins. These deconstructed genomes may interfere with the infection process of helper virus and reduce the disease severity. Further, if there is a symbiotic relationship, they may support the help virus in symptoms expression, as well as transmission. This dynamic relationship between helper virus and its deconstructed genomes should be deciphered for its further utilization.

Besides this *trans*-replication, the deconstructed genomic particles also contain one or more ORFs under the control of SGP. Thus they act as the messengers and translated into protein. The insertion of a foreign gene within this region will help to express the foreign protein. This property was explored for the construction of deconstructed RNA (dRNA)-based vector system in TMV (dRT-V) which represent the bipartite module of TMV. Where for its replication, dRNA relies on helper virus and can be utilized for multiple therapeutic proteins higher expression inclusive human monoclonal antibody (mAb) in the plants (Roy *et al.*, 2010). Another TMV based transient vector 'MagnICON' was developed by deconstructing the monopartite TMV genome into two parts; one component contains MP and CP and another one contains replicase enzyme (Gleba *et al.*, 2005). This vector-based platform is highly

recommended in various plant systems for expression of genes by replacing the CP gene with foreign genes and also capable of transgene expression regulation as well as amplification of stably transformed plants, which results in recombinant protein accumulation in plants at a higher level (Marillonnet *et al.*, 2004; Gleba *et al.*, 2007; Werner *et al.*, 2011; Gleba *et al.*, 2014; Mortimer *et al.*, 2015). Hence, for production of higher level, recombinant protein in plants MagnICON' vectors are widely used in their transient formats, due to its stability and ease to adopt in more plants (Gleba *et al.*, 2007). Thus, the technique of deconstructed TMV genome-based replicon vector was explored by several workers for the production of the biopharmaceuticals.

2.5. Functional significance of the deconstructed genome of plant viruses

2.5.1. Use for the *In planta* expression of foreign proteins

Earlier, various RNA viruses vectors have been played a significant role of foreign genes expression in plants including TMV (Donson *et al.*, 1991; Chapman *et al.*, 1992; Shivprasad *et al.*, 1999; Dohi *et al.*, 2006; Yusibov *et al.*, 2006). It is reported these RNA viral vectors shows its replication in the cytoplasmic organelle of the plant cell and results in a higher number of expression of heterologous protein in lesser time stretch without entering into the nucleus (Sainsbury *et al.*, 2008; Roy *et al.*, 2011). However, simultaneous expression of several genes in one cell with the help of plant virus vectors has been limited, due to the competition of homologous viruses containing heterologous genes.

Roy *et al.*, (2010) have utilized two components of the alike virus to engineer dRT-V, where replication of its dRNA component takes place with help of the helper virus. The results revealed that human growth hormone (hGH) and *Bacillus anthracis* lichenase-fused lethal factor (lic-LF) protein expression were of high-level with *N. benthamiana* plants co-inoculated through *Agrobacterium* using dRT-V two components. The dRT-V make use of small-sized plasmids i.e., pGRDef (6 kb), pGRD4 (10.3 kb) in comparison to pBID4 (19 kb approximately), those were advantageous and effective in downstream gene manipulation as well as with agroinfiltration recombinant target genes delivery can become speedy into plant cells.

The vectors (pGRD4) were taken in dRT-V system, as helper TMV vectors, which support functional replicase and MPs in trans to the dRNAs and also shows the expression of foreign genes within one and the similar cells. Thus, Roy *et al.*, (2010) reported successful coexpression within a single cell of two heterologous proteins,

inclusive of GFP and hGH, Lic-LF and hGH, and GFP and DsRed. However in earlier findings, expression of two heterologous proteins was accomplished either by multiple copies of the deleted version of a two-component virus (e.g., *Cowpea mosaic virus*) or expressing proteins from non-competing one component viruses (e.g., *Potyvirus* and *Potato virus X* (PVX) or TMV and PVX) (Balogun *et al.*, 2002; Sainsbury *et al.*, 2008; Larsen and Curtis, 2012). However, with the use of TMV and PVX, expression of human mAb could be observed, which require combined inoculated six cultures of *Agrobacterium* (Hiatt and Pauly, 2006; Sainsbury and Lomonossoff, 2008). Roy *et al.*, (2010) generated full-length mAb only including two *Agrobacterium* cultures in their dRT-V system. It was observed in the experiments that (pGRDefLC + pGRD4HC) was found superior over pGRDefHC + pGRD4LC for production of an assembled mAb; despite the unavailability of the optimized procedures used in the vector system to admit potential overall increase in yield of the assembled mAb.

Till now little information/ experimentations are evident regarding the expression of foreign proteins by using CGMMV genome in the plant. Ooi *et al.*, (2006) recorded that the small sequence of 99 nts carrying hepatitis B surface antigen epitopes was expressed through the insertion of read-through codon strategy at CGMMV CP ORF end and a similar pattern was followed for envelope protein expression of the dengue virus type-2 (132 nt stretches) (Jailani *et al.*, 2017). Subsequently, GFP gene insertion into MP and CGMMV coding frame of CP resulted in to weaken the expression of a larger gene sequence in *N. benthamiana* (Zheng *et al.*, 2015; Ali *et al.*, 2016). The full-virus based technique was taken out in all these studies, which manifest the following constraints: (i) ineffectual larger gene expression (ii) variability in read-through protein efficiency accord to the non -identical host-virus interactions, (iii) delayed foreign gene expression and (iv) viral pathogen generation.

Jailani *et al.*, (2017) developed a highly infective clone with the help of gene expression vector for CGMMV in cucurbits. Therein, vector's two versions were evaluated in *N. benthamiana* for their expression efficiency of GFP. The GFP expression, virion assembly and systemically symptom formation was reported in *N. benthamiana* while insertion of GFP gene is taken place at stop codon region of CGMMV CP gene without any read-through codon. Further, qRT-PCR analysis signifies twenty-three times superiority of GFP as compared to actin at ten days post-inoculation (dpi) that can also escalate up to forty-five times at 14 dpi and after then

significant decline in GFP expression was observed. Although, after deletion of CP sequences having only first 105 nts, shortened vector containing GFP in CR of CP resulted in 234 times upliftment in its expression as compared to actin at 5 dpi in *N. benthamiana* without forming any virion and symptom development of the disease. The study demonstrated that for development of a plant virus-based active expression system of a foreign protein in the plant, CP gene manipulations in the genome of CGMMV is required while preserving CP translational frame and this developed CGMMV based vectors may be possibly helpful in the research and development of edible vaccines.

In present scenario array of plant-based products are in clinical trials at variable developmental stages. To examine the functional efficiency of these plant-based vaccines, it was tested against destructive virus i.e., porcine reproductive and respiratory syndrome virus (PRRSV) in Canadian pork slaughtering industrial areas. Tran *et al.*, (2019) developed an expression vector *via* complete genome cloning of CGMMV and agro-infiltration technique to explicit the PRRSV glycoprotein 5 (GP5) neutralizing epitope (NE) genes on leaves of cucumber. The *GP5 NE* gene was integrated within the coding frame of CP and its expression utilized read-through mechanism. The expression levels attain cucumber fresh leaf weight (35.84 mg/kg) and chimeric virus particles (CVP) are also found stable. The outcome of the study proposed a propitious solution for affordable vaccine production, booming against PRRSV as oral delivery through CVP display system.

2.5.2. Use for gene silencing/ functional genomics

In transient gain- and loss-of-function studies, delivery of genes through plant virus vector, focused as a preferred strategy, especially in genetic transformation of plant species recalcitrant (Zhang and Ghabrial, 2006; Ding *et al.*, 2018). Likewise, simple cloning and inoculation techniques were employed in the delivery of a targetted gene to the entire plant in little period for infection development *via* plant virus vector. The virus vector systems are high-handed over transgenic approach regarding the genes of interest expression in plants because the high replication rate of viruses can be seen with infected cells (Choi *et al.*, 2019). Robertson in 2004 also suggested that silencing of a gene induced by virus (VIGS) vectors can be explored for transient loss-of-function studies (Ramegowda *et al.*, 2014; Choi *et al.*, 2019). However, this system has been extensively utilized since last twenty years for quick and efficacious characterization

of several plant-associated genes engaged in plant growth and development, regulatory metabolism, and to provide resistance against disease in different species of plants.

The VIGS used as a powerful tool in the field of plant functional genomics due to its speedy and accessible delivery (Lu *et al.*, 2003). It also used for the downregulation of host endogenous gene expression at the post-transcriptional level and revealed the typical phenotypic changes in the entire plant within 2-3 weeks. Therefore, for validation of plant genes associated with plant growth, plant disease immunity and control of metabolic genes, VIGS has been used broadly and it is evident that remarkable improvement has been achieved in last five years, on different hosts (Burch-Smith *et al.*, 2004; Shao *et al.*, 2008). The *Apple latent spherical virus* (ALSV), *Cucumber mosaic virus* (CMV), *Soybean yellow common mosaic virus* (SYCMV), *Tobacco Rattle Virus* (TRV), *Potato virus X* (PVX), and *Tomato yellow leaf curl China virus* (TYLCCNV), RNA/DNA plant viruses were used for the different host in VIGS system (Turnage *et al.*, 2002; Faivre-Rampant *et al.*, 2004; Nagamatsu *et al.*, 2007; Igarashi *et al.*, 2009; Bachan and Kumar, 2012; Lim *et al.*, 2016). However, very less number of vectors reported for VIGS system in cucurbits even though by taking the help of recent molecular techniques, construction of VIGS vectors for cucurbits will become easy and inevitable, which impart huge value in the apprehension of the useful qualities in functional genes of cucurbit.

The construction of CGMMV based VIGS vector and its efficiency as RNAi vector for endogenous host gene silencing in *N. benthamiana*. So far, CGMMV full-length genome-based VIGS vector was developed by Liu *et al.*, (2020) and validated for the gene function assay in cucurbits but more efficiently in *N. benthamiana*. The sense and hairpin strategy were followed for gene silencing.

Liu *et al.* (2020) have studied a newly VIGS vector evolved from CGMMV, that was proficiently induced phytoene desaturase (*PDS*) gene silencing in *N. benthamiana* and in cucurbits like cucumber, melon, etc. This vector causes infection by nurturing *PDS* sequences (69-300 bp long) while the formation of hairpin cDNAs and showed phenotypic photobleaching in *N. benthamiana* as well as in cucurbits as a resultant of *PDS* silencing. However, findings also revealed that *PDS* silencing can persist up to bi-monthly and its effect could be transmitted.

3.1. To create deconstructed genome of CGMMV with *cis*-regulatory sequence

3.1.1. To identify the cis-regulatory sequences for the genome replication and translation

3.1.1.1. Genome-wide prediction of regulatory elements in CGMMV

3.1.1.1.1.Virus isolates and genome organization

The complete genome sequence of our laboratory isolate BgDel of CGMMV (Accession no: DQ767631.2) along with other 105 isolates reported from the various parts of the world were retrieved from the NCBI website (https://www.ncbi.nlm.nih.gov/nuccore/ DQ767631.2). The genome organization of the virus was analyzed based on the literature search. The important RNA secondary structures present in the genome were predicted using the Rfam database (https://rfam.xfam.org/), and various proteins domains of the replicase enzyme were annotated using the Pfam database (https://pfam.xfam.org/).

3.1.1.1.3.RNA secondary structure prediction

Initially, the RNA secondary structure of the complete genome sequence of CGMMV isolate BgDel (Accession no: DQ767631.2) was derived using the Mfold Web server (http://mfold.rna.albany.edu/) at 37 °C folding temperature with other default parameters (Zuker *et al.*, 2003). The potential stable structures with minimal Gibbs free energy in the output file were selected. The same sequence was also run in the ScanFold-Scan program (https://github.com/moss-lab/ScanFold) with the window Size of 120 nucleotides (nt), 50 randomizations at the temperature of 37°C to predict the genome-wide distribution of RNA secondary structures (Andrews *et al.*, 2018). Further comparative analysis was made by using DotKnot (https://dotknot.csse.uwa.edu.au/) to predict RNA pseudoknots with kissing hairpins as the putative regulatory regions located at 5’ and 3’ terminal end of genome (Sperschneider and Datta, 2010). Whereas, the alignment file of complete genome sequences were derived through RNAz web-server (http://rna.tbi.univie.ac.at/cgi-bin/RNAz/RNAz.cgi) using the 120 nucleotides

of window size (Gruber *et al.*, 2010), to detect the structural conservation of the predicted regulatory elements across the reported isolates of CGMMV.

3.1.1.2. *In silico* prediction of replicase enzyme binding domain at the 5' and 3' end of the genome

3.1.1.2.1.Alignment and comparison of the sequences

The genome sequences of the 105 isolates of the CGMMV were checked and pair-wise sequence alignment was performed by using ClustalW (Thompson *et al.*, 1994) with 1000 number of bootstraps in Bioedit7.0. sequence alignment editor (Hall, 1999). The aligned sequences were further proceeded to identify the conserved sequences in the different genomic segments and to construct the phylogenetic tree by the Maximum Likelihood method using 1000 bootstrap iterations in MEGA 7.0 (Molecular Evolutionary Genetics Analysis version 7.0) tool (Kumar *et al.*, 2016)**.**

3.1.1.2.2. Modeling of the protein 3D structure and their analyses

To develop a 3D structural model of helicase and RdRp, 254 amino acids long amino acid sequences of helicase domain and that of amino acids of core RdRp were subjected to protein BLAST against the PDB (Protein Data Bank) database (http://www.rcsb.org/pdb/home/home.do; Berman *et al.*, 2000) for searching the suitable templates having almost similar crystal structures, that were used as a template for the target proteins modeling. These templates were selected based on the sequence identity, alignment percentage, positives, and gaps present between the template and the target sequences, and the models were obtained in swiss-model web-server (https://swissmodel.expasy.org/interactive/05JjhN/).

The constructed models were subjected to structural simulation/refinement and model verification by a series of tests for testing its internal consistency and reliability through various bioinformatics tools. RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php; Lovell *et al.*, 2003) and molprobity was used for determination of the stereochemical quality of the homology-based structural model by assessing parameters such as lengths, bond angles, and planarity of the peptide bonds, the geometry of the hydrogen bonds and side-chain conformations of the amino acids as a function of atomic resolution, which relies on the Ramachandran plot statistics. The non-bonded interactions between different atom-types were checked by the ERRAT

plot (http://nihserver.mbi.ucla.edu/ERRATv2/;Colovos *et al.*, 1993) to assess the distribution of different types of atoms concerning one another in the protein model for deciding model reliability. ProSA (https://prosa.services.came.sbg.ac.at/prosa.php; Wiederstein and Sippl, 2007) was also used for the validation of modeled protein structures, based on Z-scores and energy plots to indicate the overall model quality and measures total energy of the structure.

3.1.1.2.3. Modeling of RNA 3D structures and their analyses

The sequence of predicted knots/pseudoknots and hairpin loops were used to design the complex three-dimensional (3D) structures based on the template-dependent approach. Initially, models were generated in the Hdock server based on the highest sequence identity (%) with the crystal structures of the respective RNA template as in the PDB database. The quality of the models was checked using MolProbity (http://molprobity.biochem.duke.edu/) (Chen *et al.*, 2010) and RNAlyzer tool (http://rnalyzer.cs.put.poznan.pl/) (Lukasiak *et al.*, 2013). Further, their secondary structures were annotated in RNApdbee web-server (http://rnapdbee.cs.put.poznan.pl/) and compared to validate the generated models (Zok *et al.*, 2018). Based on the scores of different parameters within the acceptable range and PDB derived RNA secondary structures, the models are selected. The selected were then handed-down for molecular docking.

3.1.1.2.4. *In silico* docking between protein and RNA for predicting probable interacting partners

Initially, the RNA-protein docking using various models was performed using the Hdock server (http://hdock.phys.hust.edu.cn/), which a protein-DNA/RNA docking tool based on a hybrid algorithm of template-based as well as template free modeling (Yan *et al.*, 2017). The Hdock performs global docking to predict the interface of interacting protein and RNA binding complexes when no information about the binding site is available for the docking. Total 10 docking models with the lowest docking energy scores and the highest ligand root-mean-square deviation (RMSD) values were selected to analyze their interaction. The predicted interface and amino acids involved in the interaction was further validated in NPDock (http://genesilico.pl/NPDock) web server (Tuszynska *et al.*, 2015). NPDock (Nucleic acid-protein dock) was used to combine the application of GRAMM for global macromolecular docking, uses atomic

coordinated between molecules and results scoring with a statistical potential, clustering of best-scored structures at 5^0A RMSD cut-off, with the advanced option of interfaces selection and local refinement of the structures at 310K temperature.

3.1.1.2.5. Analysis of the Protein-RNA Interaction

The complex interaction between protein and RNA was visualized using Pymol (https://pymol.org/2/) and Discovery studio visualize (https://discover.3ds.com/discovery-studio-visualizer-), and the interacting amino acid residues and nucleotides were identified in the protein-ligand- profiler (https://projects.biotec.tu-dresden.de/plip-web/plip) and PDBe PISA v1.52 (www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver). Further refinement was made using the protein plus web tool (https://proteins.plus/).

3.1.1.3. *In silico* prediction of coat protein promoter

3.1.1.3.1. Retrieval of virus sequences and selection of promoter region

The full annotated, complete genome sequence of different tobamoviruses, retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/). Variation in their genome length was noted. Further, the translational start site (TLSS) of MP and CP was spotted. Approximately, few nucleotides (200-230 nt) from both sides, i.e., upstream and downstream region of the transcription start sites (TSS) was selected for promoter mapping. The selected sequences were further used for structural analysis.

3.1.1.3.2. Alignment and comparison of the promoter sequences

The MP-SGP and CP-SGP sequences of the selected tobamoviruses were checked and pair-wise alignment of sequences was performed by using ClustalW (Thompson *et al.*, 1994) with 1000 bootstrap value in Bioedit 7.0 sequence alignment editor (Hall, 1999). The aligned sequences were used to identify the similarity and genomic conservation within the respective SGP region of different tobamoviruses. Further, their evolutionary relationship was drawn through the construction of the phylogenetic tree based on the Neighbor-Joining method using 1000 bootstrap iterations in MEGA 7.0 (Molecular Evolutionary Genetics Analysis version 7.0) tool (Kumar *et al.*, 2016).

3.1.1.3.3. Prediction of RNA secondary structures

To identify the potential elements like hairpins, pseudoknots, etc., the RNA secondary structures of the MP-SGP and CP- SGP region of different tobamoviruses were designed using the Mfold Web server (http://mfold.rna.albany.edu/) at 37°C folding temperature keeping other default parameters constant (Zuker *et al.*, 2003). The most stable structures with the lowest Gibbs free energy in the output file were selected. The structures were also visualized using VARNA (Darty *et al.*, 2009).

3.1.1.3.4. Identification of conserved motifs

To identify the presence of various motif sequences within the SGP region, the sequences of selected tobamoviruses were submitted in the Plant Care database (http://bioinformatics.psb.ugent.be /webtools/plantcare/html/). The conserved motif sequences were identified from the output file (Lescot *et al.*, 2002). Only the presence of *cis*-elements in the '+' strand was considered for the single-stranded RNA genome.

3.1.1.3.4. Homology modeling of the three-dimensional structure of the protein

The absence of a crystal structure of tobamovirus RdRp enzyme leads us to develop its three-dimensional models for its structural and functional analysis. To implement this, the amino acid sequence of the core RdRp of CGMMV (Accession no: ABG26381) was retrieved from the NCBI reference sequence database. The protein BLAST of 109 amino acid long core RdRp against the PDB (Protein Data Bank) database (http://www.rcsb.org/pdb/home/ home.do; Berman *et al.*, 2000) resulted in maximum ~27.5% sequence similarity with the various crystal structures of the viral RdRp. Of them, few PDB models were selected based on sequence identity, alignment score, positives, and gaps present between the template and the target sequences, and used as the suitable templates for generating target proteins models in swiss-model web-server (https://swissmodel.expasy.org/interactive/05JjhN/). Among the constructed models, only one model was selected based on per-cent similarity and homology with the Q-MEAN score. The structure was further subjected to structural refinement through web tool and model verification was done by a series of tests for its internal consistency and reliability through various bioinformatics tools, viz., RAMPAGE (http://mordred.bioc.cam.ac.uk/ ~rapper/rampage. Php) and Molprobity (http://molprobity. biochem.duke.edu/) for the determination of the stereochemical quality of models (Lovell et al., 2003; Williams et al., 2018), ERRAT

plot (http://nihserver.mbi.ucla.edu/ ERRATv2/) to assess the distribution of different types of atoms concerning one another in the protein model for deciding model reliability (Colovos and Yeates, 1993), and ProSA (https://prosa.services.came.sbg.ac.at/prosa.php) for the validation of modeled protein structures, based on Z-scores and energy plots to indicate the overall model quality (Wiederstein and Sippl, 2007).

3.1.1.3.5. Modeling of the three-dimensional structure of RNA

The whole SGP region of CP and MP was divided into an upstream and downstream part from the translational start site (TLSS). Then, both the upstream and down-stream promoter sequences are segmented into various lengths, i.e., 200nt, 175nt, 150nt, 125nt, 100nt, 75nt, 50nts from TSS, irrespective of the presence of internal motifs, predicted knots/pseudoknots, and hairpin structures. The secondary structure of these segmented RNA sequences was derived from the Mfold web server. Now, a template-independent approach was adopted for designing the complex three-dimensional (3D) structures of segmented parts of both promoter regions. RNA composure (http://rnacomposer.cs.put.poznan.pl/) was used to draw the respective models of large RNA sequences along with secondary structure details in dot-bracket format (Biesiada *et al.*, 2016). The quality of the 3D models was checked by retrieving their secondary structures from 3D models in RNApdbee web-server (http://rnapdbee.cs.put.poznan.pl/) and compared with the originally generated 2D-models (Zok *et al.*, 2018). The selected 3D-models were then used for molecular docking analysis.

3.1.1.3.6. Protein–RNA docking and their analysis

Initially, the RNA-protein docking was performed using the PatchDock server (http://bioinfo3d.cs.tau.ac.il/PatchDock/patchdock.html), which a molecular docking algorithm performing based on the shape complementarity principles (Schneidman-Duhovny *et al.*, 2005). The PDB models of the RdRp and RNA molecules were submitted in this web-tool with default parameters. The recommended RMSD cutoff was restricted to 4˚A for protein-RNA interaction. A total of 20 complex PDB models were sorted based on the highest geometric shape complementarity score. Further, these complex models were submitted in the FireDock (http://bioinfo3d.cs.tau.ac.il/FireDock/php.php) for refinement and re-scoring of rigid-body protein-RNA docking results (Mashiach *et al.*, 2008). The best protein-RNA

interacting models with the lowest docking energy scores were selected to analyze their interaction and to predict their interface of RdRp binding at the SGP region. To validate the result, the best possible models were again submitted in NPDock (http://genesilico.pl/NPDock) web server, specialized for nucleic acid-protein docking (Tuszynska *et al.*, 2015). NPDock (Nucleic acid-protein Dock). The best-scored structures (complex model) with the lowest energy value clustering at 5˚A RMSD cut-off, were selected for analysis. The complex interaction between protein and RNA was visualized using Pymol (https://pymol.org/2/) and Discovery studio visualize (https://discover.3ds.com/discovery-studio-visualizer-), and the interacting amino acid residues and nucleotides were identified in the protein-ligand- profiler (https://projects.biotec.tu-dresden.de/plip-web/plip) also.

3.1.2. Cloning of deconstructed genome of cucumber green mottle mosaic virus

3.1.2.1. Designing of deconstructed genome-1 and its RNA secondary structure analysis

Initially, the deconstructed RNA genome of CGMMV was designed based on the sequence alignment of the previously reported dRNA genome structure of TMV containing 1-841 nt of 5' terminal and 5182-6424 nt of 3' terminal of the genome (Roy *et al.*, 2010). The RNA secondary structure of the wild type CGMMV and its deconstructed genome was designed using the Mfold Web server (http://mfold.rna.albany.edu/). The complete RNA genome sequence (RNA), as well as the dRNA genome of CGMMV, was used as the input file. The secondary structure of dRNA was predicted at 37 °C using Mfold version 2.3 (Mathews *et al.*,1999; Zuker, 2003). The potential stable structures with minimal Gibbs free energy in output file were selected and a comparative analysis was made to predict RNA pseudoknots, the putative regulatory regions located at 5' and 3' terminal of the genome; so determine the possibility of trans-replication of the deconstructed genome.

3.1.2.2. Development of deconstructed genome-1 by overlapping extension PCR.

Requirements:

Table 3.1. The details about the binary vectors used in this study

Vector	Plant selection marker	Bacterial selection marker	Size (kb)	Reporter	Specification
pBM1 (Modified pCambia 2300)	Kan	Kan	9.2	Nil	Modified through the insertion of with 35S promoter and ribozyme sequence (RZ) and nos terminator
pBP4 (Symptomatic infectious cDNA clone of CGMMV)	Kan	Kan	15.6	Nil	Full-length CGMMV infectious cDNA was cloned into pBM1 binary vector with BamH1 and XbaI restriction sites
pBP7 (Asymptomatic infectious cDNA clone of CGMMV)	Kan	Kan	15.6	Nil	Artificially created through site-directed mutagenesis from symptomatic CGMMV (pBP4)

The deconstructed plasmid was derived from the infectious clone of CGMMV-BgDel isolate (CGMMV-BP4), through the deletion of the original plasmid construct using Overlapping Primer Extension PCR based cloning strategy (**Fig. 3.1**). The mechanism involved two-step PCR using two long primers especially designed with high GC content and share some degree of homology at their 5'end. The polymerase chain reaction was performed using a 50 µl master mix containing 1XPCR buffer, 0.8 mM dNTPs 1 U hi-fidelity Phusion Taq DNA polymerase, 100 ng of the vector (plasmid) and 1 mM of each primer, and thermal cycling was fixed with 1 cycle at 98°C for 5 min; 20 cycles at 98°C for 20 sec, 65-70°C for 8 min (@ 1kb/30s); and a final hold at 4°C. After running the PCR reactions, the products were directly treated with DpnI enzyme @ 1U per 50µl for 1hr 30 mins without further purification and finally transformed into highly efficient competent cells of *E.coli* DH5α (10^8 to 10^9cfu/µg). Colony PCR was done to screen out the colonies containing desirable plasmid

constructs, and further, the positive colonies were verified through restriction digestion of the plasmids. The PCR-derived deletion of the constructs was further verified through the sequencing of the deconstructed genome to ensure that only the desired modifications were present.

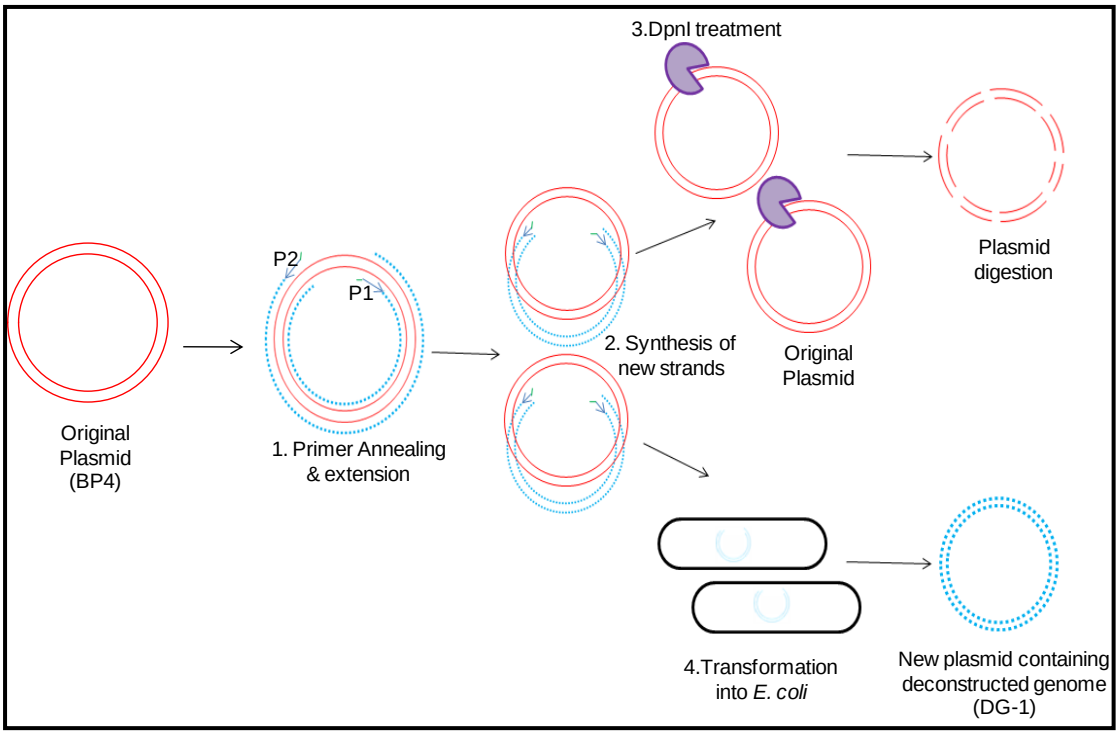


Fig. 3.1. Schematic diagram of deconstruction strategy of CGMMV infectious clone (BP4) using Overlapping Primer Extension PCR based cloning technique. It involves a two-step PCR protocol, having denaturation(95˚C), primer annealing, and extension together at (65-72˚C). The hi-fidelity Phusion Taq DNA polymerase is used to perform PCR reaction. DpnI restriction enzyme is used for the degradation of the original plasmid. After the restriction digestion of the original plasmid, the PCR amplicons were transformed into the *E.coli* competent cell to seal the nick and *in vivo* circularization of the deconstructed plasmid.

3.1.2.3. Designing of deconstructed genome-2 and its RNA secondary structure analysis

Another deconstructed genome of CGMMV was designed based on the *in silico* prediction of *cis*-regulatory elements of the CGMMV genome. Our *in silico* prediction helped in identifying certain *cis*-regulatory elements that were indispensable for replication and translation. These regulatory elements were mostly located within 5' and 3' UTRs of the genome and sub-genomic promoter region of MP and CP. These

sequences are merged to form a complete replicative deconstructed genome, i.e., DG-2. Some restriction sites were inserted into multiple locations within the DG-2 genome with aim of facilitating the insertion of multiple foreign genes. The RNA secondary structure of DG-2 was designed using the Mfold web-server (http://mfold.rna.albany.edu/) and compared with the RNA secondary structure of wild type CGMMV. The complete RNA genome sequence (RNA), as well as the dRNA genome of CGMMV, was used as the input file, and potential stable structures with minimal Gibbs free energy in the output file were selected. The comparative analysis was made to identify the conserveness of regulatory elements located at 5' and 3' terminal of the genome.

3.1.2.4.Synthesis and cloning of deconstructed genome-2

3.1.2.4.1.Synthesis of deconstructed genome-2

The deconstructed genome (DG-2) was synthesized from Bio BasicInc. Canada. The synthetic DNA product was cloned into a pUC57 vector with SacI/KpnI restriction sites. The product was supplied to us in the lyophilized form. The lyophilized DNA (4 µg) was re-suspended in 40µl of 10mM Tris buffer (pH 8.5) to keep the final concentration @100ng/µl. This original stock was further diluted in 1:10 ration using 10mM Tris buffer (pH 8.5) and 2µl of the working solution was transferred to *E. coli* JM109 competent cells and incubated at 37°C overnight with appropriate antibiotic selection (Ampiciline@100mg/ml, X-GAL@20 mg /ml and 100mM IPTG). Single colonies were selected and cultured in LB media to obtain a large quantity of plasmid DNA. To verify our construct, restriction digestion was done.

3.1.2.4.2. Cloning of the synthetic deconstructed genome-2 into pGreenII 0029 plant binary vector

Table 3.2. The details about the binary vectors used for the cloning of deconstructed genome-2

Vector	Plant selection marker	Bacterial selection marker	Size (kb)	Reporter	Specification
Modified pGreen II	Kan	Kan	3.4	LacZ	Modified through the insertion of with 35S promoter and ribozyme sequence (RZ) and NOS terminator

The synthetic gene was separated from the pUC57 through restriction digestion using fast digest SacI and KpnI restriction enzymes. The digested synthetic product was eluted from the agarose gel utilizing Wizard®SV Gel and PCR Cleanup System (Promega, Madison, WI, USA) as per manufacturer's protocol. After gel elution, the purified product was quantified in a nanodrop spectrophotometer (Thermo Scientific, USA). Then, the product was ligated into the pGreenII 0029 digested with the same SacI and KpnI restriction enzymes.

For setting up ligation reactions, the optimum concentration of the insert was calculated using the formula

$$\frac{\text{ng of vector} \times \text{kbp size of insert}}{\text{kbp size of the vector}} \times \text{insert: vector molar ratio} = \text{ng of the insert.}$$

The reaction of ligation set up as described below:

Reaction Component	**Reaction amount**
2X Ligation Buffer	5µl
Vector (50ng)	1µl
PCR product	2-3µl
T4 DNA Ligase	1µl

The final volume was made up to 10 µl with MilliQ water and the reaction mix was incubated at 16°C overnight at the water-bath.

3.1.2.5.Preparation of competent bacterial cells

Requirements:

1. **Luria-Bertani Broth (LB)**
2. **Luria-Bertani Agar (LA)**
3. **10%glycerol solution**
4. **0.1 M $CaCl_2$ solution**
5. **0.1 M $MgCl_2$ solution**

Procedure:

Competent bacterial (*E.coli*) cells were prepared to make them ready for accepting foreign DNA molecules. The bacterial cell wall was altered by chemical treatment ($CaCl_2$) so that the foreign DNA can be passed through easily. For the preparation of competent cells, the single colony in *E.coli* (strain DH5α or JM109) was picked from a fresh culture plate and transferred into 5 ml culture tube containing LB broth. The culture was grown overnight at 37°C incubator (kunher shaker) for 8-12 hrs at 220 rpm. From the overnight grown culture of *E. coli*, 500 µl was transferred to freshly prepared autoclaved 250 ml LB broth, and then the culture was kept at 37°C for 1.5- 2hrs at 250 rpm until O. D. reached at 0.6. The culture was kept in ice for 20 min and then transfer to the autoclaved ice-cold polypropylene tubes. The culture tubes were kept on 4°C ice for 30 minutes. After ice-cold incubation, culture tubes were centrifuged at 6000 rpm at 4°C for 10 mins. After then, the supernatant was discarded, and pellets were dissolved gently in 30 ml of 0.1 M $CaCl_2$ solution and incubated in ice at 4˚C for 30-60 mins, and then centrifuged again for 6000 rpm at 4°C for 10 mins. The supernatant was discarded and the pellet was dissolved again with 30ml of 0.1M $MgCl_2$ and incubated in ice for 30-60 min. After centrifugation @6000rpm for 10 mins, the supernatant was discarded and the pellet was dissolved in 2ml of autoclaved 0.1 M $CaCl_2$ solution containing 10% glycerol. The 100 µl culture of competent cells was aliquoted into ice-cold 1.5 ml of the centrifuge tube. After transferring, the tubes containing competent cells were dips with liquid nitrogen for a few seconds and stored in a -80°C deep freezer.

3.1.2.6. *E. coli* Transformation procedure

Requirements:

1.SOB medium (pH 7.0)

2. SOC medium: SOB medium containing 20 mM Glucose.

3. X-gal stock: 40 mg X-gal/ml of Dimethyl-formamide.

4. IPTG stock (100mM): 23.8 mg/ml of double-distilled water.

5. Antibiotics (Stock solutions): Ampicillin, Kanamycin (kan), Tetracycline (tet)

Procedure:

10 µl of the ligated product was mixed with 100 µl competent cells for transformation and cells were mixed by gentle tapping. Cells were then incubated in ice for 30 mins, followed by a gentle heat shock at 42°C for 60 seconds, and immediately placed back in ice for 10 mins. Then, 1 ml of SOC medium was added to the heat-shocked cell and incubated at shaker incubator @250 rpm at 37°C for 1.5-2hrs. Incubated cells were then spread on Luria-Bertani agar (LA) containing plates with the appropriate antibiotic selection. The plates were incubated at 37°C overnight.

3.1.2.7. Colony PCR

The transformed colonies were confirmed by colony PCR as described by Sambrook and Russel(2001). Sambrook and Russel (2000). The recombinant bacterial colonies were streaked on appropriate antibiotic plates were picked up with the help of autoclaved tips and incubated at the 37°C incubator overnight. From the overnight culture plate, individual colonies were picked and dissolved in 50 µl of DNase-free water. The bacterial suspension was allowed at 70°C for 10 min and quickly returned to the ice. The cell debris was removed by centrifugation at 10,000 rpm for 5 min and the supernatant was used for PCR as a template and used gene-specific primers. The PCR protocol was followed as per the following protocol: Primary denaturation at 94˚C for 3mins; followed by 35 cycles of final denaturation at 94˚C for 30sec, annealing at 52-60˚C for 30 sec (depending on primer Tm), and extension at 72˚C for 30sec to 2mins (depending on the amplicon size); and then final extension at 72˚C for 10 mins. The

result was checked through agarose gel electrophoresis with appropriate control (positive and negative).

3.1.2.8. Plasmid DNA extraction (Brinboim and Doly, 1979)

Requirements:

1. Solution I
2. Solution II
3. Solution III
4. RNase solution:
5. Proteinase K

Procedure:

The colony PCR positive colonies were grown in LB media at 37°C for 12 to 14 hours with appropriate antibiotic selection. The overnight culture (5 ml) of *E.coli* containing the desire plasmid construct was transferred into a microcentrifuge tube and centrifuged at 12000 rpm for 2 minutes. The pellet was resuspended completely in 250 µl of Solution I and kept at room temperature for 30 minutes. 250 µl of Solution II was added and kept on ice for 5 minutes. To this, 350 µl of Solution III was added, mixed gently by inverting the tube upside down, and kept on ice for 30 minutes. After then centrifuged at 12,000 rpm for 10 minutes, the supernatant was collected and precipitated with 1ml of 100 % ethanol and incubated at -20°C for one hour. The mix was centrifuged at 12,000 rpm for 10 minutes. The pellet obtained was air-dried and resuspended in 100 µl of a solution containing 0.1M sodium acetate in 50 mM Tris.HCl at pH 8.0. The 95% ethanol precipitation was repeated. The pellet obtained was resuspended in 40 µl of RNases treatment. After complete resuspension, 10 µl of 1M Tris, pH 9.0, and 10 µl of 1M $MgCl_2$ were added and incubated for 30 minutes at 37°C. To this, 1 µl of proteinase K was added and incubated for one hour at 37°C. The mixture was extracted with equal volumes of phenol: chloroform: isoamyl alcohol mixture (25:24:1). The aqueous phase was taken into a new tube and added 1/10 volume of 3 M sodium acetate at pH 5.2 and 2.5 volumes of 95% ethanol. After precipitation at -20°C overnight, it was centrifuged at 12,000 rpm for 10 minutes. The pellet obtained was washed with 70 % ethanol, air-dried, and resuspended in 50 µl DNase-free water or TE buffer.

3.1.2.9. Recombinant clone confirmation by Restriction digestion

The extracted plasmid construct was confirmed through restriction digestion. The reaction mixture contains100-1000ng of plasmid DNA, 10X reaction buffer, 1µl of restriction enzyme(s) (MBI Fermatas. Inc), and DNase-free water.

3.1.2.10. DNA sequencing

The desired DNA construct was sequenced at the automated Sanger sequencing (Eurofins, Bangalore, India). To confirm the sequences of cloned genes, to remove the cloned vector sequence from the sequence with the use of Vecscreen software, National Center for Biotechnology Information (NCBI). After removing the vector sequence and the BLAST with nucleotide database search was performed in the NCBI database (http://www.ncbi.nih.gov/index.html).

3.1.2.11. Transformation of constructs into Agrobacterium

Requirement:

1.YEP medium

2.AB medium

3. K_2HPO_4 : 0.6 gm

4. NaH_2PO_4 , $2H_2O$: 2.6 gm

***Agrobacterium tumefaciens* strain:**

The *Agrobacterium* strain GV1301 was cultured at 28-30°C with the antibiotic selection of Rifampicin@ 15 µg/mL and Kanamycin @50 µg/mL. The glycerol stock (800 ml of fresh overnight culture + 200 ul sterile 80% glycerol) was stored at 80°C.

3.1.2.11.1. Preparation of electro-competent cells of *Agrobacterium*

A single colony of the *Agrobacterium* strain GV1301 was picked and inoculated in 5 ml of LB (or 2YT) in a 15 ml secure cap tube (culture tube) added appropriate antibiotic and then grown at 28°C for overnight in the continuous shaking @200rpm. Only 0.5 ml of the overnight culture was inoculated in 100 ml of LB in a 250 ml flask and grow

at 28°C until O.D. 600 reaches between 0.6 - 1.0. The culture was transferred to ice-cold sterilized 50 ml oak-ridge tubes and then incubated in ice for 30 min inside a 4°C refrigerator for chilling, followed by centrifugation at 5000 rpm for 10 min at 4°C. The supernatant was discarded and the pellet was resuspended with10 ml of ice-cold distilled H_2O. The cells were pelleted again by centrifugation at 4000 rpm for 10 minutes at 4°C. After repeating this step finally, the pellet was resuspended in 2 ml ice-cold 10% glycerol. 100µl of this bacterial suspension was dispensed into an ice-cold 1.5 ml microcentrifuge tube on ice and stored in -80˚C with liquid nitrogen pre-chilling.

3.1.2.11.2. Agrobacterium transformation *via* electroporation

0.5-1 µg of plasmid DNA constructs were added to the 100µl of the electro-competent cell of *Agrobacterium* and mixed by gently tapping with your finger; then incubated on ice (4˚C) for 20 min. The whole mixture content was then transferred into a prechilled sterilized electroporation cuvette. The water was removed properly from the outer electrode surface of the cuvette (0.2cm). The electroporation was performed carefully at 12.5 Kv/cm, 400 ohms for 20 milliseconds, using a Bio-Rad Gene Pulser. Immediately after successful electrophoration, 1ml of SOB broth was added to each cuvette. Then, the bacterial culture (1.1ml) was transferred into a 2ml cap tube/micro-centrifuge and incubates for 2 to 3 hours on a shaking condition at 28°C. The whole culture was finally plated on to a LA media with appropriate antibiotic selection and incubated for two days at 28°C. Transformed colonies were visible on the second day of incubation. Ultimately, the colonies were confirmed by colony PCR described above procedure.

3.1.2.12. Preparation of glycerol stock

E. coli and Agrobacterium culture containing the desirable plasmid constructs were grown at 37°C for 12 hrs either on LB media with proper antibiotic selection. All the liquid cultures were incubated with constant shaking at 200 rpm, and cells were harvested from the liquid culture by centrifuging at 6,000 rpm for 5 min. Permanent stocks of bacteria were made from overnight liquid cultures by adding glycerol to a final concentration of 80 % and were stored at -70°C.

3.1.2.13. Primer designing for different experimentation

All sequences/gene-specific primers were designed based on our requirements using the full-length genome of CGMMV (Accession no.: DQ767631) as reference. Mostly length of primers was varied with the kind of assay performed. Usually, 25nt long primers were designed for RT-PCR, whereas 35-40nt long primers were preferred for PCR-based cloning, mutation, and insertion. All the primers were designed with the help of OligoCalc (http://biotools.nubic.northwestern.edu/OligoCalc.html) (Kibbe, 2007) and primer3plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/ primer3plus. cgi/)web-tool (Untergasser *et al.*, 2007). These programs were used to calculate the Tm value, GC content, and also indicate the formation of secondary structures within and between primers or primer pairs. The positioning of G or C at the 3' end of the primers was preferred. Wherever necessary, the suitable restriction sites were added to the primer sequences for restriction based cloning. The primers were synthesized commercially (GCC biotech, India). For dissolving the lyophilized product, the prescribed (as prescribed by the synthesizing agency) amount of DNase-free water was added primers to prepare the stock solution of 100 µm concentration. From the stock solution, a 10µm working solution was prepared through dilution in Milli-Q water and used directly for PCR reaction. The stock and working solutions were stored at -20˚C. The detail about the primers is given in the following table 3.3.

Table 3.3. Details of the primers used in this research programme

Primer Name	Primer sequence (5'- 3') [a] Location		Nucleotide coordinate	Amplicon size (kb)
BM 96F	TGCTAAGCAAATCCCGTTTAT	RdRp	2943-2963	
BM 486F	***CGGGATCC***GTTTTAATTTTTAAAATTAAAC	5´UTR	1-22	6.4
BM 489R	***GCTCTAGA***TGGGCCCTTACCCAGG	3´UTR	6426-6442	
BM 962F	AGGCTACCACCTCGAAAGCTTAGGGAGGGCAATCTTATGTTGAAGC	CP+NbPDS	6226-6248	0.227
BM 963R	ACCACCATCAGAAGACCCTCGAAAGGAGGGTTACCATCTAAAAAGG	UTR+NbPDS	6249-6272	
BM 979F	GGTGATGATGCTCTGGTTGACACAAGGATGCATTCT	MP	5272-5299	2.0Kb (DG-1)
BM 980R	TCAACCAGAGCATCATCACCATCGACCCTAAACTGA	MT	802-834	
BM 1171F	GGTGATGATTCGCTGATCTACCTTCC	RdRp	4591- 4616	1415nt (in BP4)
BM 1182R	ATTGGGGCTCAGTTTAGGGTCGATGG	MT	802-827	768 nt
BM 1173R	GAAGATAGGCCTCAACACAGGACCG	CP	5981-6005	(in DG1)
BM 1178F	ATG GCT TAC AAT CCG ATC ACA CC	CP	5763-5785	0.486
BM 1179R	AGC TTT CGA GGT GGT AGC CTC	CP	6245-6225	
BM 1180R	AGCCGCGTTTTCGGTAGTCTGG	CP	6200-6221	0.197
BM 1269R	GCATCGAAAGCTCGTCAGGG	*NbPDS*	129-148	
BM 1265F	GCCAACACTTGTCACTACTTTCTC	GFP	183-206	0.23
BM 1266R	GATGTTTCCGTCCTCCTTGAAATC	GFP	397-420	
BM 849F	GGTATTGTGTTGGACTCGGG	*ACT1*	219-238	0.16
BM 850R	GCTGTGGTAGTGGATGAGTAAC	*ACT1*	361-382	
BM 1267F	GCCTTCCTGGGAATTCCTTCG	RdRp	3593-3613	197
BM 1268R	GAATGCGCGGCCTTTCACAG	RdRp	3765-3784	

*****_GGATCC_ is the restriction site of BamH1, _TCTAGA_ is the restriction site of XbaI. The primer set (BM 979F and BM 980R) is used for the deconstruction of CGMMV genome, whereas the primer set (BM-962 & BM-963R) is used for the insertion of the *NbPDS* gene into the DG-2 construct, RT-PCR based detection was carried out by using terminal primer set (BM-486F & BM-489R).

3.2. To evaluate the trans-replicability and translation ability of deconstructed genome in presence of wild type CGMMV

3.2.1. In planta replication and movement of the deconstructed genome of CGMMV in the different hosts using their infectious clones.

3.2.1.1.Plant material and sample collection

Nicotiana benthamiana and bottle guard were grown in the mixer of the coco peat and perlite (soil rate). Plants were grown in a 6-inch pot in proper glasshouse condition 16 hrs daylight and 8 hrs night with 25 - 30 °C temperature. *Cucumber green mottle mosaic virus* (CGMMV), greenish mottle mosaic symptomatic lyophilized tissue powder was inoculated through the sap inoculation method. The powder of virus culture used for sap inoculation which was prepared by mixing one gram of CGMMV lyophilized culture in 10 ml of 0.1 M phosphate buffer, pH 7.2 (Supplemented with 0.15% sodium sulphite) in a pestle and mortar on ice and 0.2% of celite (BDH545) was added in the inoculums. Two weeks old fully expanded leaves were dusted with carborundum on the upper leaves and inoculated by rubbing method with the help of cotton soaked with inoculums. After inoculation plants were kept for 30 minutes in the chamber while inoculums were dried after inoculated plants were washed with distilled water and maintained under the greenhouse condition for CGMMV infection development. The greenish mottle mosaic symptoms were developed at 7 days of post-inoculation (DPI) after that virus culture was maintained regularly in the *N. benthamiana* plants in proper glasshouse condition.

3.2.1.2.Plant inoculation of deconstructed virus genome along with wild type virus strains

The cDNA construct of deletion mutant (pDG-1), as well as wild type CGMMV (pBP-4), were mobilized into the highly efficient competent cells of *Agrobacterium tumefaciens* strain GV3101 by freeze-thaw method (Weigel and Glazebrook, 2006). 1-2 µg of plasmid construct was mixed with 200µl of Agro competent cells and incubated in the ice bath for 20 min. The agro-culture containing plasmid mixture was frozen with liquid nitrogen for 20 sec followed by thawing at 37 °C for 5 min, and again immediately incubated in ice (0°C) for 20 min. Thereafter, 5ml of LB medium was added and the culture was incubated at 28 °C for 3 hours with continuous shaking (220 rpm). Then, the transformed agro cells were plated with the appropriate antibiotic-

containing plates (kanamycin and rifamycin) and incubated at 28 °C. The transformed colonies have further confirmed based on the presence of recombinant clone in agro-culture through the colony PCR with gene-specific primers. The confirmed colonies were picked and grown separately in 5ml of LB broth consisting of 50 µg/ml kanamycin and 15 µg/ml rifampicin at 28°C for 18-24h with 200 rpm. The overnight agro-culture was centrifuged at 5000 rpm for 10 mins and re-suspended in 50 ml of agroinfiltration buffer containing 10 mM 2-(*N*-morpholino) ethane sulfonic acid **(**MES, pH 5.7), 10 mM $MgCl_2$,150 µM Acetosyringonein order to obtain a final O.D. of 0.6 infiltration culture. The resuspended agro-culture was additionally incubated at 28°C for 1-3hr with 50 rpm. After incubation, the agro-culture consisting of pBP4 construct was infiltrated into 3-weeks old plants of *N. benthamiana, via* syringe infiltration on the abaxial leaf surface using 2 ml sterile syringe. The same was repeated in the cotyledonary leaves of 1-week old cucumber, bottle gourd, and watermelon seedlings. To confirm the trans-replication of the deletion mutant DG-1 (cDNA clone) of CGMMV in different host plants, its agrobacterium culture was either co-infiltrated (mixed at 1:1 ratio) simultaneously along with the wild type CGMMV (pBP4), or post-infiltrated (after 7DPI) in the CGMMV infected host plants. The infiltrated plants were maintained at the glasshouse with 16/8hr daylight conditions at 25˚C, and observation was recorded. Each experiment was repeated thrice for the replication and robustness of the data.

3.2.1.4. Extraction of total RNA from plant

3.2.1.4.1 TRIZOL method

CGMMV infected leaf samples were crushed in liquid nitrogen with a help of mortar and pestle and then add 1 ml Trizol reagent (Invitrogen, USA). After crushed, the leaf sample was centrifuged the total extract at 12,000×g for 10 minutes at 4˚C. The supernatant contains RNA and protein. After centrifugation transfer the upper layer of supernatant to a new centrifuge tube. Add 0.2 ml of chloroform per 1 ml of Tri-Reagent used for homogenization. Shook vigorously for 15 seconds and allowed standing for 2-3 minutes at room temperature. Centrifuged the resulting mixture at 12,000×g for 15 minutes at 4˚C (The mixture separates into a lower red phenol-chloroform phase, interphase, and a colourless upper aqueous phase. RNA remains exclusively in the aqueous phase. The upper aqueous phase is 50% of the total volume). Transfer the aqueous phase to a fresh tube and added 0.5 ml of isopropanol per ml of Tri-Reagent

used in sample preparation. Incubated the sample to stand for 5-10 minutes at room temperature. After incubation centrifuged the sample at 12,000×g for 10 minutes at 4˚C. Removed the supernatant and dry well and washed the RNA pellet by adding 1 ml of 75% ethanol per 1 ml of Tri-Reagent. Vortexed the sample and then centrifuged at 7,500×g for 5 minutes at 4˚C. Dried the RNA pellet for 5-10 mins by air drying or under a vacuum. Resuspend RNA in 50 µl RNAse free water.

3.2.1.4.2 Plant RNA extraction kit

The total plant RNA was extracted from the infected plant leaves by using the Promega SV total RNA isolation system (Promega, USA). Weighed the 100 mg of tissue was homogenized with liquid nitrogen with the help of DEPC treated and autoclaved motor and pestle. The homogenization of tissues of approximately 20 mg powder was used for added 1ml of lysis buffer contains beta-mercaptoethanol and added 350µl of RNA Dilution Buffer (blue). Mixed by inverting 3-4 times. Placed in a heating block at 70°C for 3 minutes. Incubation for longer than 3 minutes may result in the compromised integrity of the RNA. After incubation extracts were centrifuged at 14000xg for 10 minutes at room temperature. The supernatant was transferred to a fresh tube and added 200µl of 95 % ethanol to the supernatant and mixed gently inverting 3-5 times. Transfer the extract to the spin column, centrifuged at 14000xg for one minute. Then discarded the flow throw after centrifuged then added 600 µl RNA wash solution to the column and centrifuged at 14000xg for one minute. Prepared the DNase treatment through an incubation mix by combining 40µl Yellow Core Buffer, 5µl 0.09M $MnCl_2$, and 5µl of a DNase I enzyme. Applied 50µl on this freshly prepared DNase incubation mix directly to the membrane and kept the column in 20-25˚C for 20 minutes. Stop the DNase treatment with the help of a stop solution to add 200 µl in the column and centrifuged at 14000xg for one minute. Washed again with wash solution 600 µl and 250 µl, centrifuged at 14000xg for one minute. To dry the column centrifuged at 14000xg for 2 minutes and finally eluted the RNA, 50 µl elution buffer added to the centre of the column and centrifuged at 14000xg for one minute.

3.2.1.5. The infectivity assay of wild CGMMV

The pBP-4 infected leaf samples were harvested at 7 dpi for confirmation of virus infection. The total plant RNA was isolated from the symptomatic leaves using a plant

total RNA isolation kit (Promega, USA) as per the manufactures' protocol. For cDNA synthesis, the total RNA was used as a template for reverse transcriptase reaction using Superscripts III, (Invitrogen, USA) using CGMMV specific 3´ terminal primer (BM489R) as per manufactures described the protocol. The infectivity of the wild type virus was detected through RT-PCR using the CP gene-specific primers (BM-1178F and BM-1179R). Following the infection of wild CGMMV, the deletion mutant was infiltrated in the case of post-infiltration.

3.2.1.6. Trans-replication assay of the deletion mutant of CGMMV

To detect the trans-replication of the DG-1, the RNA was extracted from the infiltrated leaves as well as the systemic leaves of the different host plants at a different time interval (7dpi, 14dpi) and subsequently, cDNA was synthesized as per the described above protocol. The trans-replication of the DG-1 was detected through RT-PCR using the virus-specific terminal primers (BM-486F and BM-489R). For further confirmation, another RT-PCR based detection approach was adopted using virus-specific internal primer pairs (BM-1182F and BM-1173R), where the forward primer was designed from the flanking sequences (802-827of DG-1 & BP4) located just before the junction/deleted site (831nt) of the DG-1 and the reverse primer was taken from the conserved sequence located at the 5981-6005nt of BP4 and DG-1(i.e., CP gene of CGMMV). The confirmatory RT-PCR was also carried out using BM-486F and BM-489R to detect the replication rate of the DG-1 replicon.

3.2.1.7. Movement assay of a deletion mutant of CGMMV

The *in planta* systemic movement of the dRNA of the CGMMV was traced out in different temporal and spatial scales. A total of ten CGMMV infected plants was agro-infiltrated with DG-1, and leaf samples were harvested at different temporal duration (i.e., from 1dpi at 24 hrs interval) and spatial scale (from infiltrated leaves to upper systemic leaves). The RNA was extracted and cDNA was prepared as per the manufacturers' protocol. The RT-PCR based detection of the DG-1 is performed using the terminal primers BM-486F and BM-489R. The systemic moment on different leaves was checked at 14dpi to predict its distant and wider distribution within the plant. The experiment was replicated thrice.

3.2.1.8. Insertion of *NbPDS* gene into DG-1 construct, agro-transformation, and plant infiltration

A partial mRNA sequence (~ 228 nt) of the phytoene desaturase (PDS) gene of *N. benthamiana* (NCBI accession number: EU165355) was cloned at the end of CP region of DG-2 construct in the sense orientation. For that, restriction and ligation-free OPER cloning technique was adopted. The long forward (BM 962F) and reverse primers (BM 963R) were designed keeping the flanking sequences of both CP and PDS gene and that primer pairs were used for the amplification of *NbPDS* gene segments in the first round of PCR.

The first round of PCR was performed based on three steps PCR protocol (i.e., 35 cycles of denaturation at 94°C for 10 sec, primer annealing at 52°C for 30 sec, and synthesis at 72°C for 20 sec followed by one cycle of final extension at 72°C for 10 min) using a reaction mix containing 5 µl of 5x reaction buffer, 2.0 µl of 2.5 mM dNTP mix, 1.0 µl each of forward and reverses primers (1.0 micromoles), 1.0 U of Phusion *Taq* polymerase (NEB, USA) and DNase-free water to make up the volume of 25 µl reaction mixtures. The amplified primary PCR product was purified from the gel as per manufacturer protocol. In the second round of PCR, the amplified *NbPDS* gene segment (-228 nt) was used as long primers for the insertion into DG-2 construct (template) using two-step PCR protocol (consisting of 20 cycles of denaturing at 94°C for 10 sec, and primer annealing with synthesis at 65°C for 50 sec/kb followed by one cycle of final extension at 65°C for 10 min). After completion of PCR, the PCR product was digested with DpnI enzyme at 37˚C for 2 hrs followed by enzyme inhibition at 80˚C for 20 mins. The 5ul of DpnI treated PCR product was directly transformed into JM109 chemically prepared competent cells following standard transformation protocol. The recombinant colonies have further confirmed the presence of the *NbPDS* gene at the end of CP by using a gene-specific primer (BM-1178F & BM-489R) as well as sanger based primer sequencing using CGMMV 3' UTR primer (BM-489R). Finally, DG(PDS)-1 construct was transferred into *Agrobacterium via* electroporation methods. The agro mobilized constructs were infiltrated into CGMMV infected *N. benthamiana* plants using the above-mentioned procedure.

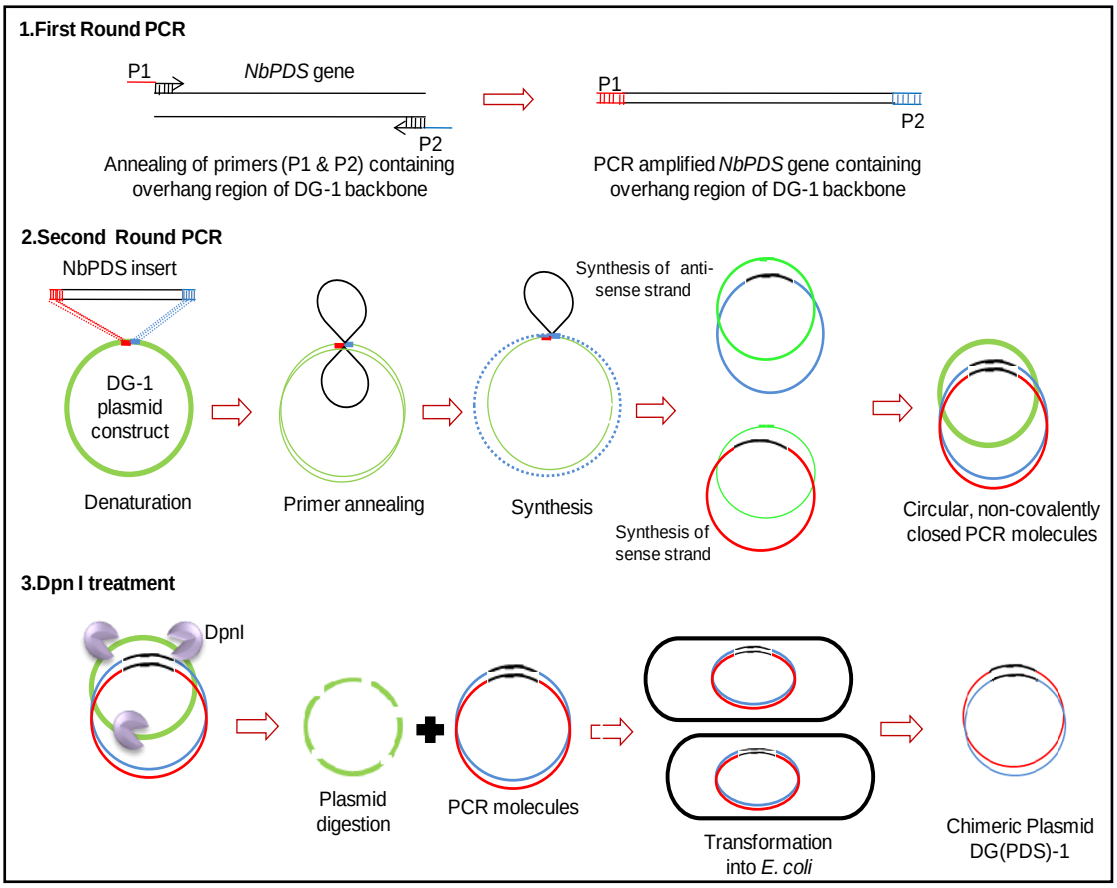


Fig. 3.2. Schematic diagram of OPER restriction and ligation-free cloning for the insertion of the foreign gene (*NbPDS*) into any larger plasmid construct (DG-1).

Step1: First round of PCR was performed to amplify the insert **(*NbPDS*)** with the forward (P1) and reverse primer (P2) containing the overlapping sequences of backbone vector (DG-1). Step2: The second round of PCR was performed to insert *NbPDS* gene within DG-1 vector backbone using first PCR amplicon as the primer, and Step 3: the second PCR product with native plasmid was treated with DpnI enzyme for the specific digestion of native plasmid (DG-1) and followed by transformation into *E.coli* to facilitate *in vivo* gap sealing of the circular, non-covalently closed PCR molecules. The final chimeric plasmid DG(PDS)-1 containing the *NbPDS* gene in the sense orientation was used as the virus-induced gene silencing vector.

3.2.1.9. Functional validation of DG(PDS)-1 construct

The phenotypic expression of gene silencing was recorded during the 15-60DPI period and observation was started from the appearance of photobleaching symptoms. The efficiency of the gene silencing vector was estimated based on the percent plant showing of photobleaching symptoms. The infiltrated plants were tested for the confirmation of replication and movement of DG(PDS)-1 construct using RT-PCR. A

duplex RT-PCR was standardized using two pairs of primers; one set contains BM 1180R and BM 1269R for the specific detection of DG(PDS)-1, whereas the other contains BM-96F and BM-1268R for specific detection of helper CGMMV, simultaneously in the same PCR reaction.

3.2.1.10. Quantification of the replication rate of the DG(PDS)-1 construct

To quantify the level of replication of the deletion mutants of CGMMV, quantitative PCR was performed. For that, DG-1 was agro-infiltrated into the previously infected *N. benthamiana*. A total of ten plants were chosen based on their similar age having the same growth pattern. The 5th leaf of the infected plants was selected for the agroinfiltration of DG-1. The small leaf disc (5mm) was cut from the infiltrated part of all ten plants regularly at 24 hrs interval starting from 1st DPI and pulled together for RNA extraction. The RNA was extracted as per the previously described protocol, quantified, and treated with RNase-free on-column DNaseI to remove the influence of the infiltrated plasmid construct. The RNA samples were then used for synthesizing the first-strand cDNAs with an oligo(dT)20 primer using with Superscripts III, (Invitrogen, USA). The RT-PCR was performed to confirm the infectivity of CGMMVinfectious clones in plants by using the internal specific primers BM 1180R and BM 1269R (Table 3.3.). The previously described PCR reaction mixture and program condition were used.

3.2.1.11. Northern blot analysis for the confirmation of the deletion RNA replicon

The RNA extracted from different experimental samples was used for the northern blot analysis. The study was performed using DIG High Prime DNA Labeling and Detection Starter Kit II (Roche Applied Science, Germany). The following materials and methods are required.

Requirements:

1.10X MOPS buffer (pH7.0)

2. 2X RNA loading buffer

3. 20X SSC buffer

4. 10X Transcription buffer

5. 10X Wash buffer

Procedures:

3.2.1.11.1.Preparation of hybridization probe

The positive-strand RNA genome-specific DNA probe was generated using the DIG-High Prime kit. For that, the targeted DNA (~225nt) that was complementary to 3' UTR sequence, was amplified from the 3' terminal of CGMMV plasmid construct (BP4) using 1180F and 489R primer. The PCR amplified products were then eluted from the gel and purified. 1µg of DNA was then added in 15µl in the nuclease-free water and denatured in boiling water for 10min; thereafter immediately kept on ice. Then 4µlDIG high prime was added to it and incubated at 37˚C for 12 hrs. 2 µl of 0.2M EDTA (pH8.0) was mixed to stop the reaction. This digoxigenin-labeled probe was used for hybridization to a membrane blotted RNA.

3.2.1.11.2. Testing of hybridization probe

3.2.1.11.2.1. Preparation of denaturalizing 1.2% Formaldehyde-Agarose Gel

For prepared 1.2% agarose gel for RNA separation, the formaldehyde was used in the gel and running buffers, as well as a sample heating step, formaldehyde prevents the RNA to form the secondary structures, thus allowing RNA to run properly on the gel. 0.6 g of agarose was added to 49 ml of 1x RNA running buffer and boiled well to dissolve the entire agarose, allow the gel to cool about 40-50°C. Add 1ml of 37% formaldehyde in gel and add 2-3 drops of easy version dye (Ameresco, USA) and mixed well to poured gel casting tray. After the gel was solidified, it was placed in the running buffer for 10mins to equilibrate it and made ready for loading the RNA samples.

3.2.1.11.2.2. RNA preparation and gel loading

2-5 µg RNA samples were mixed with an equal volume of 2X RNA loading buffer (Thermo Scientific, USA). Then, the samples were heated at 65°C in the heating blocking for 10 mins for removing the secondary structures; after heating, samples were placed on ice immediately, afterward. Finally, samples were loaded onto the gel along high range RNA (6 kb) ladder (Thermo Scientific, USA).

3.2.1.11.2.3. Gel electrophoresis in 10X MOPS buffer

The samples were run in gel with 45 V for 3hrs with the ice-cold condition to avoid the chance of RNA degradation. The 1X MOPS buffer is used for running.

3.2.1.11.2.4. Transformation of RNA from gel to membrane *via* wet blotting

After the gel run is finished, the RNA gel was rinsed twice with DEPC treated water- H_2O; then equilibrated it in 20X SSC buffer for 5 min. In the meantime, the nylon membrane (1 piece) and filter (6 pieces) were cut about the size of the gel and wet the filter paper in the transfer buffer. A bed was prepared with a continuous supply of 20X SSC buffer *via* filter membrane. Three pre-soaked sheets were placed on the bed. The pre-wetted gel was placed on top of the filter paper and rolled with a pipet or test tube or roller over the surface of the gel to exclude all air bubbles. Then, the dry nylon membrane was placed carefully on the top of the gel, aligning the gel size; again rolled with a pipet or test tube, or roller over the surface of the membrane to exclude all air bubbles. Extra buffer was added onto the membrane to avoid bubble formation. The other pre-soaked sheets of filter papers were placed onto the gel. Some blotting sheets were placed on the top carefully followed by some weight and kept it for 8-10 hrs at room temperature. The RNA was transferred from gel to membrane *via* the capillary movement of salt solution.

3.2.1.11.2.5. UV-crosslinking

After the successful RNA transfer, immediately the membrane was UV-crosslinked without prior washing. The membrane was exposed to UV-ray (250-300 nm wavelength) for 2min in a UV-cross linker/ UV- transilluminator; the step was repeated thrice. The shortwave of UV rays leads to the activation of the nitrogenous bases in RNA, mostly uracil, and helped in the formation of covalent linkages to amine groups of the nylon membrane surface. After crosslinking, the membrane can be stored dry at +4°C or proceeded for prehybridization.

3.2.1.11.2.6. Prehybridization

For the prehybridization, the DIG Easy Hyb solution was prewarmed at hybridization temperature (68° C) for 10 min; then the membrane was prehybridized with the RNA side facing inwards with prewarmed DIG Easy Hyb for 30 min with gentle shaking in hybridization flask.

3.2.1.11.2.7. Probe Hybridization.

1µl of the DIG-labelled RNA probe was denatured by boiling in the hot water bath for 10 min and then immediately kept in ice for cooling for 5min. Finally, a denatured DIG-labelled RNA probe was mixed with the new hybridization buffer and finally added into the hybridization bottle containing the membrane. The prehybridization

buffer should be removed before adding the probe. Then, it was incubated for 12 hrs or overnight at 45˚C for RNA: DNA hybridization. Continuous rotation of the hybridization bottle was required for proper hybridization. The temperature may differ from one probe to another.

3.2.1.11.2.8. Washing

After the probe hybridization, the membrane was washed properly with low and high stringency buffer. At first, low stringency wash was done with the solution containing 2X SSC, and 0.1% SDS at room temperature for 15 mins with continuous shaking to remove the unbound probes; the step was repeated twice. This was followed by high stringency wash in 0.1-0.5X SSC and 0.1% SDS solution for 15 min at 68°C to remove the non-specifically bound probes; this step was also repeated twice. After the stringency wash, the membrane was rinsed with a normal wash buffer containing 1X 0.1 (M) Maleic acid solution and 1% tween-20 for 15 min at room temperature. This step was also repeated thrice.

3.2.1.11.2.9. Blocking of membrane

To block the unspecific sites on the membrane, the membrane was treated with a 1X blocking solution, prepared by diluting 1 ml of 10 × Blocking solution with 9ml Maleic acid buffer (in 1:10 ratio) and incubated for 30 mins. After the blocking step, the membrane was washed with the washing solution.

3.2.1.11.2.10. Immunological treatment

For the specific detection of probe, 0.1µl of anti-digoxigenin-AP containing polyclonal sheep anti-digoxigenin, $F(ab)_2$-fragments conjugated alkaline phosphatase (AP) was diluted in10 ml blocking solution and added directly on the membrane; then incubated for 30 mins. After antibody treatment, the membrane was washed with 20ml of washing solution containing maleic acid at room temperature for 15 min. The step was repeated twice.

3.2.1.11.2.11. Preparation for detection

To bring into equilibrium condition, the membrane was treated with10 ml of detection buffer for 5-10 min in RT.

3.2.1.11.2.12. Immunological detection

The hybridized probes were immuno-detected with anti-digoxigenin-AP. Membrane (with RNA side facing up) was treated with 400µl of CDP-Star and rolled properly to spread the substrate evenly and to remove the air bubbles over the membrane. Then, it was placed on a development folder by covering it with a polythene sheet. One X-ray

film was also placed simultaneously, and immediately the folder was closed. It was incubated for 30 min -2 hours at + 15 to +25° C. The enzymatic dephosphorylation of CSPD by alkaline phosphatase leads to light emission at a maximum wavelength of 477 nm which is recorded on X-ray films. The film was developed with the treatment of a developer and a fixer solution.

3.2.2. Evaluation of the translational ability of the deconstructed genome of CGMMV

To study the translational ability of the deconstructed genome of CGMMV, the expression of GFP protein was analyzed. A full-length sequence of the GFP gene was inserted into the coat protein region of the deconstructed genome(s) of CGMMV. These constructs will be introduced into the *Agrobacterium* strain. The *Agrobacterium* suspension will be injected using a 1mL needleless syringe into cotyledons of cucurbits and true leaves of *N. benthamiana*. After co-inoculation with the CGMMV infectious clones, the protein expression levels of these constructs will be determined using real-time PCR and western blot.

3.2.2.1. Agro-infiltration of DG-2 construct in the CGMMV infected *N. benthamiana*

Inoculation Buffer

MES	: 10mM (Filter sterilized)
MgCl2	: 10mM (Autoclaved)
Acetosyringone	: 150µM

The *Agrobacterium* GV1301 harboring the DG-2 plasmid constructs were inoculated into LB liquid medium containing Rifampicin (15µg/ml) and Kanamycin (50µg/ml) and grown for 12-16 hours at 28°C with 200 rpm. From this 1 ml of culture was added to 30 ml inoculation buffer to set final O.D. 0.6 and grown for 2-3 hours at the same temperature with shaking (100 rpm). The induced *Agrobacterium* cells were taken in a 2 ml syringe and infiltrated into the abaxial side of the CGMMV infected *N. benthamiana* leaves without a needle. The agroinfiltrated plants were kept in a growth chamber and maintained condition 16 hrs daylight and 8 hrs nights with 25-30 °C temperature.

3.2.2.2. Detection of replication of DG-2 construct in the CGMMV infected *N. benthamiana*

The trans-replication and movement of DG-2 were detected using specific primers, i.e., BM-1265F & BM-1266R. RNA extracted from infiltrated as well as systemic tissue was used for testing. Simultaneous detection of the helper CGMMV was also carried out using 96F and 1268R primer. The PCR protocol was standardized with denaturation at 94˚C for 3 min, 35 cycles of final denaturation at 94˚C for 30 sec., annealing at 60˚C for 30 sec., extension at 72˚C for 30 sec., followed by one cycle of final extension at72˚C for 10 min. The confirmation was made based on gel electrophoresis.

3.2.2.3. Quantitative RT-PCR (qPCR)

To quantify the GFP transcripts in inoculated plants as well as control healthy transcripts qPCR was performed. The total RNA was extracted from 100 mg of tissue of plants using plant RNA extraction kit (SV total RNA isolation system, Promega, USA). The qPCR was performed using specific primers to GFP (BM-1265F & BM-1266R) and actin1 (as an internal control) from *N. benthamiana* (BM-849F & BM-850R) with the Sybr green master mix from KAPA SYBR FAST qPCR Kits (KAPA biosystem, USA). The thermal cycling program for both their primers was 95°C for 1 min, followed by 45 cycles at 95°C for 20 Sec, 60°C for 20 sec, and 72°C for 20 Sec. The reaction was performed in a qPCR machine Light Cycler 96 SW (Roche, USA).

3.2.2.4. Detection of GFP Fluorescence through confocal microscopy

Primarily CGMMV infected plants were chosen for the agro-infiltration of the DG-2 construct. From 2DPI onwards, plants were analyzed for GFP expression through confocal microscopy. Since then at the periodical observation of GFP expression was taken at 2 days interval. The DG-2 infiltrated plants were showing GFP expression in the inoculated area. Only CGMMV infected plants, DG-2 infiltrated plants, and PBS buffer inoculated (mock) plants were chosen as the control. GFP expression was confirmed by confocal microscopic the wavelength of 490-510 nm.

3.2.2.5. Plant protein extraction

Inoculated *N. benthamiana* and systematic leaf tissue (0.8 g) were ground to a fine powder in liquid nitrogen using an autoclaved pestle and mortar in a cold room (pre-chilled the mortar and pestle at -70°C). Dispensed the tissue powder into a centrifuge

tube with the help of a spatula. The powder was resuspended by inversion mixing in 300 µl of protein extraction buffer (0.1 M KH_2PO_4, pH 6.5, 0.5 mM PMSF, and 10mM β-mercaptoethanol). Extracts were centrifuged for 20 mins at 15,000 rpm at 4°C. Removed the supernatant was stored at -20°C.

3.2.2.6. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli, 1970)

Requirements:

Acrylamide bisacrylamide solution (30 %)

Separating gel (10 %) for 10 ml

Stacking gel (4 %)

Sample buffer (2X)

Running buffer (pH 8.3)

Staining solution

Destaining solution

A separating gel was cast in between two gel plates and a small volume of stacking gel was poured over it and an appropriate comb was placed and kept for polymerization. Plant protein samples were prepared an equal volume of the sample buffer and kept for 5 mins at boiling water. Electrophoresis (Bio-rad, USA) was done at 100 V, till the dye front reached the bottom of the gel. After electrophoresis, the gel was kept in the staining solution for 2 hr and the excess stain was removed by keeping in the de-staining solution till the bands were clear.

3.2.2.7. Western blotting (Towbin *et al.*, 1979).

Towbin's buffer (Transfer buffer)

Tris HCl	: 0.25 M
Glycine	: 0.192 M
Methanol	: 20 %
SDS	: 0.03 %

TBS (pH 7.5)

Tris HCl : 0.1 M

NaCl : 0.9 %

TTBS

Tween 20 : 0.1 % in TBS

Blocking solution

Gelatine : 5 % in TBS

Diluting solution

Gelatin : 3 % in TBS

Procedure:

The protein samples were separated by SDS-PAGE and the plant proteins were transferred to a nitrocellulose membrane (Sartorius, Germany) using an ATTO, Japan dry electroblotting apparatus at 100 V for 45 min. The membrane was incubated in the blocking solution overnight at 4°C and washed twice with TTBS and once with TBS. The blot was then treated with primary antiserum with appropriate dilution for 1 hrs at 37°C. The membrane was then washed three times with TTBS and once with TBS. The blot was then treated with the secondary antibody, goat anti-rabbit IgG conjugated with alkaline phosphates' (AP). After two washes with TTBS and once with TBS the reaction was detected by adding the substrate (BCIP/NBT).

4.1. Objective: To create deconstructed genome of CGMMV with *cis*-regulatory sequence

4.1.1. Identification of the cis-regulatory sequences for the genome replication and translation

4.1.1.1. Genome-wide prediction of regulatory elements in the CGMMV genome

4.1.1.1.1. Mining the regulatory elements in the CGMMV genome from Pfam and Rfam database

Based on our Rfam database search, a stretch of pseudoknots (RF01072) forming upstream pseudoknot domain (UPD) was identified in the 6289-6319 nucleotides (nt) region of the 3'UTR of CGMMV. This RNA element was reported to be similar to other positive-stranded RNA viruses, like tobacco mosaic virus, turnip yellow mosaic virus, beet virus Q, barley stripe mosaic virus, etc, and which is thought to be necessary for efficient transcription and translational. Further, a search on the protein family (Pfam) database, various domains of the complex replicase enzyme was classified. Mainly three domains, viz., viral methyl-transferase (pfam01660), Viral (Superfamily 1) RNA helicase 1 (pfam01443), and RNA dependent RNA polymerase (RdRP) (pfam00978) were identified and found to be spanned within 50-426aa, 860-1113aa, 1196-1642aa, respectively (**Table 4.1.**). A highly conserved UvrD motif of DEAD-like helicase superfamily was also identified within 848-945aa of the helicase domain, and the core RNA-dependent RNA polymerase (RdRp) domain was distributed within 1407-1524aa of RNA dependent RNA polymerase.

4.1.1.1.2. Prediction of the RNA-secondary structure of CGMMV genome

For the prediction of putative secondary regulatory elements, the 6424 nt long genome sequence of CGMMV was submitted to Mfold and ScanFold web-server. The predicted structures which contain at least one base pair with Zavg score < -2 have been extracted from results that ultimately identified 27 Structural elements with their secondary structure and respective Z-score and minimum free energy value. The same was further analyzed by using Dot-Knot (https://dotknot.csse.uwa.edu.au/) with 200 nt window size at 37˚C to validate the elements. Numerous secondary elements were identified

within the CGMMV genome (**Fig. 4.1**). These structures are the only indicators of the complex genome; but not enough for specifying their functional significance.

Table 4.1. Predicted domains of replicase enzyme complex from the protein family database

Domain*	Accession	Description	Amino acid (aa) position	E-value
Vmethyltransf	pfam01660	Viral methyltransferase; This RNA methyltransferase domain is found in a wide range of ssRNA	50-426	3.77e-44
Viral helicase	pfam01443	Viral (Superfamily 1) RNA helicase.	860-1113	1.63e-50
DEXQc UvrD	cd17932	DEXQD-box helicase domain of UvrD which is a highly conserved helicase.	848-945	2.14e-03
RdRP	pfam00978	RNA dependent RNA polymerase	1196-1642	5.22e-141
Core RdRp	cd01699	Core RNA-dependent RNA polymerase (RdRp) domain is essential for the catalytic activity	1407-1524	9.97e-04

*****The replicase enzyme of CGMMV contains three major domains, *viz*., methyltransferase, helicase, RNA dependent RNA polymerase. The helicase contains a DEXQD-box type catalytic site. The catalytic sites RdRp is situated in its core region. The information was collected based on the Pfam search of amino acid sequences.

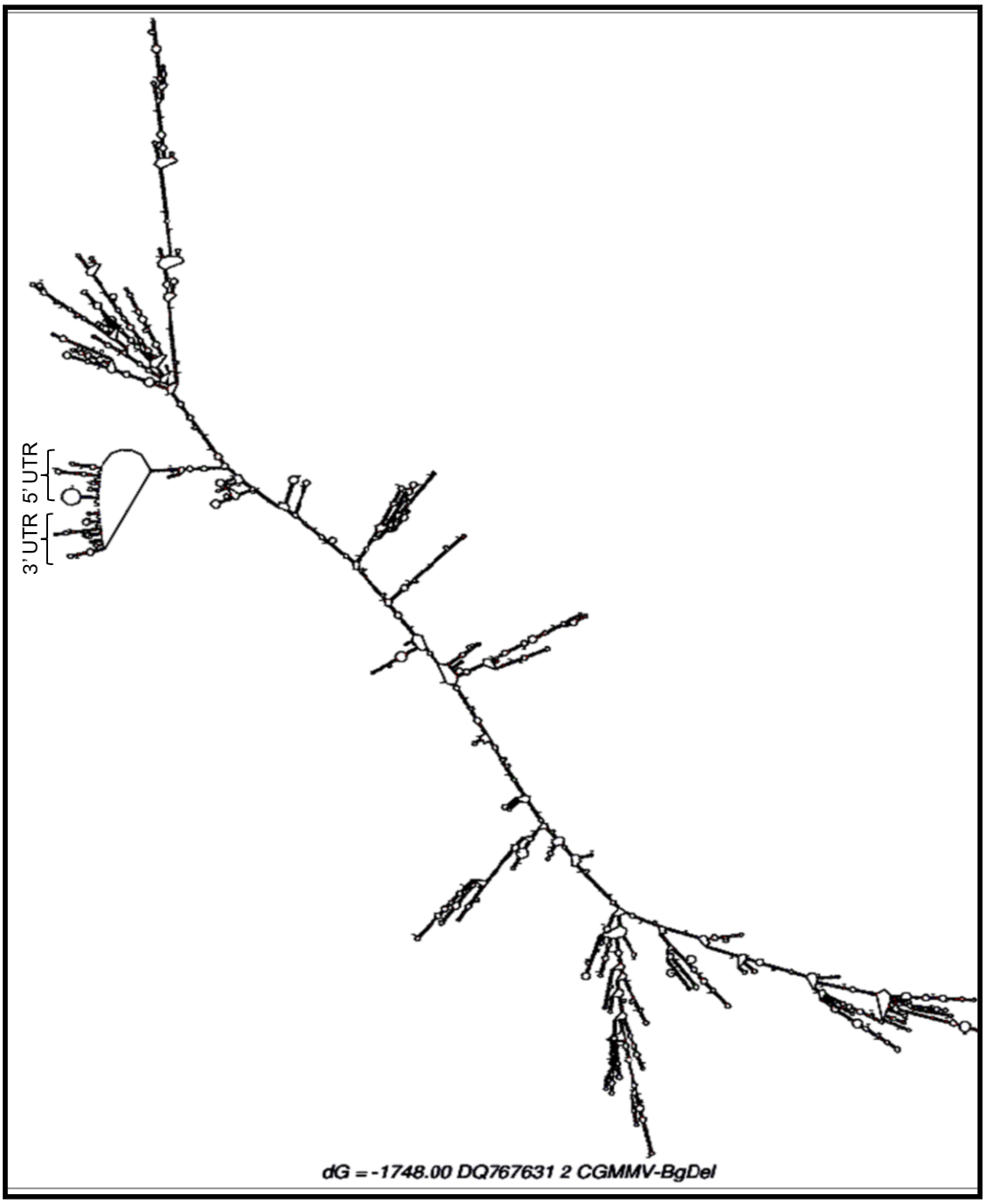


Fig.4.1. The RNA secondary structure of the CGMMV genome (6424 nt) is generated by using Mfold web server with minimum free energy. It depicts the numerous secondary elements are present in the genome with conserved 5' and 3' UTRs. The stability and structural conservation of these elements are essential to recognise as the potential one.

4.1.1.2. *In silico* prediction of replicase enzyme binding sites at the 5' and 3' terminal end of the genome

4.1.1.2.1. Comparative analysis of terminal regions of CGMMV

The alignment and comparison of 95nt of 5'-terminal ends of 105 isolates of CGMMV show a high degree of conservation among the isolates. The sequence diversity is varying from 0-0.098% with an overall diversity of 0.016%. A stretch of conserved $(AAAC)_3$-$(AAC)_5$ compound perfect repeats are existed within the first 19-41 nt of the UTR; $(AAC)_2$ again repeated twice from 54-59 nt. The similar $(CAA)_3$ repeats are also found within 84-94 nt, but in the imperfect form. Similarly, the 3' terminal ends of CGMMV isolates are also highly conserved in terms of their sequence similarity. Only 0-0.094% variability with an overall mean variation of 0.032% was recorded. It also contains some small compound repeat sequences, *viz*., $(UCU)_2$-$(UGG)_2$ located within 6258-6271nt followed by a hexa-repeats of C_6 (6398-6403). These simple and compound sequence repeats have been found in all isolates of CGMMV. These microsatellites provide excellent evidence of evolution and assessing genetic variation within the population.

4.1.1.2.2. Secondary structures of 5' and 3' UTRs of CGMMV

To identify potential regulatory regions in the highly conserved 5'-and 3'-terminal ends of the CGMMV genome, some typical RNA secondary structures were predicted. The most initial 74nt region of the 5' terminal end was predicted to contain two typical stem-loop (SL) structures **(Fig. 4.2)**. Of them, SL-1 is very large containing a contiguous stretch of ~45 nt that were almost conserved, representing a long stem with the canonical A-U intramolecular base pairing and a large loop containing the conserved AAAC and AAC repeats. This G-deficient hairpin structure at 5- UTR region of CGMMV genome can be considered to be the equivalent to omega ladder of TMV. This ladder is followed by a small SL structure, i.e., SL-2 which was identified within the 60-72 nt region of 5'-UTR, possessing the AUG start codon for replicase enzyme **(Table 4.2)**. Interestingly, the loop of all two SL structures is enriched with conserved $(CAAA)_{n=1}$ sequence. It might have any functional significance with virus replication and translation.

Comparably, the 3' terminal region is more complex with a typical tRNA-like structure and some pseudoknot structures. About 55nt long consensus sequence of tRNA-like structure (TLS) was folded into a long stem-loop starting with a non-

canonical G-U base-pairing and consists of two arms arising from an internal bulge, where one arm is the resemblance of the tRNA anticodon branch. Just upstream to this TLS, three consecutive pseudoknots (PK-1, PK-2, and PK-3) are present, showing kissing hairpin interaction. A complex pseudo-knot structure with two double helix stem (S1, S2) with a kissing hairpin loop was also located downstream of the TLS (**Fig.4.2)**. Of them, the sequence in the S2 was not conserved always; sometimes absent in some isolates of CGMMV.

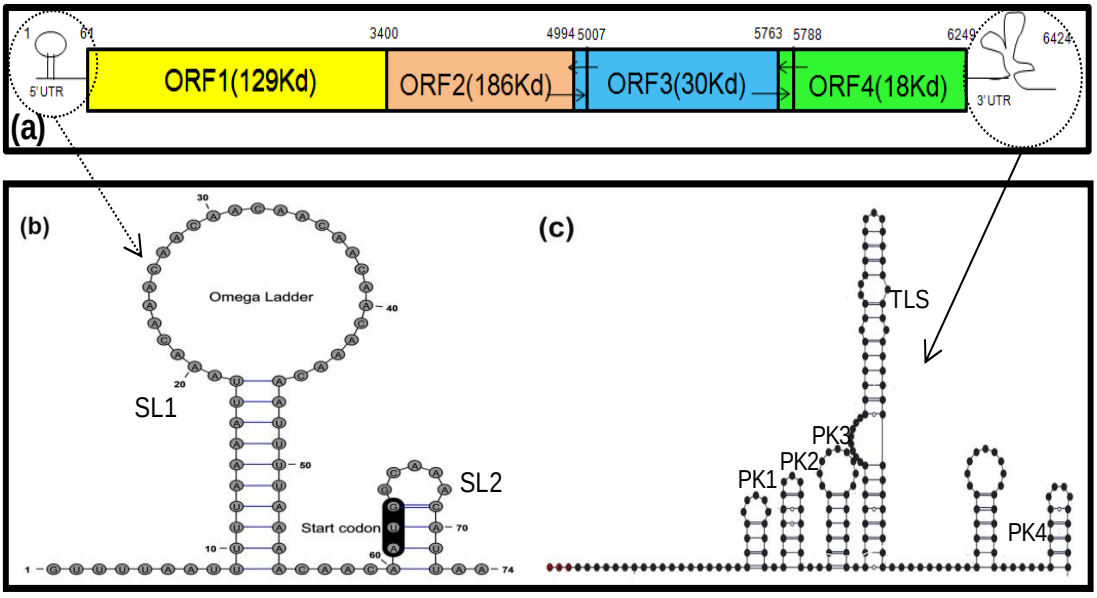


Fig.4.2. The RNA secondary structure of 5' and 3' UTR of CGMMV genome. (a) The complete genome map of CGMMV depicting the presence of the regulatory elements in the 5' and 3' terminal ends. (b) The 5'UTR of CGMMV possesses a large stem-loop (SL) structure which is named as the omega ladder in as of TMV. It is rich in CAA repeats that have a crucial role in the binding of replicase enzyme. The Omega ladder is followed by another small SL structure that carries the translation start site for replicase enzyme. (c) The 176 nt long 3'UTR of CGMMV is highly conserved; it is folded to a complex tRNA-like structure (TLS) at the most 3'terminal end that is the preferred site for the binding of replicase enzyme. In the upstream of TLS, three pseudoknots (PK1, PK2, PK3) were identified. Their structure as well as sequence conservation were also important for replicase enzyme functioning during negative-strand synthesis.

Table 4.2. Predicted secondary structures identified in the 5' and 3' UTRs of CGMMV

No	Structure	Sequence	Secondary Structure	Location in genome (nt)	MFE (kcal/mol)
1	Omega ladder (SL1)	GUUUUAAUUUUUAA AAUUAAACAAACAA CAACAACAACAACA AACAAUUUUAAAA	.....((((((((((..... .....))))))))))	5' UTR (1- 55 nt)	-4.75
2	Sl2	CAACAAUGGCAAAC AUUAAUG	....((((.....))))....	5' UTR (56-76 nt	-1.2
3	PK1	UGGUGGUGCACACC AA	(((((.....))))).	3' UTR (6266-6281 nt	-6.8
4	PK2	AGUGCAUAGUGCUU U	(((((...)))))..	3' UTR (6282-6396 nt)	-1.0
5	PK3	CCGUUCACUUAAAU CGAACGG	((((((.....))))))	3' UTR (6298-6318 nt)	-6.3
6	TLS1	GGUUUGCGGAAACC UCUCACGUGUGA	(((((([[[.)))))..(((]]] .))) (((((....)))))..... .	3' UTR (6329 - 6355 nt)	-13.0
7	PK4	AAGGGGUUCGAAUC CCCCCUUUUCCCUG GGUAAGGG	((((((.....[[[[])))))).. ((((]]]].))))	3' UTR (6385-6420 nt)	-19.9

*These RNA secondary structures were identified based on Mfold, Scan-fold, and the dot-knot web server. The minimum free energy (MFE) of these structures represents their stability. Stem-loop (SL-1, SL-2) are found in the 5'UTR, whereas, pseudoknots (PK1, PK2, PK3, PK4) along with tRNA-like structure (TLS) are the important regulatory component located at 3'UTR.

4.1.1.2.3. Three-Dimensional modeling of the RNA secondary structures

The three-dimensional structures of RNA pseudoknots and stem-loop structures were designed based on a template-dependent approach. Initially, models were generated through the Hdock server based on the highest sequence identity (%) with the crystal structures of the respective template. Although the majority of the templates have appeared as the complex nucleic-protein crystal structures in the PDB database, only RNA components were selected for 3D modelling. Except for the tRNA-like structure, all other RNA structures were showing >50% sequence identity, with a maximum 77% identity by pseudoknot 1 with its template chain A of 5V7Q. PDB. Other than sequence identity, a high TM-score varying from 0.45 to 0.86 was also provided to address the

topological similarity between predicted RNA structures and their template. Furthermore, low RMSD values (below 20˚A) of all the generated RNA 3D models determined based on phosphate as the representative set of atoms centred around the 28˚A sphere were signifying their desirable quality (**Table 4.3**). The only tertiary structure of the tRNA-like structure was curated in the RNAComposer web server (http://rnacomposer.cs.put.poznan.pl/) based on a template-independent manner using their predicted secondary structure (Biesiada *et al.*, 2016). The RMSD of this model was at par with that of the model generated through the template-dependent approach. On retrieval of secondary structures from the RNA PDB model in RNApdbee, the higher resemblance with the previous predicted secondary structure helped us to select for molecular docking.

Table 4.3. 3D-Models of RNA secondary structures of 5' and 3' UTR of CGMMV genome and their quality parameters

Model	Template	Sequence identity (%)	TM-score	RMSD (˚A)	Molecule Types of template	Source
Omega ladder	4V19.pdb (Chain A)	65.9	0.72487	7.38	Mito-ribosomal 16S rRNA	The large subunit of the mammalian mitoribosome of Sus scrofa
Sl2	2N0R.pdb (Chain A)	55	0.75620	2.08	RNA	RNA structure of synthestic construct
PK1	5V7Q.pdb (Chain A)	76.9	0.86364	2.57	23S rRNA	Large ribosomal subunit from Mycobacterium tuberculosis
PK2	6G90.pdb (Chain 1)	66.7	0.65978	19.51	U1 snRNA	Prespliceosome structure of Saccharomyces cerevisiae
PK3	2KUU.pdb (Chain A)	50	0.79321	7.24	K10 TLS RNA-Chain A	Structure of K10 TLS RNA (GC mutant in the upper helix) of Drosophila melanogaster
TLS-1	4AOB.pdb (Chain A)	57.1	0.73831	9.29	SAM-I Riboswitch	SAM-I riboswitch containing the T. solenopsae Kt-23 in complex with S-adenosyl methionine
PK4	5XYM.pdb (Chain A)	71	0.44596	7.81	23S RNA	Large ribosomal subunit of Mycobacterium smegmatis

* Three-dimensional models of RNA secondary structures were generated based on template-dependent approach using H-dock webtool. The model RMSD is calculated based on RNAlyzer tool. The detail about the template was collected from Protein databank.

Table 4.4. 3D-Models of helicase and core RdRp domain of replicase enzyme of CGMMV and their quality parameters

Model	Template	Sequence identity (%)	Residue	Query coverage (%)	Secondary structure			Residues in Ramachandran plot			Z score	QMEAN	GMQE
					Disordered (%)	α-helix (%)	β-strand (%)	Favoured region (%)	Allowed region (%)	Outlier region (%)			
Helicase	3vkw.1.A	57.31	254	100	13	23	26	93.4	5.7	0.40	-6.57	-1.47	0.80
Core RdRp	5ziu.1.A	27.45	109	86	12	37	27	89.8	9.2	1.0	-0.73	-2.89	0.56

*3-D modeling of helicase and RdRp was done by using swiss-model web-tool. Its structural details were derived by using Phyre2. The Ramachandan plot, Z-core, QMEAN score was generated by using RAMPAGE, Molprobity, ProSA, ERRAT plot to define the structural suitability of these models.

4.1.1.2.4. Homology modeling of helicase and core RDRP domain

Based on the template search, a total of 7136 and 283 templates were found from the PDB database for helicase and core RDRP domains, respectively. Of them, helicase showed high sequence identity (57.31%) to the crystal structure of the superfamily 1 helicase from tomato mosaic virus (3vkw.1.A, and 3wrx.2.B) with 100% coverage. The 3vkw.1.A (Nishikiori *et al*., 2012) is used as the template for generating the three-dimensional model without Mg^{+} ligand. The re-building of the model in Phyre2 confirmed that a total of 254 residues have been modeled with 100% coverage and 100.0% confidence based on the single highest scoring template (3vkw.1.A). Further, the secondary structure prediction of the protein revealed that it is composed of 23% α-helix and 26% β-strand **(Fig. 4.3)**. Similarly, core RDRP has an acceptable sequence identity (27.45%) to the crystal structure of RdRp (5zit.1.A/ 5ziu.1.A), and genomic polyprotein (5xe0.1.A) of human Enterovirus D68 (Li *et al.*, 2020), with about 86% coverage. The 3D model of RDRP was generated using the crystal structure of RdRp (5zit.1.A/5ziu.1.A) as a template based on homology modeling**.**

The quality of the generated protein model was then evaluated based on Ramachandran plot, z-score, GMQE score along with Overall Quality score as generated by ERRAT program and energy plot as in ProSA. The Ramachandran plot for the helicase model showed 93.4% residues are in the favoured region, whereas for RdRP, 89.8% residues were in the favoured region with 9.2% are in the allowed region. The distribution of ~90% residue in the favoured region ensured the overall acceptability of both these models. Additionally, the global model quality estimate (GMQE) for both models scored between 0 to 1(0.8 for helicase, and 0.56 for RdRp core) indicating model stability **(Table 4.4**.). Furthermore, the overall quality score generated in ERRAT program for helicase and RdRp was 91.453 and 77.083, and a lower Z-score (less than zero) supporting the reliability of the homology modeling.

The overall homology of the core RdRp model matched with its template (5zit.1.A) with RMSD: 0.834, and Q-score: 0.202. Of 109 aa, 104 aa of core RdRp was fully superimposed on the 224 aa to 334 aa in the final alignment. The modeled structure of helicase (254 aa long) was aligned between 860-1113aa of 129 KDa protein and composed of two sub-domains, *viz*., 1A, and 2A, with multiple catalytic motifs (**Fig. 4.3**.). Whereas, the RdRp was overlapped with the part of palm (195-246, and 286-375) and finger (264-286 aa) subdomains of the template, and composed of 3 conserved structural motifs, *viz.*, motif A (residues 02-14), motif B (residues 70-84), motif C (residues 92-102), out of seven motifs (A-G) in all RdRp (**Fig. 4.4**), that expected to play a pivotal role in catalysis (Li *et al.*, 2020).

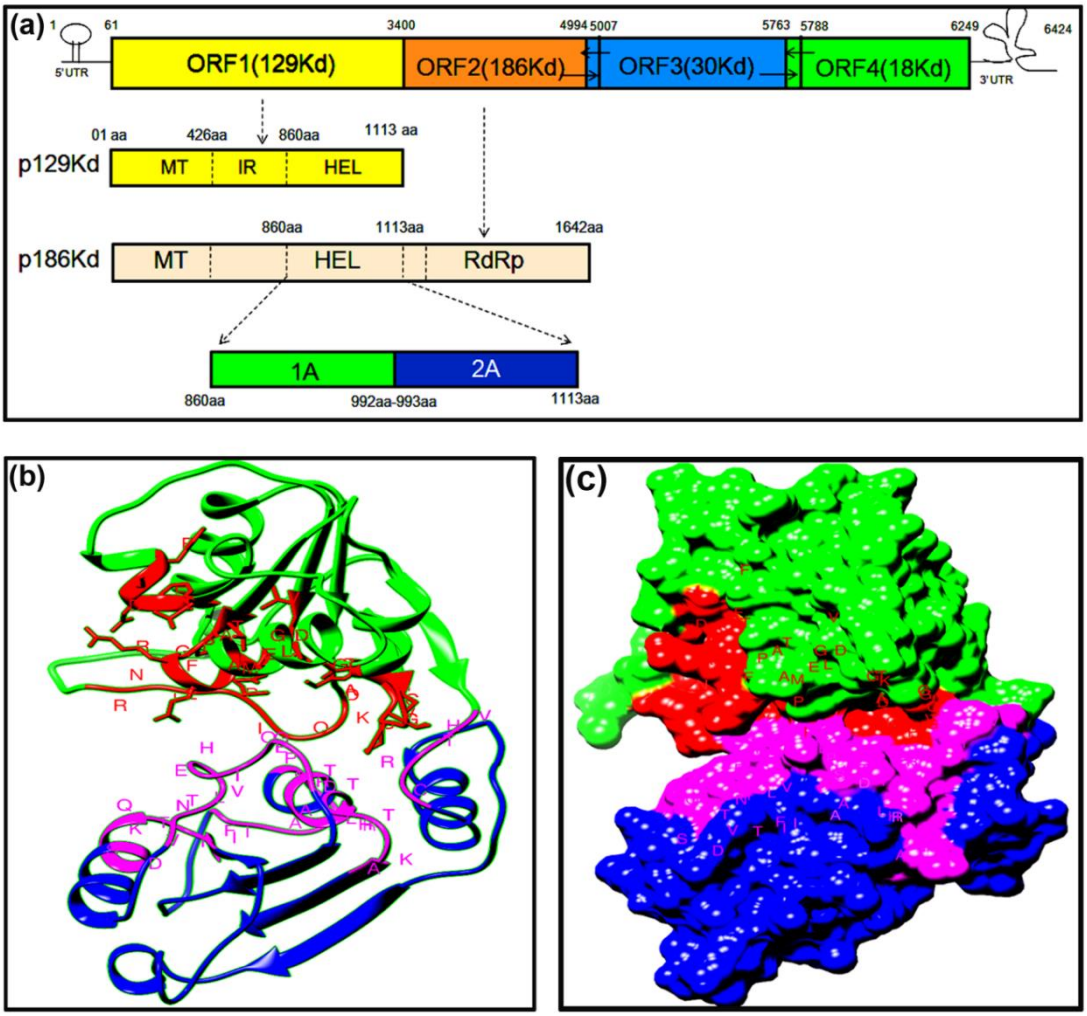


Fig. 4.3. The three-dimensional model of helicase domain (Hel) of CGMMV replicase enzyme complex. (a) The genome map of CGMMV depicts the expression of 129 kDa (from ORF1) and186 kDa (from ORF2) protein that polymerizes to form the replicase complex. The helicase domain is located within 860-1113 aa of both protein with its two sub-domains i.e., 1A (860-992 aa) and 2A (993-1113 aa) (b) The tertiary structure of helicase domain contains multiple catalytic motifs in its two sub-domains that are conserved in Superfamily I helicase enzyme. The 1A (green) has conserved motifs like GVPGCGKT, TPGREAA, TFDSF, VDEGLM, GDAKQIPFINR (coloured in red) whereas the 2A sub-domain (blue) has VTHRC, ITFTQSDK, VNTVHE, IQGET, LIARDSP, VALTRHTK conserved motifs (coloured in magenta). (c) The surface-exposed model of helicase domain with its two subdomains and conserved catalytic motifs that form a central groove for binding with nucleic acids.

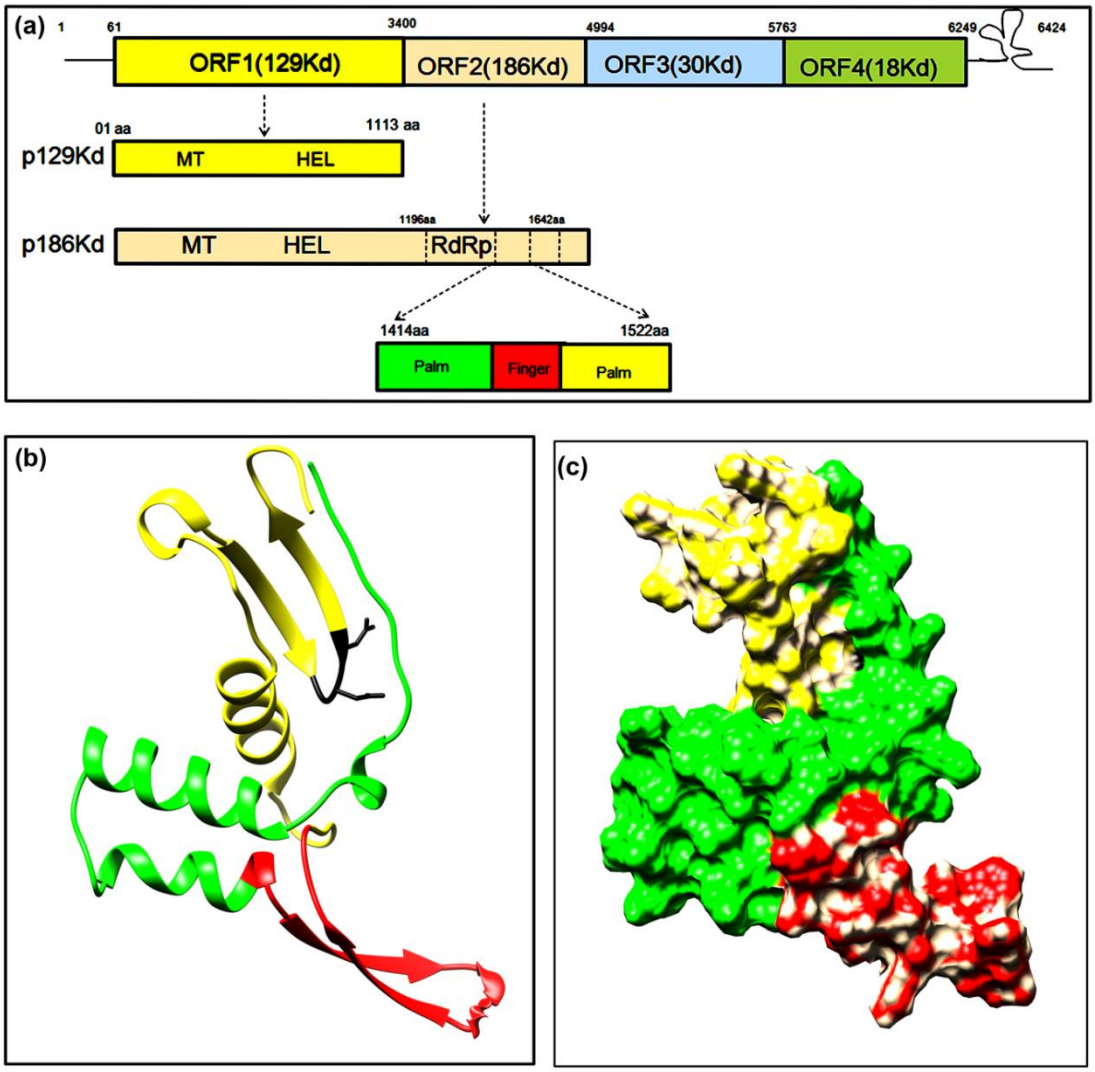


Fig.4.4. Three-dimensional model of core RNA dependent RNA polymerase domain (RdRp) of CGMMV replicase enzyme complex. (a) The genome map of CGMMV depicts the expression of 129 kDa (from ORF1) and186Kda (from ORF2) protein that polymerizes to form the replicase complex. The RdRP is part of the 186 kDa protein with its core catalytic sites are located within 1414aa to 1522aa. (b) The tertiary structure of the core RdRp domain consisting of palm sub-domain (1-41aa) represented in yellow colour, finger sub-domain (42-62aa) represented in red colour, followed by another palm sub-domain (63-109aa) represented in yellow colour. The conserved GDD motif was shown in black colour. (c) The surface-exposed model of core RdRp with different sub-domains (represented with a different colour) to the active protein surface.

4.1.1.2.5. Interaction between helicase and core RDRP domain with the RNA regulatory elements

The *in-silico* interaction of different domains of replicase enzyme with the regulatory elements in the 5'UTR region depicted an identical pattern. In the omega ladder, both the helicase and RdRp bound specifically to the $(AAAC)_3$-$(AAC)_5$ repeat sequence. The formation of strong H-bond at the junction of the compound repeat at both ends of the hairpin loop signified the importance of certain nucleotides (CAAA-CAA & AAC-AAC-AAAC) in the formation of complex interaction with LYS19, LEU36,38,42, MET45, PHE58, ARG70, LYS71, SER72, GLY73, ASP74, VAL75, THR76, 77, PHE83, ASP106 amino acids of RdRp to the omega ladder (**Fig. 4.5**). In contrast, the stem region of the SL1 structure located (between 56-76 nt region) just following the omega ladder was mostly preferred by both the helicase and RdRp domain. Interestingly, helicase was found to be specifically interacting with CAAC nucleotide stretch and RdRP strongly binds to the AAUG stretch covering the start codon in one arm; whereas in the opposite arm, UAAU is highly conserved for binding with them.

The consecutive three pseudoknots followed by a t-RNA like structure in the complex of 3'UTR region have interacted differently with helicase and RdRP domain. In pseudoknot 1, the G-rich stem region containing $(UGG)_2$ repeat sequence was preferred by RdRp for binding, whilst helicase was predominantly interacted with following UGGUGCA nucleotide stretch in the hairpin loop region through hydrogen, electrostatic, and hydrophobic interaction. Similarly, such specification was identified in pseudoknot2 where some nucleotides of both the stem (GUGC) and terminal-loop (AUA) region were interacted with RdRP, though GUGCUU nucleotide stretch of the stem was preferred by helicase. Again, the binding of helicase to specific nucleotide stretch UAAAUCG in the hairpin loop of pseudoknot3 was observed, and that of RdRp to GAACGG nucleotides in the stem/arm region was noticeable (**Fig.4.6)**. It signifies the binding preference of helicase to the unwinded strands of virus RNA. On the pseudoknot 4 region, entrancingly, the strong interaction of RdRP and helicase was visualized with the hexa-repeats of C_6 located in the S1 double helix stem and $U_3C_3G_3$ repeat sequence of the S2 double helix stem region, indicating the potential role of these nucleotides in the initiation of replication.

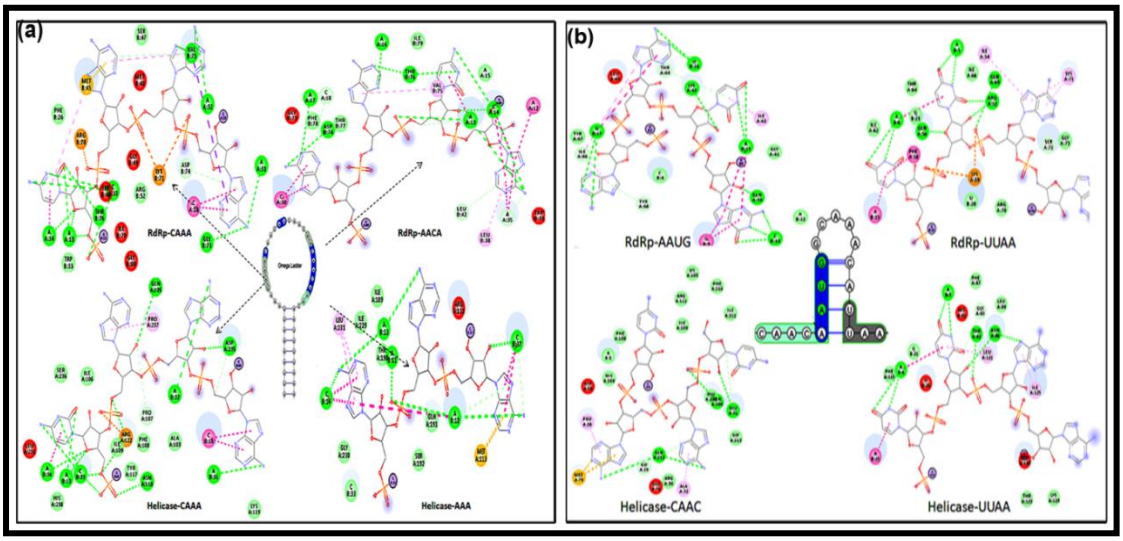


Fig.4.5. Binding of helicase and RdRP domain of replicase enzyme with the regulatory elements of 5' UTR. The interaction between different amino acids of helicase and RdRp with the definite nucleotides in the genome specifies their functional importance in replication. (a) The CAA repeats of the omega ladder are mostly preferred nucleotides for binding by both the domains. (b) The stem-loop (SL)-2 is also the important target for binding of replicase as the translation start site for its synthesis is located within it.

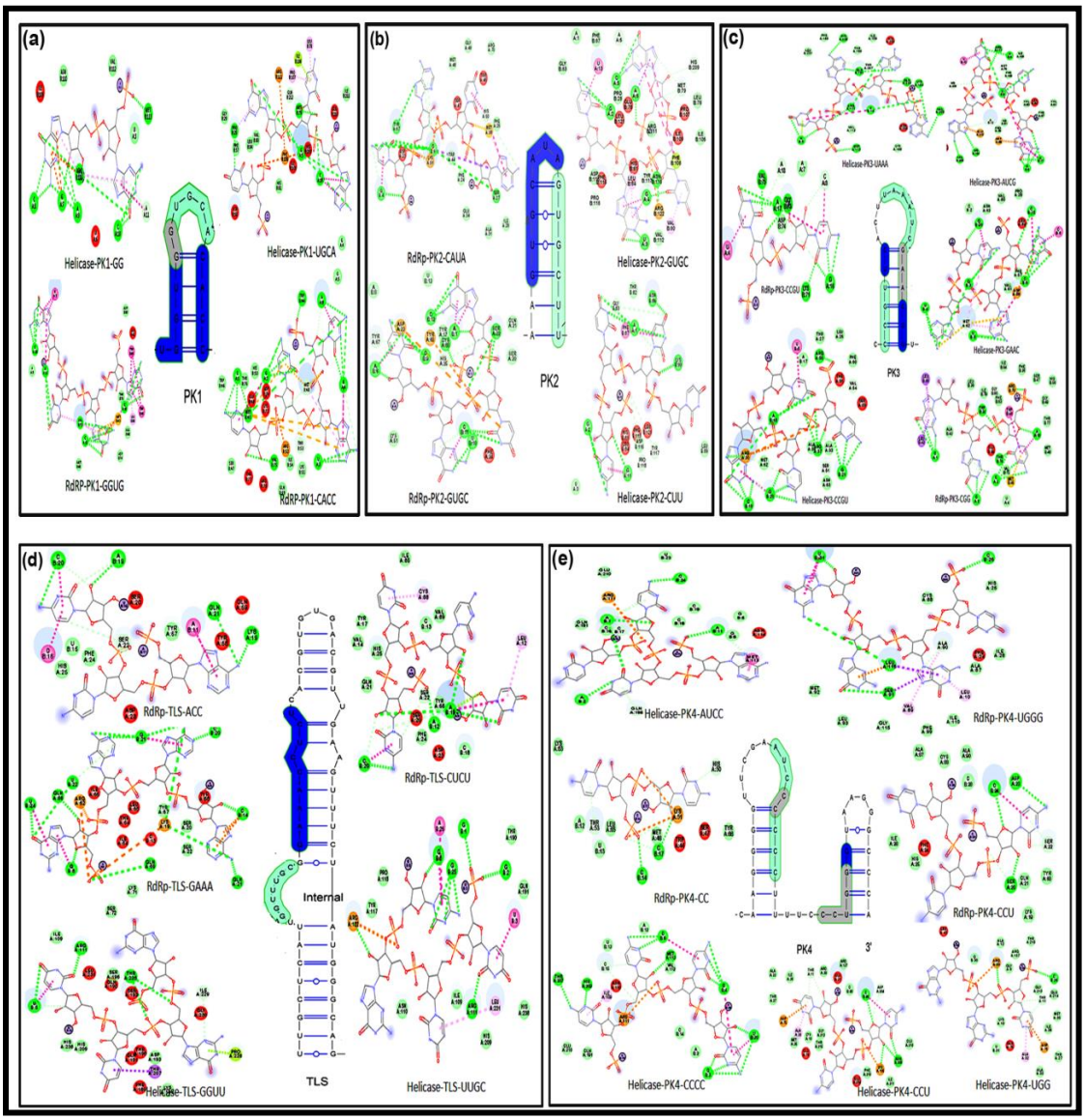


Fig.4.6. Binding of helicase and RdRp domain of CGMMV with the regulatory elements of 3' UTRs. The interaction between different amino acids of helicase and RdRp with the definite nucleotides of various regulatory elements is specified here. (a) The di-repeats of UGG repeats in the PK1 is mostly preferred for binding by both the domains of replicase. (b) The adenine clump (AUA) in the terminal end of PK-2 is also an important target for the binding of replicase. (c) The structural and sequence conservation of PK-3 is important for the synthesis of a negative stand. (d) The long complex tRNA-like structure is the initial site where RdRp binds first to start the synthesis of the negative-strand of the genome.

4.1.1.3. *In silico* prediction of subgenomic promoters for movement protein and coat protein

4.1.1.3. 1. Virus isolates and their sequences

A total of 8 different tobamoviruses viz., CGMMV (DQ767631), CFMMV (AF321057), CMoV (NC_008614), KGMMV (NC_003610), TMV (FR878069), ToMV (AJ243571), WGMMV (MH837097), ZGMMV (AJ252189) (Supplementary file 1) were selected in this study. Of them, six are pathogenic to cucurbits (**Table 4.5.)**, whereas the other two (TMV, & ToMV) are pathogenic to the solanaceous hosts. Interestingly, CGMMV, CMoV, and WGMMV possess overlapping ORFs, i.e, replicase overlaps with MP and MP overlaps with CP. In contrast, these ORFs, *viz*., replicase, and MP, as well as MP and CP are separated by a small spacer sequence in CFMMV, KGMMV, and ZGMMV. This variation in genome organization may have some role in the sgRNA production.

Based on previous studies of the MP-SGP and CP-SGP in other viruses, a stretch of 400 nts of CP-SGP that included 200 nt of 3' terminal end sequences of MP and 200 nt of 5'terminal of CP ORF were sliced out with respect to the translation start site (TLSS). Similarly, another stretch of 460 nts in the MP-SGP that included 230 nts of 3' terminal end of replicase (ORF2) and 230nt of 5'terminal end sequence of MP ORF with respect to MP-TLSS (+1) was selected for the analysis. In the context of poorly characterized RNA virus promoters, the translation start site was considered as the basis of marking the location of promoters. Although, conventionally, in the case of the DNA genome, the transcription start site (TSS) was used as the base to represent promoters as TSS was mostly well characterized. But for the plant RNA viruses, especially for CGMMV, the transcript mapping of the sgRNAs was mostly remain undeciphered; thus, the transcription start site (TSS) remained also undefined. Now, to avoid the ambiguity for the characterization and identification of subgenomic promoters, the translation start site (+1) was used as a baseline.

Table 4.5. The genome organization of different cucurbit infecting tobamoviruses

Sl NO	Virus species	Acronyms (Accession)	Genome size (nt)	Translation start site (nt)		Promoter type	
				MP	CP	MP	CP
1.	*Cucumber fruit mottle mosaic virus*	CFMMV (AF321057)	6562	5064	5846	-	7nt SP
2.	*Cucumber green mottle mosaic virus*	CGMMV (DQ767631)	6424	4994	5763	13nt OL	26nt OL
3.	*Cucumber mottle virus*	CMoV (AB261167)	6485	5066	5826	14nt OL	29nt OL
4.	*Kyuri green mottle mosaic virus*	KGMMV (AJ295948)	6514	5073	5866	3nt SP	5nt SP
5.	*Zucchini green mottle mosaic virus*	ZGMMV (AJ252189)	6513	5075	5865	3nt SP	5nt SP
6.	Watermelon green mottle mosaic virus*	WGMMV (MH837097)	6,482	5064	5824	14nt OL	29nt OL
7	*Tobacco mosaic virus*	TMV (FR878069)	6392	4903	5709	17nt OL	4nt SP
8	*Tomato mosaic virus*	ToMV (AJ243571)	6383	4906	5703	17nt OL	4nt SP

*unclassified tobamoviruses; the information collected from the NCBI database. Acronyms: Overlapping (OL), spacer (SP), nucleotides (nt)

4.1.1.3. 2. Relationships of CGMMV MP-SGP and CP-SGP regions with the other cucurbit infecting tobamoviruses

The nucleotide sequences analysis of the MP-SGP and CP-SGP of tobamoviruses showed a significant difference among them. The MP-SGP of CGGMV shared the maximum identity of 66.2% with that of CFFMV and 66.0% with CMoV, followed by 65% with WGMMV; and only 45-46% with TMV and ToMV. Thus, in the phylogenetic tree, CGMMV MP-SGP formed a distinct cluster from clade 1 containing CFMMV, KGMMV, and ZGMMV, and clade 2 composing CMoV and WGMMV (**Fig. 4.7a**). Similarly, the CGMMV CP-SGP had a maximum 49.7% sequence identity with that of CMoV, and 49% with WGMMV, but shared only 32.8 and 35% identity with that of TMV and ToMV, respectively ((**Fig. 4.7b**). The phylogenetic tree based on CP-SGP also depicted three distinct clades similar to MP-SGP, where CGMMV was positioned separately from clade 1 and clade 2. Notably, both phylogenetic trees represented a similar evolutionary pattern of CGMMV. The sequence comparison and subsequent phylogenetic analysis of the MP-SGP and CP-SGP region indicated the possible structural difference among the members of different clades.

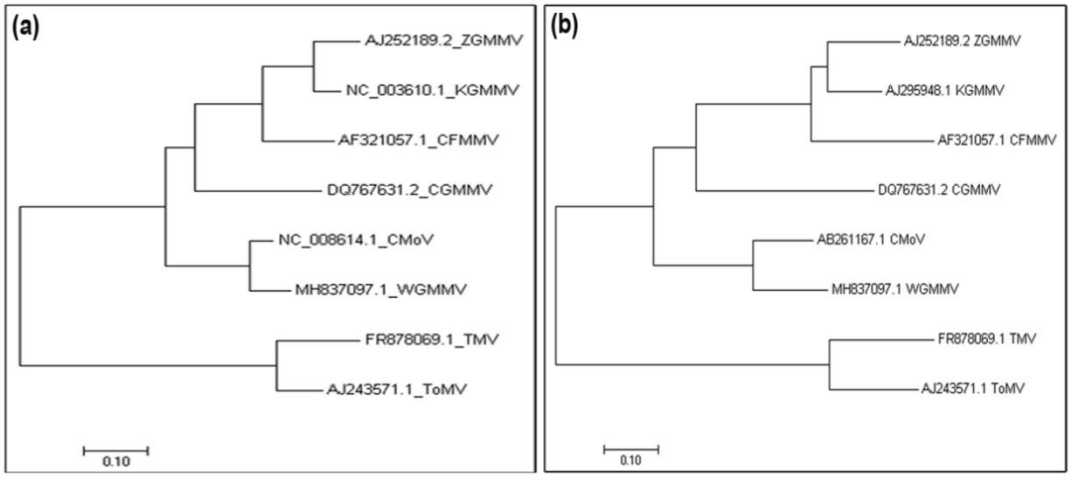


Fig.4.7. Phylogenetic relationship among the six different cucurbit infecting tobamoviruses based on the subgenomic promoter region (SGP) of movement protein (MP) and coat protein (CP) represents a similar evolutionary pattern of CGMMV as it is placed in a distinct clade in both cases. (a) The evolutionary history represents the relationship between the MP-SGP of different cucurbit infecting tobamoviruses (b) The evolutionary relationship between the CP-SGP of different cucurbit infecting tobamoviruses. The trees are inferred by busing the Neighbor-Joining method.

4.1.1.3.3. Distribution of cis-elements within MP and CP SGP regions

The mining of cis-elements in the SGP sequence had shown their presence in varying numbers among the cucurbit infecting tobamoviruses. A total of 15 cis-elements were identified in the MP-SGP region (**Fig.4.8a**), and that of 20 elements were distributed in the CP-SGP region (**Fig.4.8**b). Of them, the CAAT box and TATA box elements were the most common and present in the highest copy number. Additionally, the jasmonic acid (MeJA)-responsive CGTCA-motifs/ TGACG-motifs, salicylic acid-responsive TCA-motifs (CCATCTTCTT), abscisic acid-responsive ABRE motifs (ACGTG) was also observed in these regions of some viruses. Further, auxin-responsive TGA-elements (AACGAC), Gibberellin-responsive GARE-motifs (TCTGTTG) were also evident in a few cases. The existence of such regulatory elements in the promoter region pointed out the plausible interactions with various host trans-acting factors during sgRNA production.

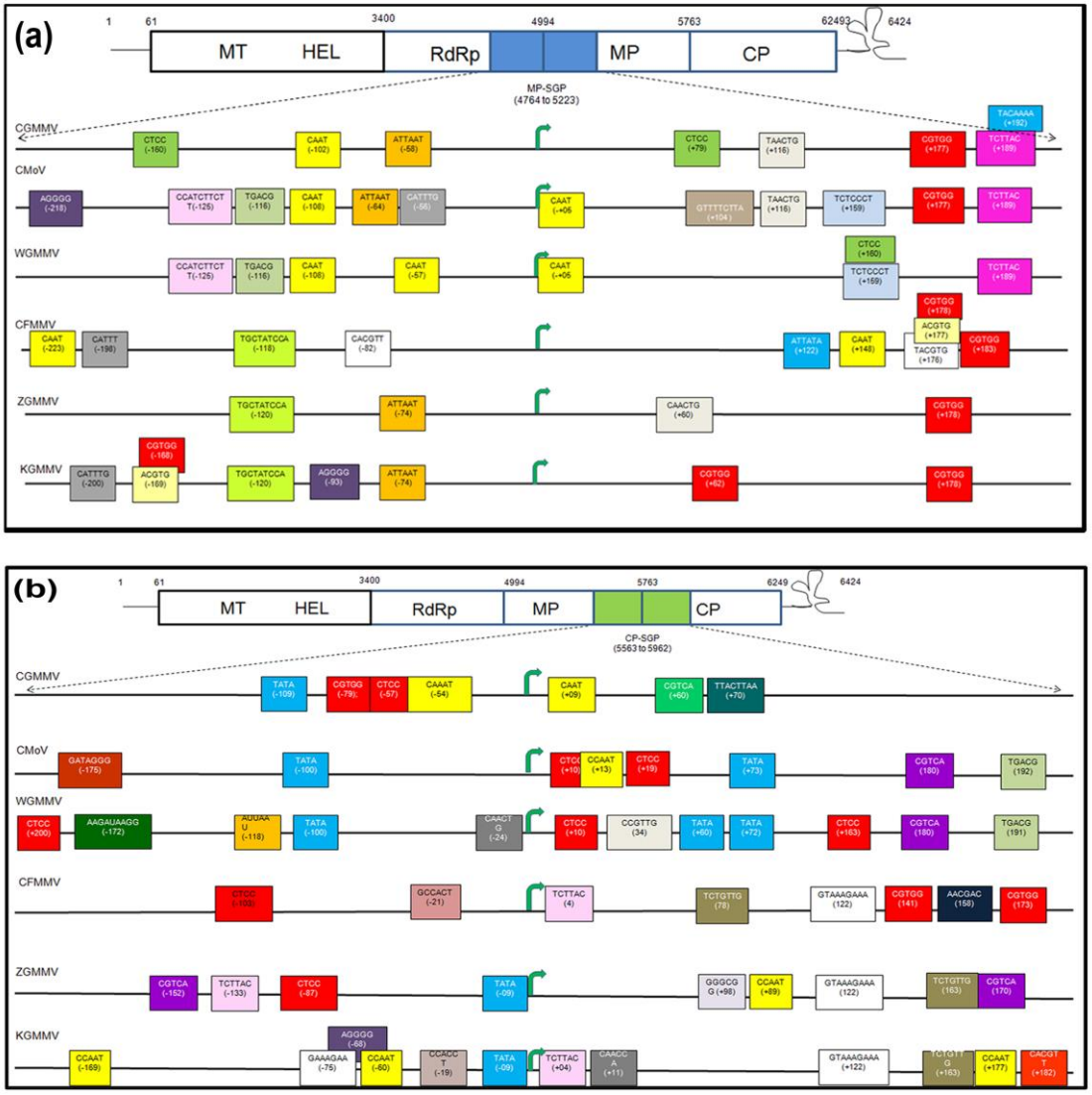


Fig.4.8. Distribution of putative cis-elements within the subgenomic promoter region (SGP) of movement protein (MP) and coat protein (CP) of different cucurbit infecting tobamoviruses identified using plant care database web-server. The colour indication is used to represent different elements. The coordinates of the elements are indicated with respect to the translation start site (green vertical arrow) in parenthesis. (a) Showing the location of MP-SGP in the CGMMV genome and the distribution of cis-elements in MP-SGP. A total of 15 different cis-elements were identified within MP-SGP regions of different tobamoviruses. The occurrence and distribution of similar regulatory elements within MP-SGP of CGMMV, CMoV, and WGMMV indicate the structural similarity among them, whereas CFMMV, ZGMMV, and KGMMV have a distinct pattern. (b) Showing the location of CP-SGP in the CGMMV genome and the distribution of cis-elements in CP-SGP. A total of 20 different cis-elements were identified in the CP-SGP region of six cucurbit-infecting tobamoviruses. CGMMV CP-SGP possesses only 7 cis-elements; although they are common to others, their distribution pattern depicts significant variation.

4.1.1.3.4. RNA secondary structures of MP and CP SGP among cucrbit infecting tobamoviruses

The predicted RNA secondary structures of MP-SGP and CP-SGP of various tobamoviruses were drawn to confer their structural similarity. The MP-SGP region of CGMMV spanning from 4764 to 5223nt possesses many long SL structures, and showed extensive structural similarity with CMoV, and WGMMV, but differs from other cucurbit infecting tobamoviruses. Usually, TLSS of MP in CGMMV along with CMoV and WGMMV was pitched in a terminal loop of an SL structure (SL-2) that is surrounded by many SL structures. Of them, SL-1 in the upstream region held small but structurally conserved/similar pseudoknots, observed within -110 to -60 nt with respect to the TLSS . Other than their structural similarity, they may possess some sequence similarity also. They might have a crucial role in enhancing the sgRNA transcription or may support the binding of RdRp enzyme. Likewise, in the downstream part, consecutive four large stem-loops (SL-3, SL-4, SL-5, SL-6) were identified in the case of CGMMV, which were identical to the counterpart of CMoV (**Fig.4.9a**), but their functional significance is yet to established in regulating the MP sgRNA production. Adversely, TLSS of MP in CFFMV, ZGMMV, and KGMMV is situated in the internal bulge or central loop of a long SL structure, and they are also structurally diverged from the group of cucurbit-infecting tobamovirus having overlapping ORFs (replicase and MP) in this SGP region.

Similarly, the CP-SGP region of CGMMV (5563-5962nt) possessed SL structures with many internal bulges and loops. The TLSS (+1) was situated in one such internal loop/bulge, and it was also evident in the case of other tobamoviruses, *viz*., CMoV, CFFMV, WGMMV, KGMMV. Further, two short hair-pins might be the putative enhancers located just within -60 nt upstream of TLSS, showed structural conservation among the tobamoviruses (**Fig.4.9b**). Although the length distribution of the large SL structure possessing the TLSS (+1) along with putative enhance elements is varying, still its emergence from a large central loop that indicated the putative sites of RdRp binding for the sgRNA production. This length variation corresponds with the difference in the actual promoter length of these tobamoviruses.

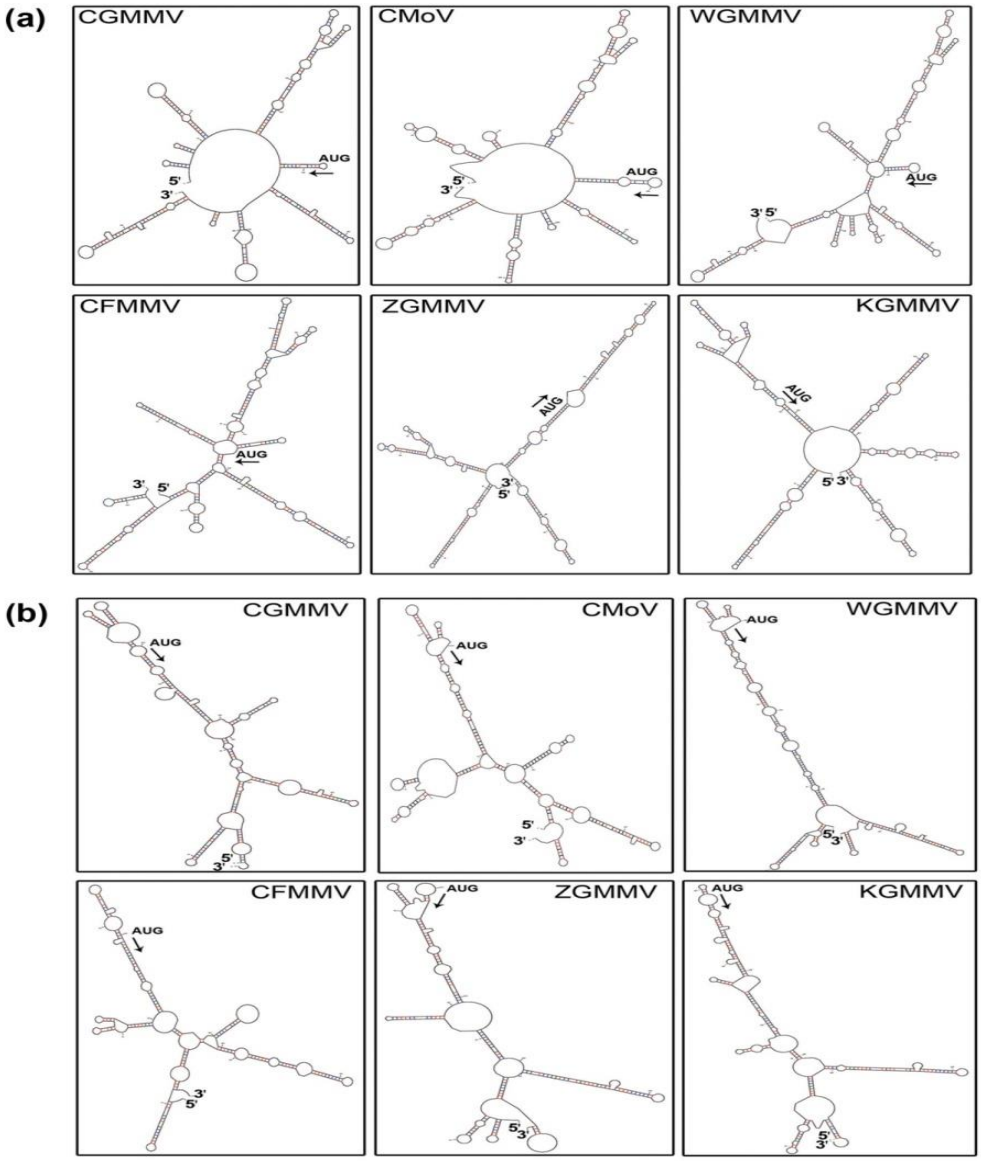


Fig.4.9. RNA secondary structures of the subgenomic promoter region (SGP) of movement protein (MP) and coat protein (CP) of different cucurbit-infecting tobamoviruses generated from M-fold web server using template RNA sequence. The arrow indicates the location of the translation start site (TLSS) (+1) with AUG start codon (indicted with an arrow). (a) In MP-SGP of CGMMV, TLSS (+1) is located within a terminal loop of SL structure (SL-5) which is surrounded by other SL elements both in the upstream (SL-1, SL-2, SL-3, SL-4) and downstream (SL-6, SL-7, SL-8, SL-9) regions that are indistinguishable from the secondary structures of CMoV, and WGMMV. The CFFMV, ZGMMV, and KGMMV have a different pattern of structural elements with localization of TLSS (+1) within the internal bulge. (b) The secondary structure of CP-SGP of CGMMV has a certain degree of similarity with that of CMoV and WGMMV. It contains two small hairpins that emerged from a central loop located within 60 nt upstream with respect to TLSS (+1). These elements might have a key regulatory function in sgRNA synthesis. ZGMMV also has a similar pattern, but CFMMV and KGMMV have distinct features in this region. The start and end of the single-stranded RNA are indicated with 5´ and 3´.

4.1.1.3. 5. Three-dimensional modeling of RNA structures

The three-dimensional structures of RNA segments, representing various elements in the upstream as well as down-stream part of MP-SGP and CP-SGP regions were curated using a complementary strand of the viral SGP sequence in the RNA Composer web server based on a template-independent approach. This web-server facilitated the automated 3D modeling of long RNA sequences based on their secondary structures (Popenda *et al.*, 2012). The structural conservation of the 3D models was confirmed based on the retrieved secondary structures from the respective RNA PDB model using RNApdbee. The structural resemblance of RNA 3D models in the 2D form with the previously predicted secondary structure of RNA ligands helped us to select for molecular docking study.

4.1.1.3.6. Interactions of RdRp with MP SGP elements in the CGMMV genome

To identify key regulatory motifs in the MP-SGP region, the binding ability of RdRp to the counterpart in the negative strand was explored *via* molecular docking. In the upstream region, RdRp interacted with various sites spanning from -1 to and -136 nt (**Table 4.6.**). Most of these sites were scatted and possess small SL like structures and internal bulges. RdRp strongly preferred the double helix stem region of SL structures. Based on interaction analysis, four long nucleotide stretch (I, II, III, and IV) were identified within -136 to -114 nt (AAUUACCACGAAUAAAGGUAAA), -64 nt to -56 nt (AAAACAUUAA), -53 nt to -32 nt (AACAAACACAUUCAUAAACUCA), and -26 nt to -17 nt (CGCGGAAAAG) position. During scanning, RdRp showed strong binding affinity (with lowest binding energy) for -26 nt to -17 nt site through the formation of strong hydrogen bonds with LYS16, TYR17, LYS19, SER20 aa of the motif A in the palm subdomain of RdRp; thus could be predicted as the putative key elements in the core promoter. Similar strong interactions were noticed at the -32 nt to -53 nt and -56 nt to -65 nt position (**Table 4.6.**).

Similarly, in the downstream part (+1 to +200 nt), very strong binding of SER20, SER22, TYR68, ALA90, LEU116 aa of the motif A and C of RdRp was visualized in the internal bulge (CAACUUU) located at +33 to +39 nt near to the TLSS (+1), This could be prognosticated as the putative core promoter. Besides this, different binding interfaces were scattered in different locations (**Fig. 4.10a**). Of them, a long stretch (AGAAUUCAUAAA) was identified at 99 nt to 110 nt that folded into the complex helix with SL structure, where strong interaction with the specific amino acids (CYS88, ALA90, MET92, ASP96, LYS100, LYS114, LEU116) of RdRp was evident (**Table 4.6.**). Overall, our *in silico* analysis of the MP-SGP indicated that the SL structures distributed within -136 nt of the upstream and +110 nt of

downstream are the important elements for the binding with the RdRp domains; thus proposed to be the putative full promoter for MP protein expression, whereas -64 nt to + 39 nt regions relative to TLSS (+1) could be the core promoter, that is sufficient enough for obtaining a basic level of expression.

4.1.1.3. 7. Interaction of RdRp with CP-SGP elements in the CGMMV genome

The molecular docking of RdRp with the upstream of CP-SGP revealed five potential sites for RdRp binding. Interestingly, RdRp was found to interact strongly in a nearby region (within ±20 nt) of the CP-TLSS (+1), mostly around the conserved pseudoknot where TLSS was located. Thereafter, some motif sequences were identified to be distributed in the upstream and downstream region of CP-SGP, where a strong affinity of RdRp binding was visualized. Of them, the nucleotide stretch (UUGGAGGUUUAGCCUCCA) was located at -60 to -43 nt zone, and another stretch (GGUCAA) was located at -114 nt to -109 nt region, was pitched at the internal SL elements (**Fig. 4.10b**). The strong binding of RdRp with its palm and finger subdomains signified its functional importance, thus could be proposed as the boundary for the active promoter elements in the upstream region. Likewise, in the downstream region, binding of RdRp at three different sites (9-21 nt; 97-114 nt, and 136-144 nt) in CP-SGP was observed. The strongest binding of RdRp was evident at +9 to +21 nt, which was just adjacent to CP-TLSS. Further, strong binding was visualized with the double helix of SL structure, situated within +110 to +140nt zone. The amino acids like LYS16, LYS19, GLN21, SER22, LYS51, PHE58, THR64, LEU65, CYS88, ALA90, ASP57, ALA60, GLY61, LYS63, and LEU65 of RdRp showed a potential strong affinity for this binding (**Table 4.7.**). This analysis helped us to predict some of the RNA regulatory elements that were identified to be the potentiality site for binding with palm and finger sub-domain of RdRp. These could be proposed as the essential regulatory elements that were necessary for CP-sgRNA synthesis.

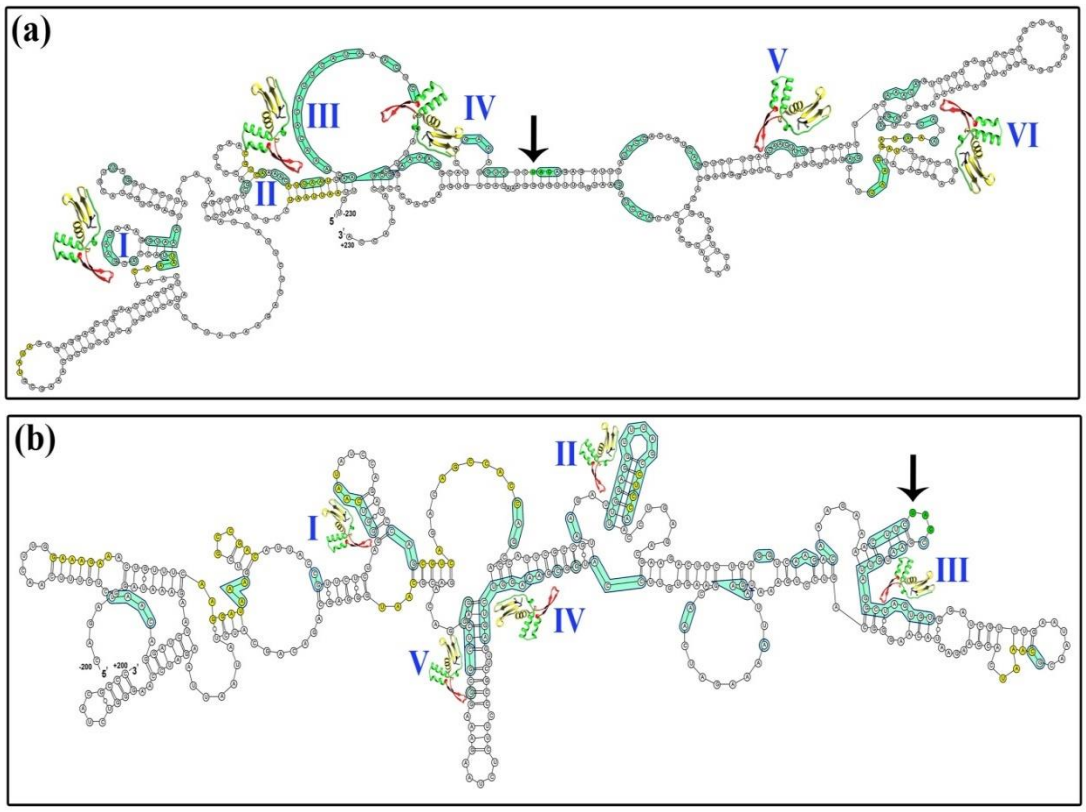


Fig. 4.10. Interaction of RdRp with the different structural elements of the subgenomic promoter region (SGP) of movement protein (MP) and coat protein (CP) of CGMMV. The RNA secondary structures of the SGPs in negative-strand were generated from Mfold web server and used to present the interaction between 3D model of RNA and core RdRp. The arrow mark indicates the location of the translation start site (TLSS) (+1) with UAC as the start codon (green colour, indicated with an arrow), and the number in roman designate the putative interacting sites. (a) A total of six RdRp binding sites were predicted within MP-SGP of CGMMV; four (I to IV) are located in the upstream, and two (V and VI) in the downstream region of SGP. (b) A total of five locations were predicted for the RdRp binding within CP-SGP of CGMMV, of them, two putative sites (I, II) are in the upstream region and three (III, IV, V) are in the downstream region. The detail of the interacting nucleotides of RNA and amino acids of RdRp were presented in 4.6 and 4.7, respectively. The nucleotides in yellow represent the different *cis*-elements (identified based on plant care database) distributed within SGP regions.

Table 4.6. Interactions of RNA dependent RNA polymerase (RdRp) with the regulatory elements of movement protein (MP) subgenomic promoter (SGP) region in the CGMMV genome

Regulatory elements (Co-ordinates)	Interacting nucleotides: amino acids (co-ordinates) of different interacting sites*
I (-136 to -144)	A(-136):GLU11; A(-135):ASP107, ASP13; A(-134):ASP107; U(-133):SER15; C(-129): TYR67; A(-128):LYS51, ALA90, MET92; G(-126):HIS25, SER22; A(-125):SER22, PHE24; A(-124):SER22; U(-123):TYR67; G(-118):LYS63; U(-117):GLN69; A(-116):GLN56, ARG52, LYS19; A(-115):LYS19; A(-114):SER15, SER15, LYS16
II (-64 to -56)	A(-64):GLU11, ILE9, LEU10; A(-62):PHE83; A(-62):ASP96; C(-61):SER91, ALA90; U(-58):ALA31, LYS35; A(-57):GLU34; A(-56):ASP40
III (-53 to -32)	A(-53):TRP44; A(-52):SER20, ILE41; A(-50):MET45; A(-49):CYS104; A(-48):CYS104, ASP107, VAL75; C(-47):LEU109, ASP107, ASP13; A(-46):ASP13, LYS100 ; C(-45):GLU11, ASP96; A(44):LEU116,ILE32; U(-43):GLN56; U(-42):ALA31; C(-41):SER27; A(-40):MET92, SER27; U(-39):MET92, ASP23; A(-38):LYS51; A(-36):THR64, LEU65; C(-35):THR64; C(-33):LYS51; A(-32):LE66,TYR67
IV (-26 to -17)	C(-26):TYR17, LYS16; G(-25):SER20; C(-24):ALA90; G(-23):LEU116; A(-21):LYS19, LYS16; A(-20):LYS19 ; A(-19):LYS19; A(-18):SER72; G(-17):LYS71
V (+33 to +39)	C(33):LEU116; A(34):LEU116; A(35):LEU12, ALA90; C(36):SER20; U(37):SER20; U(38):TYR68; U(39):SER22
VI (+99 to +110)	A(99): LEU116; G(100): LEU116; A(102): CYS88, ALA90, MET92; C(105): LYS114; A(106): LYS35; A(108): ASP96; A(110): LYS100

*Value in parenthesis indicating the coordinates of nucleotides in the RNA with respect to the translation start site (TLSS). The nucleotides located upstream of TLSS (+1) have (-) markings; nucleotides located downstream of TLSS have (+) markings. The interactions were obtained from the interaction analysis between core RdRp and RNA ligands derived from complementary sequences of MP-SGP region using the Patch-dock web server.

Table 4.7. Interactions of RNA dependent RNA polymerase with the regulatory elements of coat protein (CP) subgenomic promoter (SGP) region in the CGMMV genome

Regulatory elements (Co-ordinates)	Interacting nucleotides: amino acids (co-ordinates) of different interacting sites*
I (-114 to -109)	G(-114):ASP57; G(-113):GLN56; U(-112):LYS19, GLN69; C(-111):LYS19; A(-110):VAL14; A(-109):SER20, LYS16
II (-60 to -43)	U(-60): ILE41; U(-59): PHE24; G(-58):CYS88; G(-57):SER22; A(-56): TYR68; G(-55):VAL75; G(-54):VAL75, ASP7, TYR67; U(-53):LYS51; U(-52):LYS51; U(-51):ARG52; A(-50):LYS51 G(-49):MET30,SER27; C(-48):ASP23, PHE24; C(-47):VAL75, SER22, SER22; U(-46):MET45, VAL75, ARG70, CYS88; C(-45):TRP33,TRP46; C(-44):TYR111; A(-43):LEU116
III (+09 to +21)	G(9):LYS35; U(10):LYS35; U(11):ILE54; A(12):LYS35; G(13):SER72; G(14):ASP106; C(15):VAL14; U(16):ASP96, LYS100, SER15; A(17):ASP96, LYS97, LYS16; G(18):LYS16; U(19):GLN56, LYS16; G(20):THR64, GLN56;U(21):THR64
IV (+97 to 114)	C(97):GLU11; C(98):GLU8, ILE9, GLU11, LEU10; A(99):GLU8, LEU116; G(101):GLY115; C(103):CYS104; A(105):TRP33; A(106):VAL75, THR76; A(107):MET45, VAL75, TRP46; G(108):MET45; G(109):ASP74; U(110):ASP74; C(111):PHE58, LYS71; U(112):GLN56; G(113):PHE58; A(114):PHE58
V (+136 to +144)	G(136):GLN59; G(138):LYS63; C(139):LEU65; U(140):TYR67,TYR68; C(141):GLN69, LYS19,GLN69,ARG52; G(143):SER72; G(144):SER72

*Value in parenthesis indicating the coordinates of nucleotides in the RNA with respect to the translation start site (TLSS). The nucleotides located upstream of TLSS (+1) have (-) markings; nucleotides located downstream of TLSS have (+) markings. The interactions were obtained from the interaction analysis between core RdRp and RNA ligands derived from complementary sequences of CP-SGP using the Patch-dock web server.

4.1.2. Development of deconstructed genome of cucumber green mottle mosaic virus

4.1.2.1. Development of deconstructed genome (DG)-1 based on CGMMV genome

4.1.2.1. 1.Designing of deconstructed genome-1 and its RNA secondary structure analysis

A deconstructed genome-1 (DG-1) of CGMMV was firstly designed based on the previously reported dRNA of TMV as the reference. The deconstructed RNA of TMV, named ΔHINC151 (Knapp *et al*., 2005) retained 1-841nt (i.e., 258 aa) of the N- terminal and 5182-6395 nt of the c-terminal of its genome. Based on the sequence alignment, the dRNA of CGMMV was designed to have only the 1- 834nt of the N- terminal part and 5271 to 6424 nt of the c-terminal part (**Fig.4.11a**). The rest of the internal sequence presenting the helicase, RdRP, and movement protein was deleted to abolish the significant functional property like self-replication and cell to cell movement. But, to ensure its trans-replication with the help of wild type CGMMV, its RNA secondary structure was analyzed in the M-fold web server and a total of 35 RNA folded structure of DG-1 was at 37°C. Of them, three structures presenting the lowest free energy value (ΔG = -562.10) were chosen, and all of them were showing conserved hairpin and pseudoknot like structures at the 5' and 3' UTRs of the genome (**Fig. 4.12b**) in comparison to the wild type CGMMV genome (**Fig. 4.12a)**. The deletion of the internal coding region did not affect the structural integrity of the DG-1 that is necessary for the replication. The dot plot map of the structure also signified the minimum free energy-based stability of the genome.

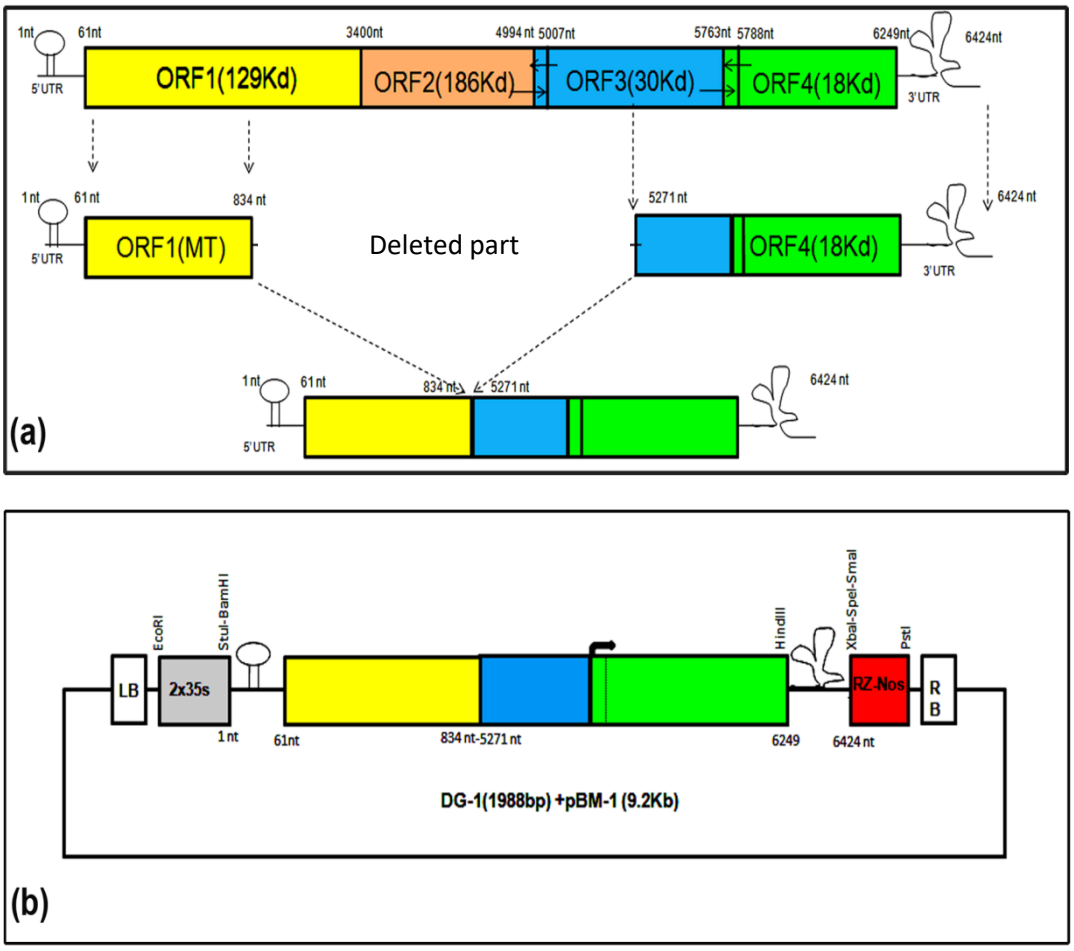


Fig.4.11. Designing of deconstructed genome-1 of CGMMV. It was designed based on the genome organization dRT-V of TMV (Roy *et al*., 2010). It contains 1-834 nt of 5' terminal and 5271-6424 nt of 3' terminal of the genome. (a) The DG-1 (1988 nt) was created through PCR-based deletion of BP4 infectious clone of CGMMV. A significant part of replicase and movement protein was removed. (b) The DG-1 was cloned into a modified binary vector, pBM1 containing the 2X35S promoter after the left border (LB), and ribozyme sites (Rz) followed by a Nos terminator just before the right border (RB). Two-step PCR with high fidelity Taq polymerase is used for creating DG-1.

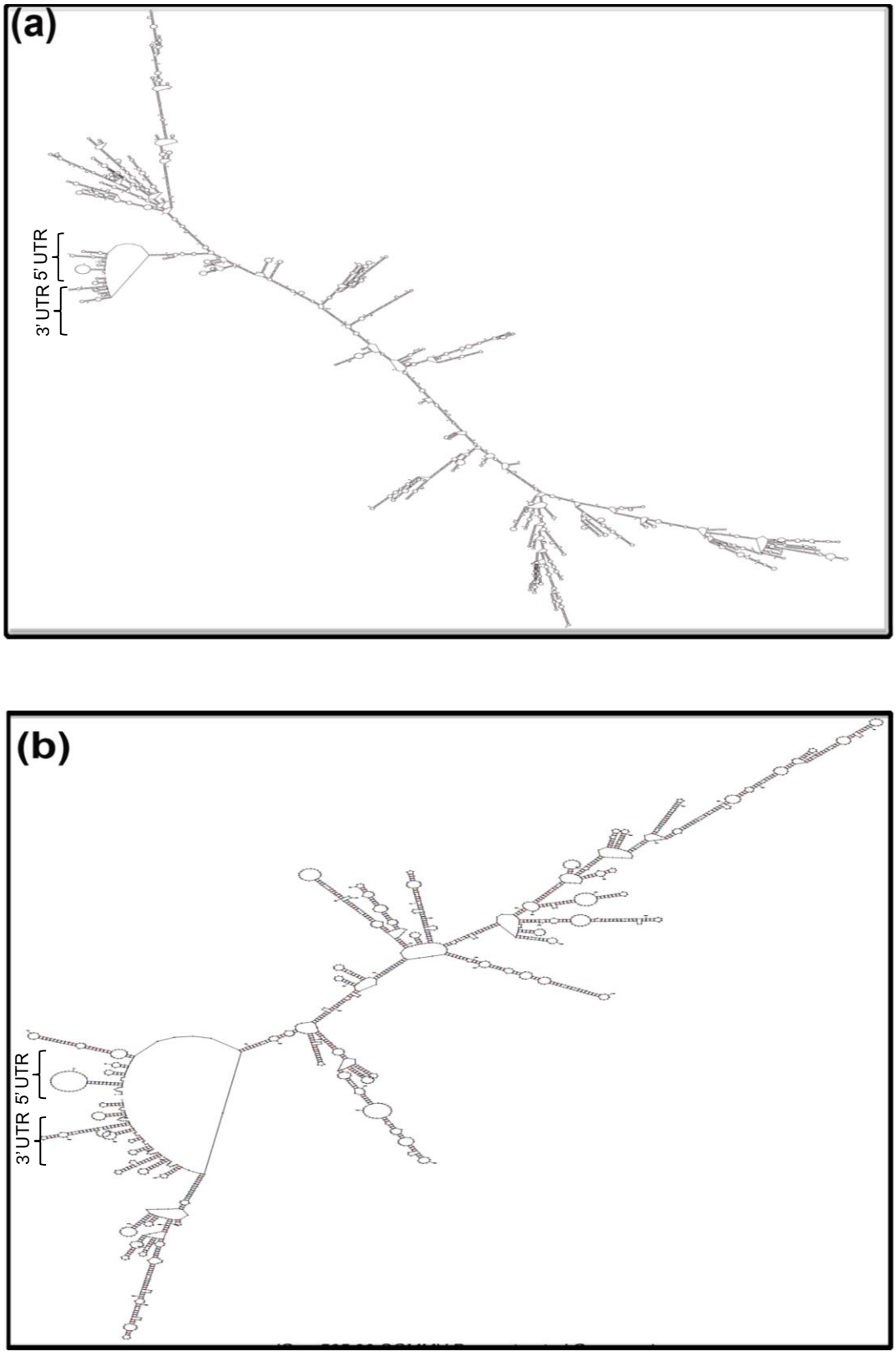


Fig. 4.12. The RNA secondary structure of the deconstructed genome-1 (DG-1) of CGMMV is generated through the Mfold web-server. The structural conservation of 5' and 3' terminal end of DG-1 (~2.0 Kb) (b) depicts similarity with that of wild type genome of CGMMV (6.4 Kb) (a)

4.1.2.1.2. Cloning of deconstructed virus genome

The overlapping extension PCR technique was used for the generation of site-specific deletion within a particular DNA sequence (**Fig.4.11a**). The special primers set with partially overlapping ends were used at high annealing temperature (72˚C) to ensure the site-specific binding at target sequence and accurate deletion and subsequent ligation of the plasmid DNA constructs; thus mostly preferred as the restriction digestion and ligation free cloning technique. The synthesized PCR amplified molecules wrapped around the plasmid and circularized with overlapping ends but remained non-covalently closed. The original plasmid molecules were degraded with DpnI enzyme treatment, and leftover, non-covalently closed, circular PCR amplicons were transformed into the *E.coli* competent cell to seal the nick at both strands using *in vivo* endogenous recombinase activity of the bacterial cell. Very few colonies were obtained and confirmed based on colony PCR and plasmid-based PCR (**Fig. 4.13a**). The constructs were further verified using restriction digestion using BamHI and XbaI with wild pBP4 plasmid as a positive control (**Fig. 4.13b**). The DG-1 (1988nt) was cloned into pBM-1 plasmid which was a modified pCambia backbone containing 2x35s promoter near to left border and ribozyme (Rz) site followed by a *Nos* terminator before the right border (**Fig.4.11b).** was confirmed through restriction digestion with BamHI and XbaI.

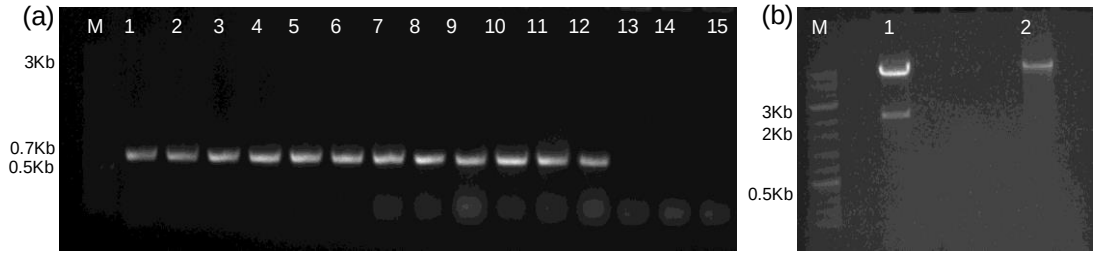


Fig. 4.13. Development of deconstructed genome-1 using CGMMV genome (BP4 cDNA infectious clone). (a) The colony PCR based confirmation of the deconstructed genome using BM-1182F and BM-1173R primer which produces 768 bp amplicon for the DG-1 and 5204bp long amplicon for BP4. Colonies 1-12 are tested positive with rest (13-15) are negative. (b) The restriction digestion of the DG-1 plasmid construct with BamH1 and XbaI restriction enzyme. 1 denoting the digested plasmid with ~2.0 Kb insert (DG-1) and 2 is the uncut plasmid.

4.1.2.2. Development of deconstructed genome (DG)-2 based on CGMMV genome

4.1.2.2.1. Designing and synthesis of the DG-2 Construct

An attempt had been made to design another construct that possesses a significant part of the methyltransferase (MT) domain to enhance the replication rate of CGMMV genome-based deletion constructs. The design was prepared in such a way that the whole CGMMV genome sequence was parted into major three segments, viz., the N-terminal part containing 5'UTR with 87% of MT domain (1-1173nt), a central part with putative MP-subgenomic promoter ranging from 4900 nt to 5080 nt, and the C' terminal part carrying CP coding sequences with CP-SGP and3' UTR (5613 nt to 6424 nt) of the genome. These segments were joined together, to form deconstructed genome-2 (DG-2) that was presented as a miniature CGMMV. Based on the secondary structure analysis of CGMMV complete genome by using various web-tools, some consensus regulatory sequences were identified within 1-1123 nt region keeping-terminal ends; thus to keep all these sequences, 1173nt, i.e., 371aa of MT domain. Further, 90nt both in the upstream and downstream of MP-TLSS (+1) was selected based on our *in silico* prediction of MP-SGP sequences. The core MP promoter was retained to utilize this promoter for functional genomic study. Lastly, a long stretch of genomic segments from the 3' terminal end (5613-6424 nt) containing CP coding sequences along with 150nt in the upstream was retained to maintain the full functionality of CP-promoter. The conserved nucleotides necessary for the formation of pseudoknots and tRNA-like structure in the 3UTR was also included without any modification, as the structural conservation of 3'UTR is necessary for binding of RdRp domain and subsequent replication of the genome. Finally, the RNA secondary structure of the DG-2 was generated from Mfold to check the structural integrity of the newly designed genome and its similarity with wild type CGMMV; so that to achieve higher replication and translation rate. Notably, numbers of restriction sites were incorporated inside the DG-2 genome sequence to make it more flexible for exploiting multiple purposes. The 714 nt sequences of enhanced green fluorescent protein (gene) was also included after 105nt from CP-TLSS (+1), which was previously identified as the active core promoter length of CP. Ultimately, the ORF search on this deconstructed genome depicted the existence of 3 functional overlapping ORFs; ORF1 (1245nt) encoding 98% part of the Methyltransferase domain, ORF2 (276 nt) encoding 68% of MP, and ORF3 (1239nt) encoding CP fused with eGFP (412aa). The detail of the genome map (2961nt) of DG-2 was presented in **Fig. 4.14a**. The RNA secondary structure analysis of DG-2 confirm its stability and similarity with the wild type CGMMV (**Fig. 4.15**). Finally, the DG-2 sequence was synthesized together with 2X35s promoter and a ribozyme (Rz) site

followed by a nopaline synthase (*Nos)* terminator (**Fig. 4.14b**.) through outsourcing from Bio Basic Canada Inc.

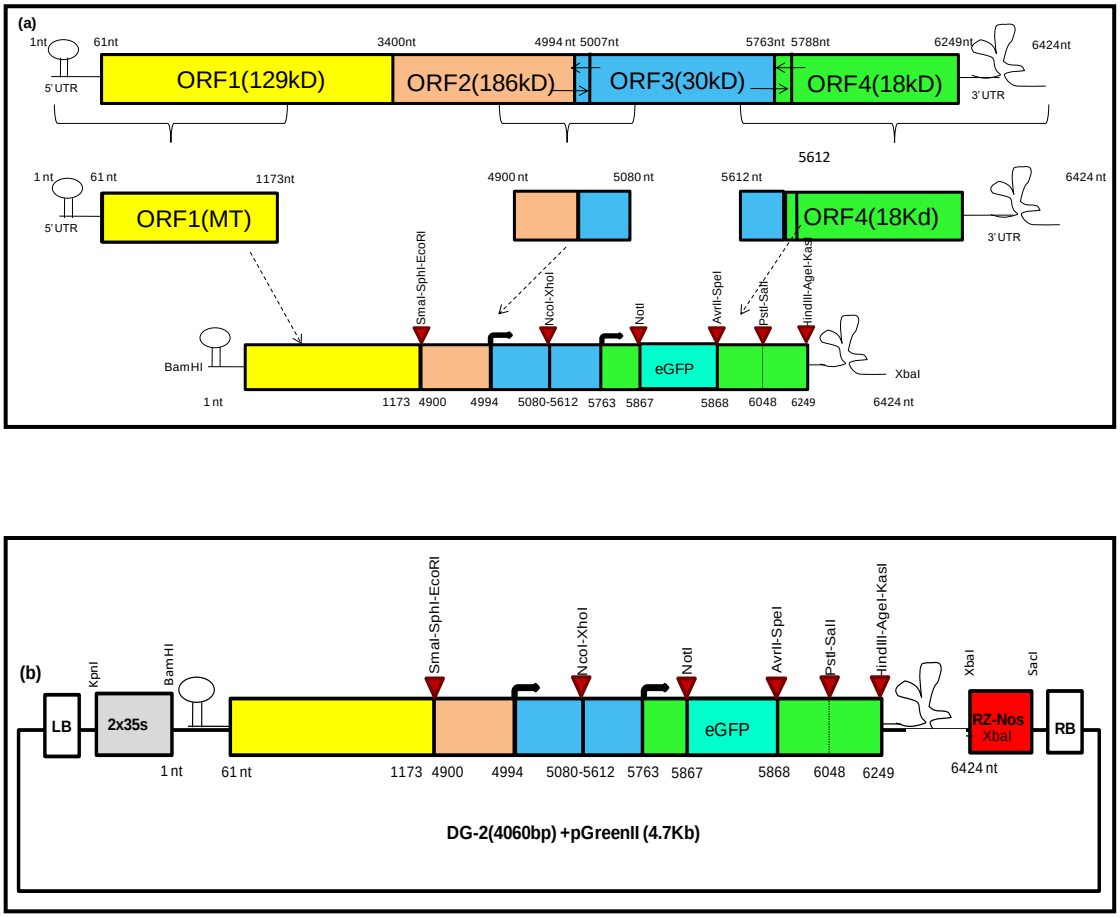


Fig. 4.14. Designing and cloning of deconstructed virus genome-2 of CGMMV. (a) Design of deconstructed genome-2 based on CGMMV full genome. The deconstructed genome-2 was designed based on our *in silico* prediction. It contains 1173 nt of 5' terminal and 5612 nt of 3' terminal of the genome with an internal fusion of putative MP-subgenomic promoter region (4900-5080 nt). (b). cloning of deconstructed genome-2 of CGMMV into the pGreenII vector was done along with 2x 35S promoter in the upstream of 5' terminal end, and a ribozyme (Rz) site-nos terminator in the downstream of 3' UTR. The entire cassette was inserted in between the left border (LB) and right border (RB) of pGreen II backbone with KpnI and SacI restriction sites.

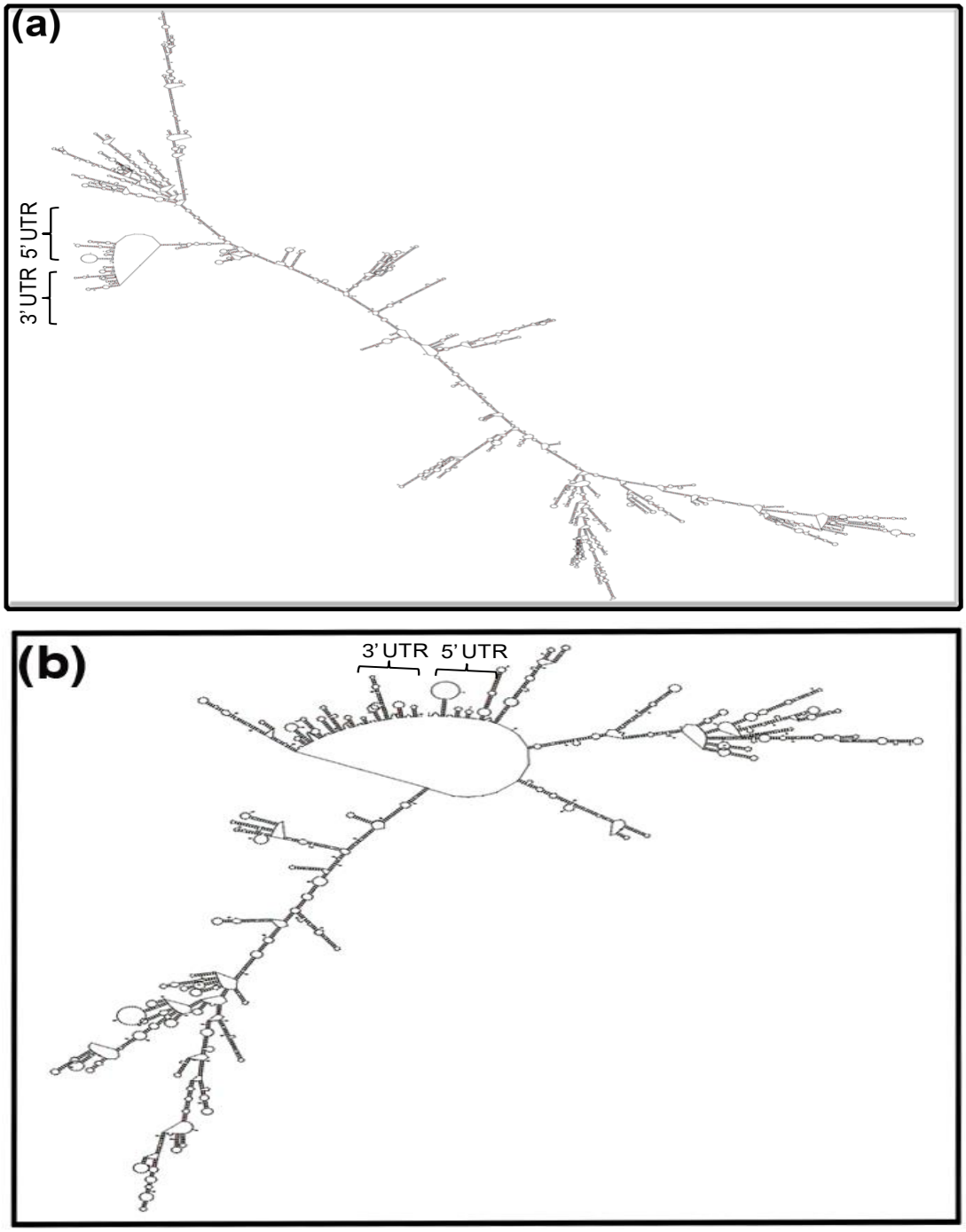


Fig. 4.15. The RNA secondary structure of the deconstructed genome-2 (DG-2) of CGMMV is generated through the Mfold web-server. The structural conservation of 5' and 3' terminal end of DG-2 (2.9Kb) (b) depicts similarity with that of wild type genome of CGMMV (6.4Kb) (a).

4.1.2.2.2. Cloning of DG-2 in pGreenII binary vector and transformation into Agrobacterium GV3101

The synthetic genome construct (4062nt) containing the 2x35s promoter, DG-2 genome sequence, and ribozyme (Rz) site with a Nos terminator was sliced out from the pUC57 vector (2.7Kb) backbone with KpnI and SacI restriction sites (**Fig. 4.16a)**, and then directly cloned into pGreen II (0029) with the same restriction sites (**Fig. 4.16b)**. It removed the entire MCS from pGreenII (0029) backbone (4632 bp); thus will facilitate us to utilize these sites, further, for DG-2 genome modification

The colony PCR positive clones were further verified through restriction digestion using BamH1 and HindIII restriction enzymes (**Fig. 4.16c)**. It released desirable 2.78kb fragments from the vector (4.7Kb) backbone. The confirmed plasmid constructs were transferred into the Agrobacterium GV-3101 together with pSoup helper plasmid *via* electroporation. The co-transformation of pGreenII and pSOUP was done by using an equal concentration of both, and only Kanamycin (not tetracycline) and Rifampicin were used for the selective screening of the transformed Agrobacterium colonies. The transformed Agrobacterium colonies were confirmed through colony PCR(**Fig. 4.16d)**. The positive colonies were maintained for further use.

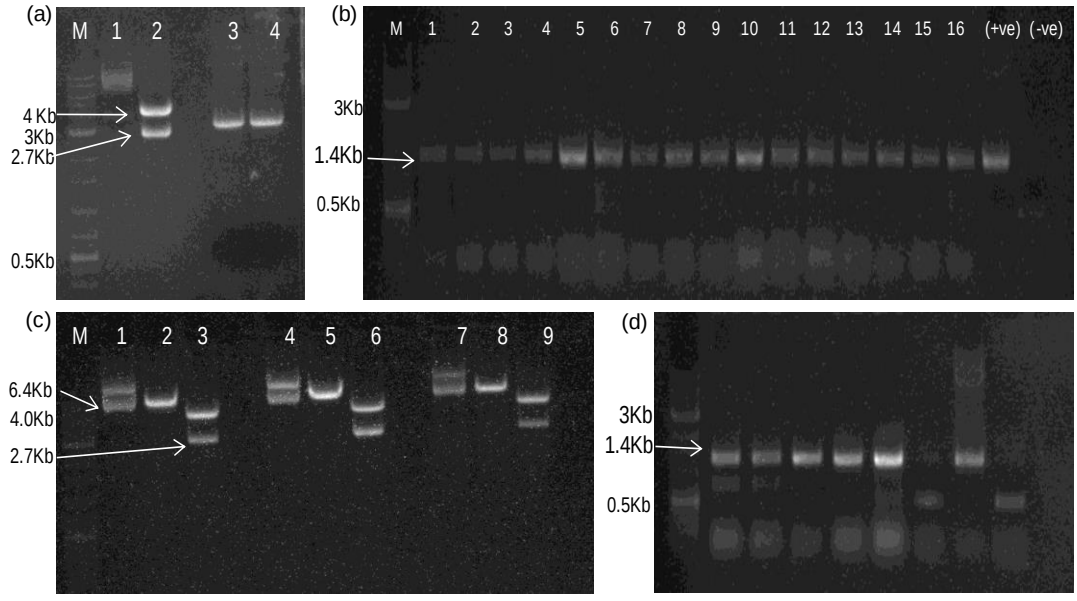


Fig. 4.16. Cloning of deconstructed genome-2 synthetic construct from pUC57 to pGreenII 0029 vector. (a) The DG-2 was taken out from the pUC57(2.7kb) through restriction digestion restriction enzyme KpnI and SacI restriction enzyme. The digestion leads to the released 4062 bp long insert carrying 2x35s promoter, DG-2 genome (2961nt), and a ribozyme (Rz) site with

Nos terminator. (b) Colony PCR for the detection of DG-2 constructs in pGreen II depicted to ensure the insertion of the entire synthetic gene cassette into pGreenII 0029 vector (4.6Kb) with KpnI and SacI restriction sites. The BM-1178F and BM-489R primer binding at 5' terminal of CP and 3' terminal end of 3'UTR of CGMMV, respectively, amplify the 1.4kb insert of DG-2 carrying eGFP (714 bp) fused to CP (486 nt) and 3'UTR (176 nt) against 662 nt long amplicons of wild CGMMV. (c) Restriction digestion of DG-2 from pGreenII backbone using BamH1 and HindIII released 2.786 Kb large internal fragments from the entire cassette and proved the stable insertion of DG-2 inside the pGreenII backbone.

4.2. Objective: To evaluate the trans-replicability and translation ability of deconstructed genome in presence of wild type CGMMV

4.2.1. *In planta* replication, movement and expression ability of the deconstructed genome-1 of CGMMV in the different host using their infectious clones

4.2.1.1. Detection assay of DG-1 in different hosts when co-infiltrated and staggered-infiltrated with wild type CGMMV

Initially, the terminal primers (486F and 489R) were used for the detection of replicating dRNA of DG-1, and the replicating behavior of the replicon was studied in the infiltrated leaves *N. benthamiana* at a different time interval. The DG-1 replicon was not detected from the inoculated tissue when infiltrated alone; thus indicating the loss of infectivity and self-replicating ability of DG-1. The replicon was trans-replicated with the help of wild type CGMMV, only when DG-1 is infiltrated in the CGMMV infected leaves. But, dRNA replicon was not identified once DG-1 co-infiltrated together with wild type CGMMV. It indicated that the CGMMV did not support the replication of DG-1 immediately after co-inoculation (**Fig. 4.17**). To validate this, the experiment was repeated twice.

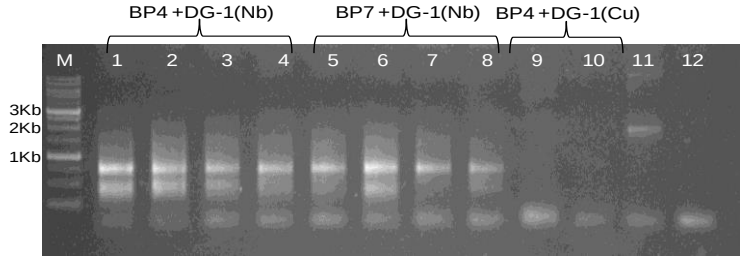


Fig. 4.17. Detection of DG-1 constructs in *N. benthamiana* and cucurbits when co-infiltrated along with helper CGMMV constructs (BP4, and BP7). RT-PCR was performed by using BM-486F and BM-489R primer. DG-1 replicon was not detected when co-infiltrated with the 'helper' CGMMV. Lane **1**: Infiltrated (0) leaf (*N. benthamiana-1* infected with symptomatic BP4); **2**:1st systematic leaf (*N. benthamiana* -1infected with symptomatic BP4); **3**: Infiltrated (0) leaf (*N. benthamiana*-2 infected with symptomatic BP4); **4**:1st systematic leaf (*N. benthamiana* 2 infected with symptomatic BP4); **5**: Infiltrated (0) leaf (*N. benthamiana* -1 infected with asymptomatic BP7); **6**: 1st systematic leaf (*N. benthamiana* -1 infected with the asymptomatic clone); **7**: Infiltrated (0) leaf (*N. benthamiana* -2 infected with asymptomatic BP7 clone); **8**: 1st systematic leaf (*N. benthamiana* -2 infected with asymptomatic clone BP7); **9**: 1st systemic leaf (Cucumber-1); **10**: 1st systemic leaf (Cucumber-2); **11**: positive control (DG-1 plasmid); **12**: negative control.

Further, the DG-1 plasmid construct was infiltrated in the *N. benthamiana,* after 7days of post-infiltration of CGMMV. The DG-1 replicon was detected from the infiltrated leaf as well as in the systemic leaf of CGMMV infected *N. benthamiana* (**Fig. 4.18**). But its accumulation within the infected tissue varied greatly; the lack of uniform distribution of replicons within infiltrated tissue might be due to the disruption of movement protein. Here, both the symptomatic as well as asymptomatic strains of CGMMV acted as the helper virus for replicating the DG-1.

Furthermore, a specific duplex PCR-based detection system was tried to develop for the simultaneous detection of the DG-1 along with its helper CGMMV and to avoid the amplification of multiples fragments that had been faced during the use of terminal primers (i.e., BM-486F & BM-489R). For that, a new primer set was designed by a combination of three primers, viz., BM-1171F, BM-1182F, and BM-1173R. The BM-1171F and BM-1173R were used to detect the helper CGMMV amplifying 1415bp long amplicon, whereas BM-1182F with BM-1173R is used for the specific detection DG-1 replicon (**Fig. 4.19**). The BM-1171F did not bind to DG-1 as its coordinates (4591- 4616nt) is deleted from the CGMMV; whereas BM-1182F can able to bind to wild type also, but it generated 5204 nt amplicons (for BP4) that can be avoided by restricting the extensions time of PCR within 1min.

After PCR, the desirable bands were visualized along with non-specific amplification; thus this duplex PCR system was found to be not suitable for the specific and simultaneous detection of helper CGMMV and DG-1 replicon.

To confirm the trans-replication of DG-1 in CGMMV infected cucurbits, the agrobacterium culture of DG-1 was infiltrated into the fleshy cotyledon of various cucurbits, viz., cucumber, bottle gourd, and watermelon, 7days after the infiltration of symptomatic CGMMV infection clones (BP4). The RNA was extracted from the symptomatic systemic leaves, and RT-PCR based detection depicts multiple non-specific amplifications. Although the amplification of DG-1 was visualized (**Fig. 4.20)** when duplex RT-PCR was performed by using BM-1171F, BM-1182F, and BM-1173R, it was undetected with the terminal primers (i.e., BM-486F & BM-489R). Thus, trans-replication of DG-1 replicon by CGMMV remained unconfirmed in cucurbits. The trans-replication of DG-1 was not evident in any CGMMV infected cucurbitaceous hosts like bottle gourd, cucumber, and watermelon, where the trans-replication of DG-1 was not supported by its wild types (**Fig. 4.21**). Lack of proper delivery methods might be the reason. Further clarification is also necessary.

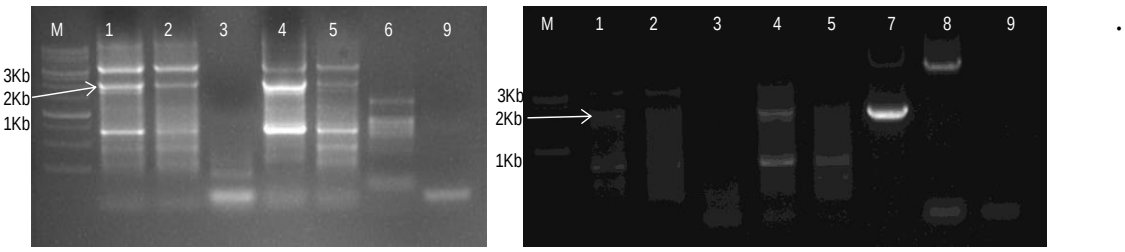


Fig. 4.18. RT-PCR based detection of DG-1 constructs, when infiltrated into the CGMMV infected *N. benthamiana* using primer BM-486F and BM-489R. Amplification of DG-1replicon from the CGMMV infected *N. benthamiana*. Lane 1, 2 represents the DG-1 infiltrated leaf and the first systemic leaf of BP4 infected *N. benthamiana*. Lane3 is the only DG-1 infiltrated leaf of healthy *N. benthamiana.* Lane 4, 5 represents the DG-1 infiltrated leaf and the first systemic leaf of BP7 infected *N. benthamiana*. Lane 6 is the Only BP4 (CGMMV) infected *N. benthamiana.* Lane 7 is the DG-1 plasmid control, Lane 8 is the BP4 plasmid control, and Lane 9 is the reagent control (negative).

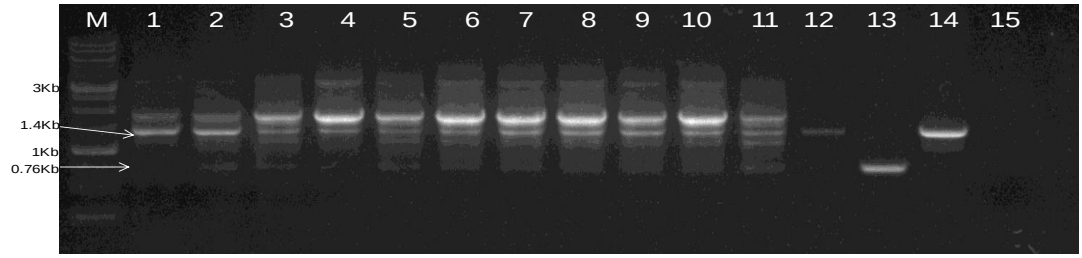


Fig. 4.19. Duplex-PCR based detection of DG-1 replicon along with helper CGMMV when DG-1 infiltrated into the CGMMV infected *N. benthamiana.* The RT-PCR protocol was standardized by using a combination of three primers BM-1171F, BM-1182F, and BM-1173R**.** Lane 1 to 11 represents the co-existence of DG-1 replicon along with BP4 in different *N. benthamiana* plants. Lane 12 is the only BP4 infected *N. benthamiana*. Lane 13 is DG-1 plasmid control; lane 14 is the BP4 plasmid control, and lane 15 is the negative (reagent) control.

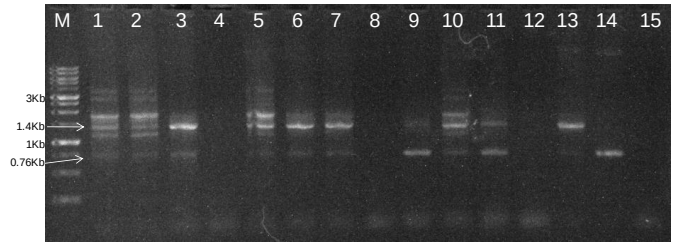


Fig. 4.20. Duplex-PCR based detection of DG-1 replicon along with helper CGMMV when DG-1 infiltrated into the CGMMV infected *N. benthamiana* and cucurbits. The RT-PCR was performed using a combination of three primers BM-1171F, BM-1182F, and BM-1173R**.** Lane

1 to 3 represents the CGMMV infected cucumber plants with healthy plants in lane 4; lane 5-7 represents the CGMMV infected bottle gourd plants with healthy plants in lane 8; lane 9-11 represents the CGMMV infected watermelon plants with healthy plant in lane 12. Lane 13 is DG-1 plasmid control; lane 14 is the BP4 plasmid control, and lane 15 is the negative (reagent) control.

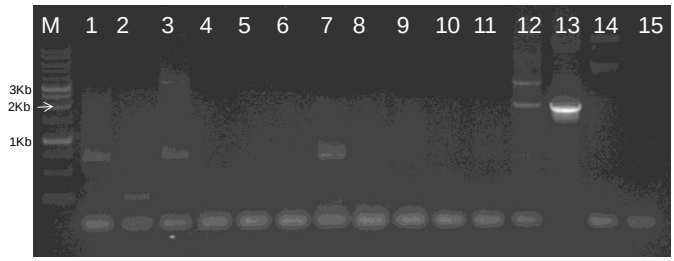


Fig. 4.21. RT-PCR detection of DG-1 replicon in CGMMV infected cucurbits in comparison with CGMMV infected *N. benthamiana* using BM-486 F & BM-489R. DG-1 replicon was not detected in lane 1-11, despite its infiltration in the various CGMMV infected cucurbits. Lane 1 to 3 represents the CGMMV infected cucumber plants with its healthy control in lane 4; Lane 5-7 represents the CGMMV infected bottle gourd plants with its healthy control in lane 8; Lane 9-11 represents the CGMMV infected watermelon plants. The DG-1 replicon was only detected in lane 12 depicting the CGMMV infected *N. benthamiana*. Lane 13 is DG-1 plasmid control; lane 14 is the BP4 plasmid control, and lane 15 is the negative (reagent) control.

The experiments were repeated thrice and a total of 6 plants per replication were tested by RT-PCR. DG-1 was detected in all 18 plants that were infected with either symptomatic (BP4) and asymptomatic (BP7) CGMMV infectious genome construct **(Table 4.8.)**. Thus it was concluded that both the symptomatic (BP4) and asymptomatic (BP7) CGMMV was efficient to support DG-1 for its replication and movement. Whereas, DG-1 was unable to replicate individually.

Table 4.8. RT-PCR based detection of DG-1 constructs infiltrated in the CGMMV infected *N. benthamiana* and cucurbits

Host	BP4[@] + DG1	BP7[$]+ DG-1	BP4	BP7	DG-1	Healthy[#] Control
N. benthamiana	18/18[*]	18/18	0/18	0/18	0/18	0/18
Cucurbits	0/18	0/18	0/18	0/18	0/18	0/18
Bottle Gourd	0/18	0/18	0/18	0/18	0/18	0/18
Water melon	0/18	0/18	0/18	0/18	0/18	0/18

*No of plants showing positive result/ total number of plants tested (in tri replicate)

[@]BP4 is the name of the symptomatic infectious clone of CGMMV

[$]BP7 is the name of the asymptomatic infectious clone of CGMMV

Healthy control represents the buffer (without agro-construct) infiltrated plant

4.2.1.2. Transreplication and movement assay of DG-1 in the CGMMV infected *N. benthamiana*

4.2.1.2.1. Replication of DG-1 in CGMMV infected *N. benthamiana*

To prove the replication of the DG-1 within the infiltrated tissue, the progression in the replication rate of DG-1 was examined at a different time interval, and the increasing trend of replication was evident with a semi-quantitative PCR. The DG-1replicon was reached at a detectable level within 48hrs of infiltration. Starting from infiltration, its concentration increased to attend the plateau within 5th to 6th DPI and either remained stable or declined gradually. The production of subgenomic RNAs and other progeny RNAs from the wild-type genome remained almost constant through this period as evident from Fig. **4.22.b**.; thus signified the very limited influence of DG-1 on the functionality of wild type CGMMV.

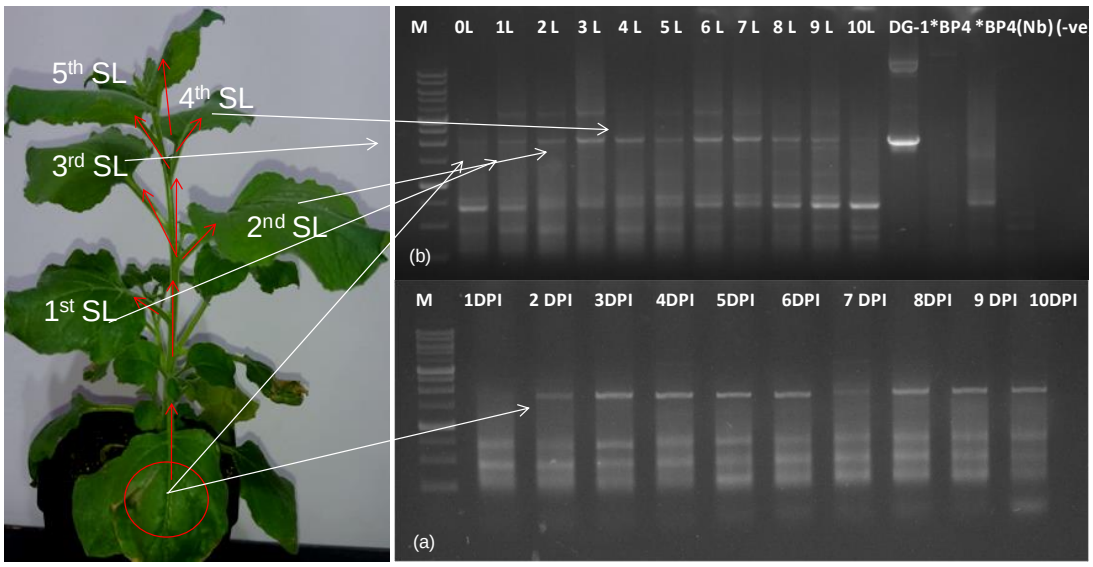


Fig. 4.22. Trans-replication and systemic movement of DG-1 replicon in the CGMMV infected *N. benthamiana*. (a) The replication DG-1in the infiltrated leaf (0L) of CGMMV infected *N. benthamiana* showed a gradual increment in the concentration of DG-1 up to 6DPI; after then its concentration either remain stable or decreased gradually. (b) After replication, the DG-1 replicon was moved systemically into the systemic leaves, and then subsequently spreads into the whole plant. It was tested up to 10th SL of *N. benthamiana*.

4.2.1.2.2. Systemic movement of the deconstructed genome

To determine the *in planta* systemic movement of the DG-1 replicon, the different leaves of N. *benthamiana*, positioned upward from the infiltrated one were tested. These leaves were designated as L1 to L10 based on their relative position with respect to infiltrated (L0) one. The RNA was extracted separately, from the individual leaves at 14DPI and tested based on RT-PCR using BM-486F and BM-489R primer pairs. The DG-1 replicon (~2.0Kb) was detected

to be distributed in all the studied leaves (1L to10L) of *N. benthamiana* (**Fig. 4.22a**). This systemic movement of DG-1 replicon was supported by both the wild type 'helper' virus, i.e., symptomatic (BP4) and asymptomatic (BP7) CGMMV.

Both this trans-replication and movement assay of DG-1 as performed based on RT-PCR using BM-486F and BM-489R primers that generates ~2.0Kb amplification, along with the amplification sub-viral genomic fragments. As the DG-1 has 100% homology with the wild type CGMMV genome, thus very difficult to differentiate. To resolve that issue, a foreign sequence was chosen to insert within the DG-1 construct.

4.2.1.3. Vector function assay of DG at 7DPI of CGMMV infected *N.benthamiana*

4.2.1.3. 1. Insertion of partial *NbPDS* within DG-1 construct

To evaluate the efficiency of DG-1, the partial fragment of the NbPDS gene was inserted into the DG-1 plasmid construct in the sense orientation, after the CP coding frame. A small partial sequence (227 bp) of *NbPDS* gene was amplified from its mRNA transcript and inserted within the genome through an overlapping extension PCR strategy. A long primer pairs were designed that have two parts containing the sequences of the *NbPDS* gene along with overhang parts (P1& P2) to the sequence of the viral vector/VIGS vector. Two round of PCR was performed. In round one, the primers were used to amplify the desirable *NbPD*S gene that retains the overhang flanking regions (P1 & P2). The P1 is homologous to the target region of the vector sequence whereas P2 is complementary to the subsequent region of the viral vector. The purified PCR amplicons were then used as the primer for the next PCR, in which the overhang parts (P1 and P2) facilitated the specific binding of the insert to the virus vector. In this PCR, a third primer (P3) was added that acted like a reverse primer containing the few complementary sequences of joining ends of insert and vector. P3 was used to enhance amplify the desired VIGS vector, which contains PDS gene of interest incorporated in the final PCR product as a circular PCR molecule (PDS + vector) with nicked ends. The circular PCR molecule contains a circular plasmid which was treated with DpnI enzyme resulted in plasmid molecules were degraded and circular PCR molecule transformed into the E.coli competent cell and resulted in the appearance of few colonies. All colonies were further confirmed by colony PCR resulted in all positive colonies (**Fig. 4.23a**), and further validated through restriction digestion of the plasmid with BamH1 and XbaI restriction enzyme (**Fig. 4.23b**),. On restriction digestion, 2.227Kb fragment was released from DG (PDS)-1 which was comparable with 2.0 Kb size of DG-1. The newly generated construct was named DG(PDS)-1.

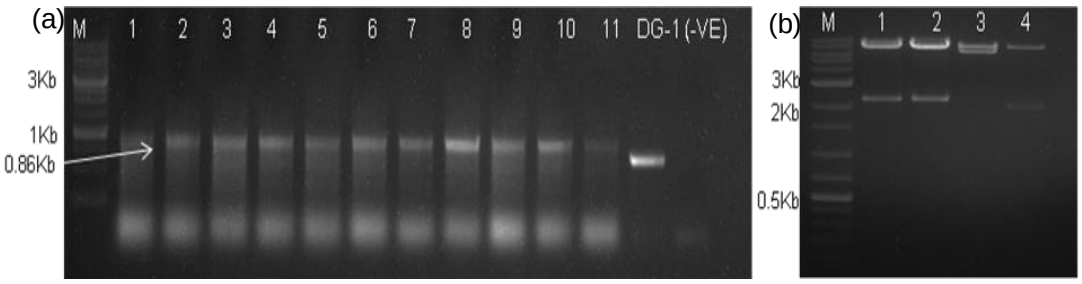


Fig. 4.23. Insertion partial *NbPDS* gene into the DG-1 construct. (a) The colony PCR based confirmation of the deconstructed genome using BM-1178F and BM-489R primer which produces 862 bp amplicon for the DG(PDS)-1 and 662bp amplicon for DG-1. Colonies 1-11 are tested positive. (b) The restriction digestion of the DG-PDS1 plasmid construct with BamH1 and XbaI restriction enzyme. Lane 1, 2 denoting the digested plasmid with ~ 2.2 Kb insert, lane 3 is the digested BP4 plasmid with 6.4Kb insert, and lane 4 is the digested DG-1 plasmid with 2.0Kb insert.

4.2.1.3.2. Functional assessment of DG (PDS)-1 in the CGMMV infected *N. benthamiana*

Phenotypic changes:

The DG (PDS)-1 construct was agroinfiltrated into the CGMMV infected *N. benthamiana* plant in the very early (two leaves) growth stage. It resulted in the development of photobleaching phenotype in about 30% of plants that were showing typical phenotypic changes starting from the 15 DPI and extended up to 45 dpi (**Fig. 4.24**). Initially, few small yellowish-white spots were appeared in the lamina and started expanding gradually. Most of the photobleaching appeared around the veins. Although these phenotypic changes are visible only on few leaves of a plant, it was distributed randomly up to the top leaves of the plant, indicating the long-distance systemic movement of DG(PDS)-1 construct. The frequency and intensity of gene-silencing are higher when symptomatic CGMMV (BP4) was used as the helper, in comparison to asymptomatic CGMMV (BP7) as the helper virus (**Table 4.9.)**. It also suggested the possible difference in the replication rate of DG(PDS)-1 to its helper CGMMV. This outcome exhibited the phenotypic evidence for the functional significance of the DG-1 construct as the successful VIGS.

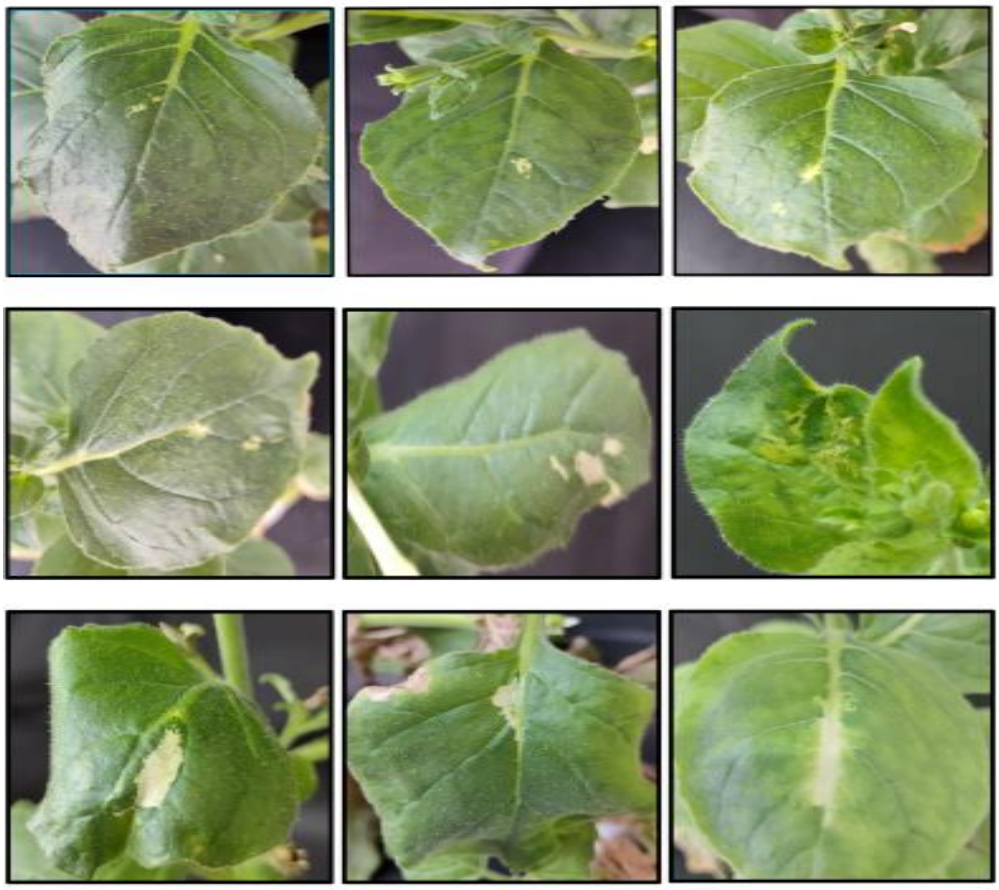


Fig. 4.24. The evidence of photobleaching symptoms on the different leaves of *N. benthamiana* produced by DG(PDS)-1. The symptoms are scattered and observed on the few leaves only.

Table 4.9. Phenotypic assessment of gene silencing using DG-1 replicon in the CGMMV infected *N. benthamiana* at 15-60 DPI

Treatment	No of plant showing photobleaching symptoms/ Total no of plants infiltrated (%)
BP4 + DG(PDS)-1	6/20(30)
BP7 + DG(PDS)-1	3/20(15)
Healthy Control	0/10(0)

[1]The frequency of gene silencing was represented based on the per-cent of plants showing photobleaching phenotype (in replication 1, and 2)

[2]The effectiveness of gene silencing was quantified as the per-cent of leaves showing photobleaching phenotype.

Detection of replication and movement of DG (PDS)-1 in the CGMMV infected *N. benthamiana*

To confirm the replication and movement of DG (PDS)-1 with the help of wild type CGMMV in *N. benthamiana*, the production of RNA replicon from DG (PDS)-1 construct was detected by RT-PCR. To confirm the replication rate of DG (PDS)-1 with respect to its helper CGMMV within infiltrated tissue, a duplex PCR was standardized using two sets of primers, in which one primer set (96F & 1268R) was used for the specific amplification of helper CGMMV, whereas another primer set (1180F & 1269R) was specific for DG(PDS)-1. DG (PDS)-1 was specifically detected by amplifying the CP fused with PDS fragment by using CP forward (BM-1180F) and PDS reverse primer (BM-1269R). The specific amplification of DG (PDS)-1 was detected within 0 to 9DPI period (**Fig 4.25**), but the intensity of amplicon was significantly less than the well-established helper CGMMV. Further, over the time-course analysis, the replication rate was stable up to 7DPI, but gradually decreases afterward It indicated the pattern of trans-replication of DG (PDS)-1 with respect to its helper CGMMV.

As per our observation, the DG-1 was replicated within the infiltrated tissue. with the initiation of its replication, it was moved systemically from infiltrated leaf (0L) to the first systemic leaf (1L) within 2-3DPI and subsequently reached the 2nd systemic leaf after 3DPI. In this manner, it was systemically distributed within different parts of the plant as evident from the photobleaching of leaves. Although the long-distance systemic movement of DG-1 was evident, the local cell to cell movement within a leaf lamina was not uniform as visualized from their day-wise distribution pattern in the 1st and 2nd systemic leaf. Thus, it could be plausible that the movement protein of helper CGMMV may not assist in the uniform cell to cell local distribution of the deconstructed replicon within the infected tissue of the leaf.

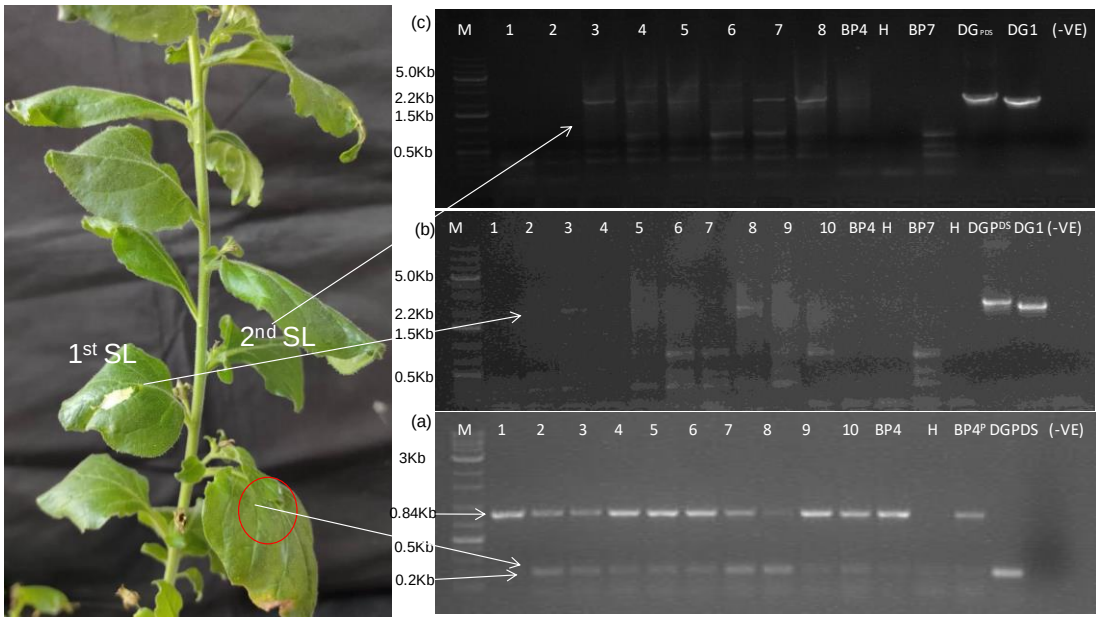


Fig 4.25. The trans-repliction and movement of DG(PDS)-1 with the help of wild type CGMMV in *N. benthamiana*. (a) Duplex PCR based detection of CGMMV and DG (PDS)-1 replicon in the infiltrated tissue of *N. benthamiana*. The RT-PCR was performed by using BM-96F, and BM-1268R primer pairs specifically detecting CGMMV infection, whereas another primer set (BM-1180F and BM-1269R) is used for the specific detection DG (PDS)-1 replicon. The replication pattern of DG (PDS)-1 was detected 0 to 9DPI period. Here, Lane1-11 represents the samples (leaf disk) harvested at 0DPI, 1DPI, 2DPI, 3DPI, 4DPI, 5DPI, 6DPI, 7DPI, 8DPI, 9DPI, and 9DPI from the 10 specified leaves of 10 infiltrated plants. Lane 11 is the only symptomatic (BP4) CGMMV infected plant. Lane 12 is the healthy plant; whereas lane 13, 14 is the BP4 and DG (PDS)-1 plasmid control with a reagent (negative) control in lane 15. (b) The DG(PDS)-1 replicon was detected from 2-3 DPI onwards in the 1st systemic leaves that were tested up to 1-10DPI, whereas it is absent in helper type CGMMV infected plant (BP4, BP7), and health plant (H). (c) Similarly, DG(PDS)-1 replicon was detected in 2nd systemic leaf from 3 DPI onwards and tested up to 8 DPI. In the whole study DG(PDS)-1 and DG-1 were used as the plasmid control and (-) ve was reagent control.

4.2.1.3.2. Quantification of the replication rate of the deconstructed genome

To check the replication pattern of DG(PDS)-1 replicon, its relative quantification was assessed with the help of qRT-PCR keeping the host Actin gene (*NbActin*) as the reference gene. Simultaneously, the relative quantification of helper CGMMV (BP4) was also done to check the impact of DG(PDS)-1 on its helper virus. The 5mm leaf disks were taken out from a fixed infiltrated leaf of 10 different plants at a 24hrs interval starting from 0DPI and continued up to 6DPI. RNA was extracted from the harvested tissue using an RNA extraction kit. The quantification of DG(PDS)-1 was performed using the primers (1180F and 1269R) designed from the CP-PDS gene of the DG(PDS)-1 construct, whereas, BP4 was quantified using a primer set (BM-1267F and BM-1268R) designed from ORF2 of CGMMV, that was absent in

the DG(PDS)-1 construct. Based on the quantitative analysis, it was observed that the DG-1 was replicated over the 0-6DPI time course, and its concentration increased up to 4.5 fold over the actin within 5DPI of infiltration; thereafter decreases gradually or remain in the steady-state (**Fig 4.26)**. It indicated the minimal support from the helper virus to carry forward the trans-replication of DG(PDS)-1. On the other hand, immediately after the infiltration of DG(PDS)-1, the replication rate of BP4 was decreased tremendously and reduced up to 0.0096 and 0.917 fold within the first two days (1DPI, 2 DPI) of the infiltration of DG(PDS)-1 construct in comparison to its concentration at 0DPI. It signified the influence of DG(PDS)-1 replicon on the replication of BP4, immediately after its infiltration. But, from the 3DPI onwards, the concentration of BP4 either increases or remains stable. Thus, although the replication of help CGMMV was affected by DG(PDS)-1 replicon in the initial few days of infiltration, its self-replication ability helped in maintaining its own-replication rate later on. This comparative analysis indicated the requirement of standardization of time frame for delivering the deconstructed genome and its helper virus. Further, refinement in the delivery system was also felt.

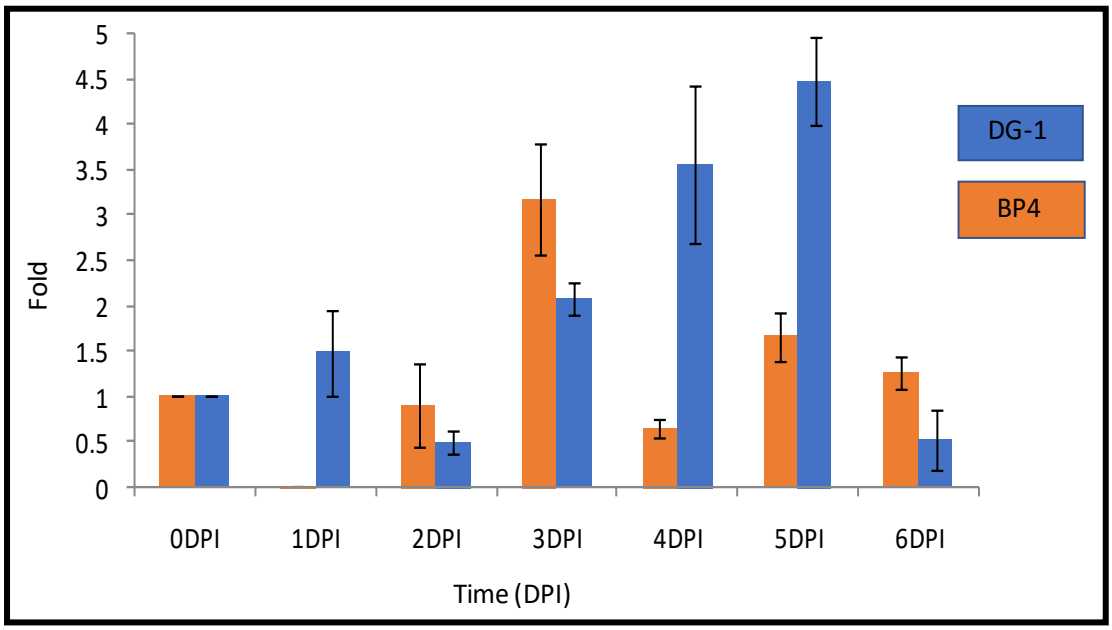


Fig 4.26. The quantitative presentation showed the replication pattern of DG(PDS)-1 replicon and its helper CGMMV (BP4) in *N. benthamiana* over the 0-6 DPI temporal scale. The test was performed using RNA samples extracted from infiltrated leaves of 10 different plants. The comparative analysis of fold change in the concentration of DG-1 replicon and helper CGMMV was calculated with reference to their concentration at 0DPI of DG(PDS)-1 in the CGMMV infected plant. The actin gene of *N. benthamiana* was used as the reference.

4.2.1.3.4. Detection of the deconstructed genome by Northern blotting

To confirm the trans-replication and movement of the DG(PDS)-1 replicon, Northern blot was performed using the pure RNA extracted from the different leaf samples of *N. benthamiana*. The GC rich 3' terminal of CGMMV was used as the preparation of 225 nt long probes. To determine the probe sensitivity, the total RNA was diluted in different concentrations and dot blot was performed and the probe sensitivity was detected up to picogram concentration. Further, the same probe was used for detecting the DG(PDS)-1 replicon along with viral gRNA and sgRNA. The replicon was detected in the infiltrated as well as the first systemic leaf of *N. benthamiana*; proving the replicating nature of DG(PDS)-1 with the help of wild type CGMMV. The expression of DG(PDS)-1 is very limited in comparison with its helper genome. The excessive expression of two sgRNAs was evident in the Northern blot (**Fig 4.27)**

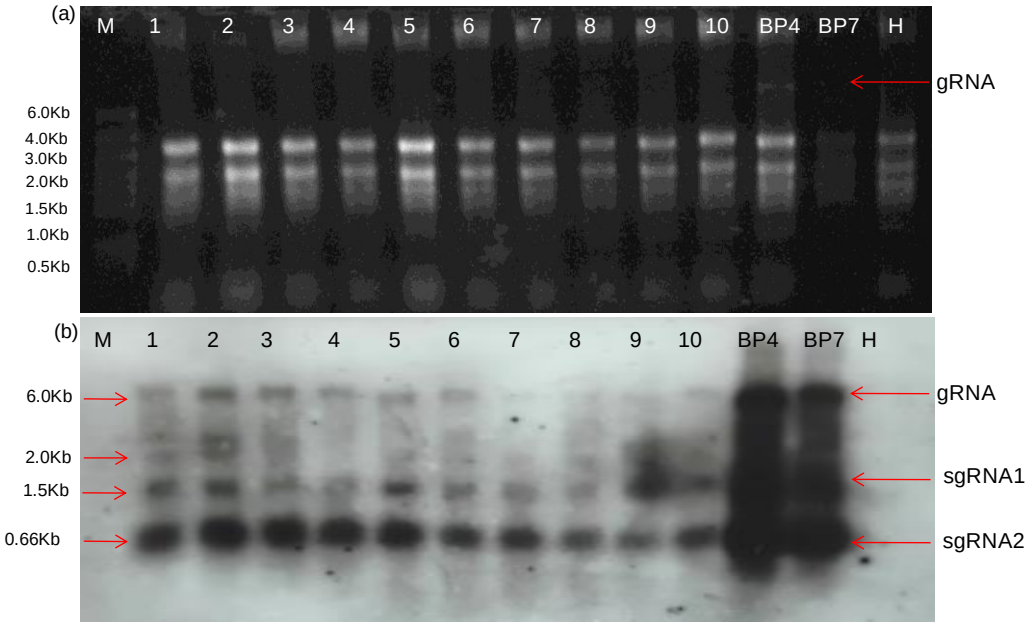


Fig 4.27. The accumulation of DG-1 replicon (2.0kb) was evident from the Northern blot assay. (a) The separation of RNA in the denaturing gel. The RNA was extracted from the 1st to 10th leaves (as numbered) of CGMMV infected *N.benthamaina* infiltrated with DG(PDS)-1. The BP4 and BP7 represent the symptomatic and asymptomatic CGMMV infected *N.benthamiana* without DG(PDS)-1 infiltration, whereas the RNA of health plant sample (H) is used as the control. The gRNA and sgRNA denote the genomic RNA and subgenomic RNA, respectively. (b) Pictorial depiction of northern blot indicated the presence of DG(PDS)-1 in the 1st and 2nd systemic leaf (Marked by 2.0Kb).

4.2.2. *In planta* replication, movement, and translational ability of the deconstructed genome-1 of CGMMV in the different hosts using their infectious clones.

4.2.2.1. Trans-replicational ability of DG2 construct

The agro-culture containing the DG-2 construct in pGreenII along with pSoup helper plasmid was used for the infiltrated into the CGMMV infected *N. benthamiana* plant. Usually, DG2 was infiltrated at 7DPI of wild type CGMMV constructs (BP4/BP7). To predict the trans-replication of DG2 construct with the help of CGMMV helper virus, the RT-PCR based detection of the infiltrated leaf samples using GFP specific primers (BM-1265F, and BM-1266R) showed the presence of DG-2 within infiltrated tissue, whereas it was not detected from the only DG2 infiltrated tissue as well as only CGMMV infected tissue (**Fig. 4.28**a). It nullified the possibility of self-replication of DG-2, unless the support of wild type CGMMV. Further, to check the long-distance systemic movement, a similar RT-PCR was performed using both, infiltrated and systemic leaf samples; showed the presence of DG-2 within infiltrated tissue, but absent in the first systemic leaves (**Fig. 4.28**b), thus indicated the restricted systemic movement of DG-2 in *N. benthamiana*. Further, the experiment was repeated with both symptomatic as well as asymptomatic CGMMV; a similar result was obtained. Finally, to prove its trans-replication within infiltrated tissue, the gradual increase in the replication rate was visualized over time when tested at a 2days interval (**Fig. 4.28**c), keeping pCambia 1302 as a positive control. The trans-replication of DG-2 with its restricted systemic movement provided a new dimension in our research.

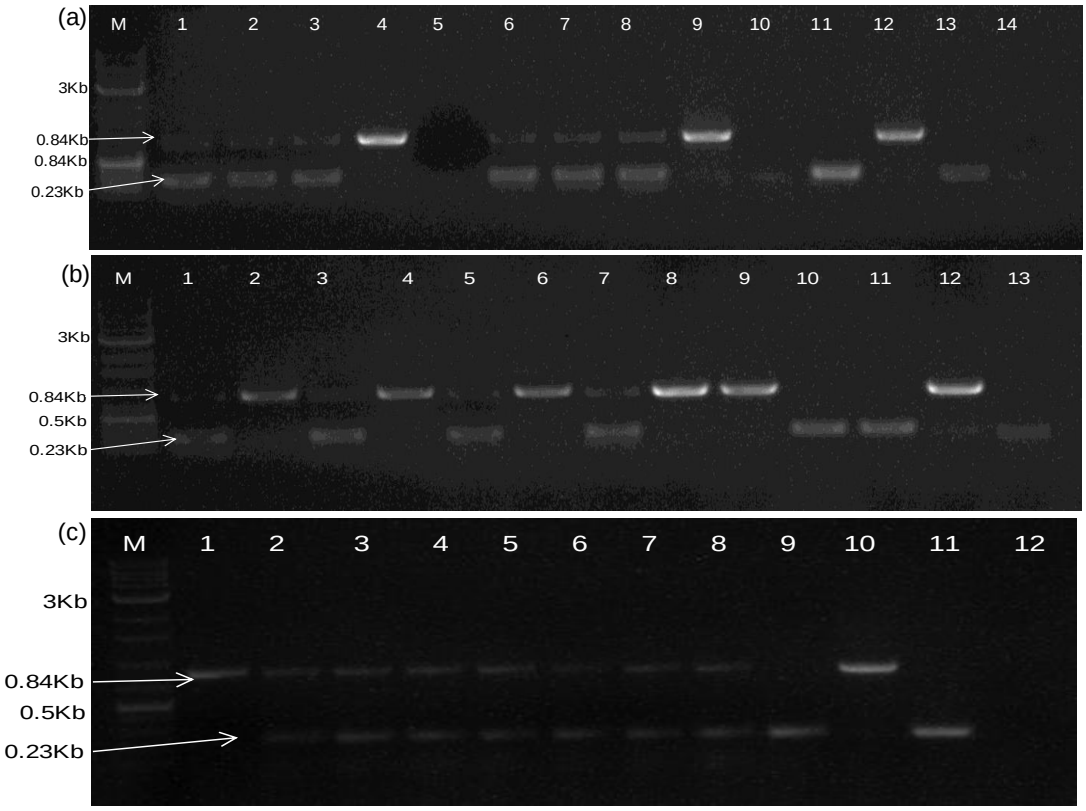


Fig. 4.28. Duplex RT-PCR based detection of deconstructed genome-2 within CGMMV infected *N. benthamiana* using two primer sets; BM-96F and BM-1268R are used to predict the infection of helper CGMMV, whereas, BM-1265F and BM-1266R are used to detect the DG-2 replicon. (a) The presence of DG-2 replicon within infiltrated tissue signifies its trans-replication with the help of CGMMV, as it was not detected from the only DG2 infiltrated tissue (in absence of CGMMV) as well as only CGMMV infected tissue (in absence of DG2). Here, Lane 1-3 is the symptomatic (BP4) CGMMV infected *N. benthamiana* plants agro-infiltrated with DG2, Lane 4 and Lane5 represents only BP4 and DG2 infiltrated leaf, respectively. Lane 6-8 is the asymptomatic (BP7) CGMMV infected *N. benthamiana* plants agro-infiltrated with DG2, Lane 9, and Lane10 represents only BP7 and DG2 infiltrated leaf tissue, respectively. Lane 11, 12, 13 depicts the DG2 plasmid control, BP4 plasmid control, and pCambia 1302 plasmid control with a negative control in lane 14. (b) To check the long-distance systemic movement of DG2, a similar RT-PCR was performed using RNA isolated from the infiltrated and systemic leaf samples; gel picture showed the presence of DG-2 within infiltrated tissue (lane-1,3, 5, 7), but absent in the systemic leaf (lane-2,4,6,8) as well as in the BP4 infected plant (Lane 9). Here, lane 10, 11, 12 is the pCambia 1302 plasmid control, DG2 plasmid control and BP4 plasmid control, respectively; (c) The gradual increment in the replication rate of DG2 was visualized over the time when tested at 0DPI (1), 2DPI (2), 4DPI (3), 6DPI (4), 8DPI (5), 10DPI (6), 12DPI (7), 15DPI (8), and compared with pCambia 1302 at 5th DPI (9). Here lane 10 and 11 are the BP4, and DG2 plasmid control with reagent control in lane 12.

Formulation of Standard curve for the absolute quantification

For the absolute quantification of DG-2 replicon, a standard curve was formulated against the standard template of known concentration. The template (230 nt long) was PCR-amplified using GFP specific primers (BM-1165F and BM-1166R) and further purified amplicon was used. The 70ng/µl stock was serially diluted in a 1:10 ratio and continued up to 10 times to get the lowest 7×10^{-9} ng//µl. The lowest dilution staring from 7×10^{-4} ng//µl to 7×10^{-9} ng//µl was used for qPCR reaction, selecting reaction as the standard. The result was displayed in the form of a straight line having slope (m)=-1.655, intercept(c)=9.09, with a correlation of coefficient (R^2)=0.99. This standard curve of known concentration was further used for the quantification of unknown samples.

Quantification of GFP transcript

To prove the trans-replication of DG-2 with the help of wild type CGMMV, qRT-PCR was performed to quantify the amount of DG-2 replicon in different temporal scales. The harvested tissues from the infiltrated area in the 2-15DPI period was used for the quantification using the GFP specific primers (BM-1165F and BM-1166R), keeping actin as the reference control. The average Ct value was varying from 19.40 to 18.46 within this time that was significantly higher than the respective reference actin gene (**Table 4.10.)**. The expression of GFP was calculated to be fold higher than the actin gene, and its concentration slowly increases within this prescribed time-course. It proved the higher accumulation of DG-2 within the infiltrated tissue, over host house-keeping genes. Further, to calculate the load of DG2 replicon, the Ct values were plotted in the previously formulated standard curve (Y=(-)1.655X+9.09, with R2=0.99), where Y is Ct value of the sample, and X is the log value virus titer. The original concentration of DG-2 replicon was calculated based on the antilog of X. The calculated load of DG-2 replicon was ranging from 0.067467 ng/g of leaf tissue in the 2nd DPI to 0.098899 ng/g of leaf tissue in the 15th DPI. Due to the restricted localization within specialized host tissue, the yield was comparatively lesser than the wild type CGMMV as per the previous observation in our laboratory (data not shown).

Table 4.10. Quantitative estimation DG-2 replicon in the infiltrated leaves of CGMMV infected *N. benthamiana*

Samples	Ct value	Log value (x) of Ct	Antilog value	Concentration (pg/µl)	Expected load of DG2 replicon (ng/g tissue)
2DPI	19.40	-6.26245	0.013026	0.134934	0.067467
4DPI	18.51	-5.67078	0.01963	0.195598	0.097799
6DPI	19.04	-6.08957	0.014684	0.15535	0.077675
8DPI	19.03	-6.08736	0.014707	0.156168	0.078084
10DPI	18.86	-5.80198	0.017924	0.167858	0.083929
12DPI	18.49	-5.70588	0.019158	0.194759	0.097379
15DPI	18.46	-5.57614	0.020961	0.197798	0.098899

*DPI indicates days post infiltration of DG2 at which the leaf samples were harvested. The threshold cycle (Ct) represents the average of three replicates. The log value of X was derived from the formula of the standard curve (Y= -1.655X+9.09, with R^2=0.99), where Y= Ct value. The original viral load was calculated from the antilog value of X with base value=2. The concentration of DG-2 RNA replicon was estimated in pg/µl, that was derived into the expected load (ng/g tissue).

4.2.2.2. Trans-replicational ability of DG2 constructs

Confocal microscopy

The CGMMV infected *N. benthamiana* plant agroinfiltrated with the DG2 construct was exhibited colour changes in the infiltrated tissue. When these infiltrated tissues were examined in Confocal microscopy, the expression of GFP was observed. The harvested tissue from the infiltrated area at different points of time showed the expression of GFP staring from 2DPI and extended up to 12 DPI or more (**Fig. 4. 30a**). But, the prominent expression of GFP was visualized at 6-8DPI, thereafter, gradually decline but remain visible up to 12DPI or more (**Table 4.11.)**. Interestingly, the expression of GFP within the leaf was scattered, and remain restricted within certain specific tissue (**Fig. 4.29**). It indicated that the DG-2 replicon was restricted within some specialized non-vascular-tissue of N. *benthamiana* leaf, and its inability to cross the barrier restricts them from long-distance movement.

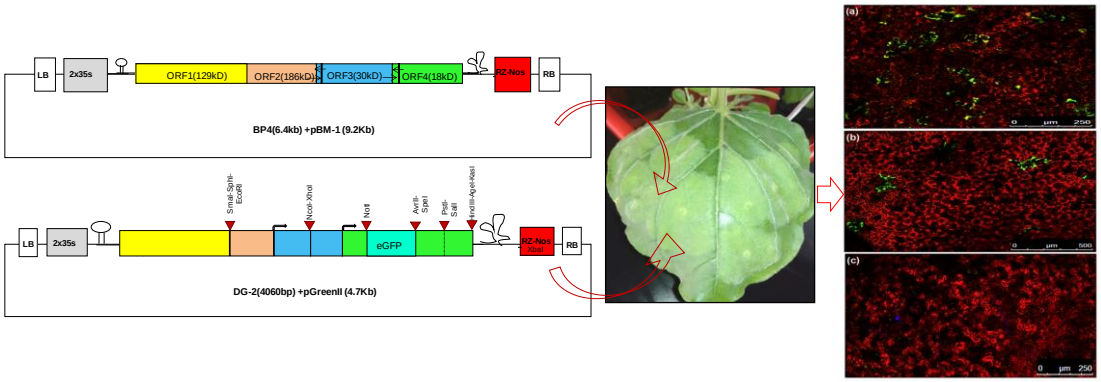


Fig 4.29. The restricted localization of DG-2 within the cellular system of leaf lamina, showing scattered expression of GFP in yellowish-green (a, b), whereas it was absent in the healthy tissue showing red colour (c). The expression was visualised in confocal laser microscopy. DG-2 construct was agro-infiltrated in the CGMMV infected *N. benthamiana*, after 7 days post infiltration of CGMMV infectious clone (BP4)

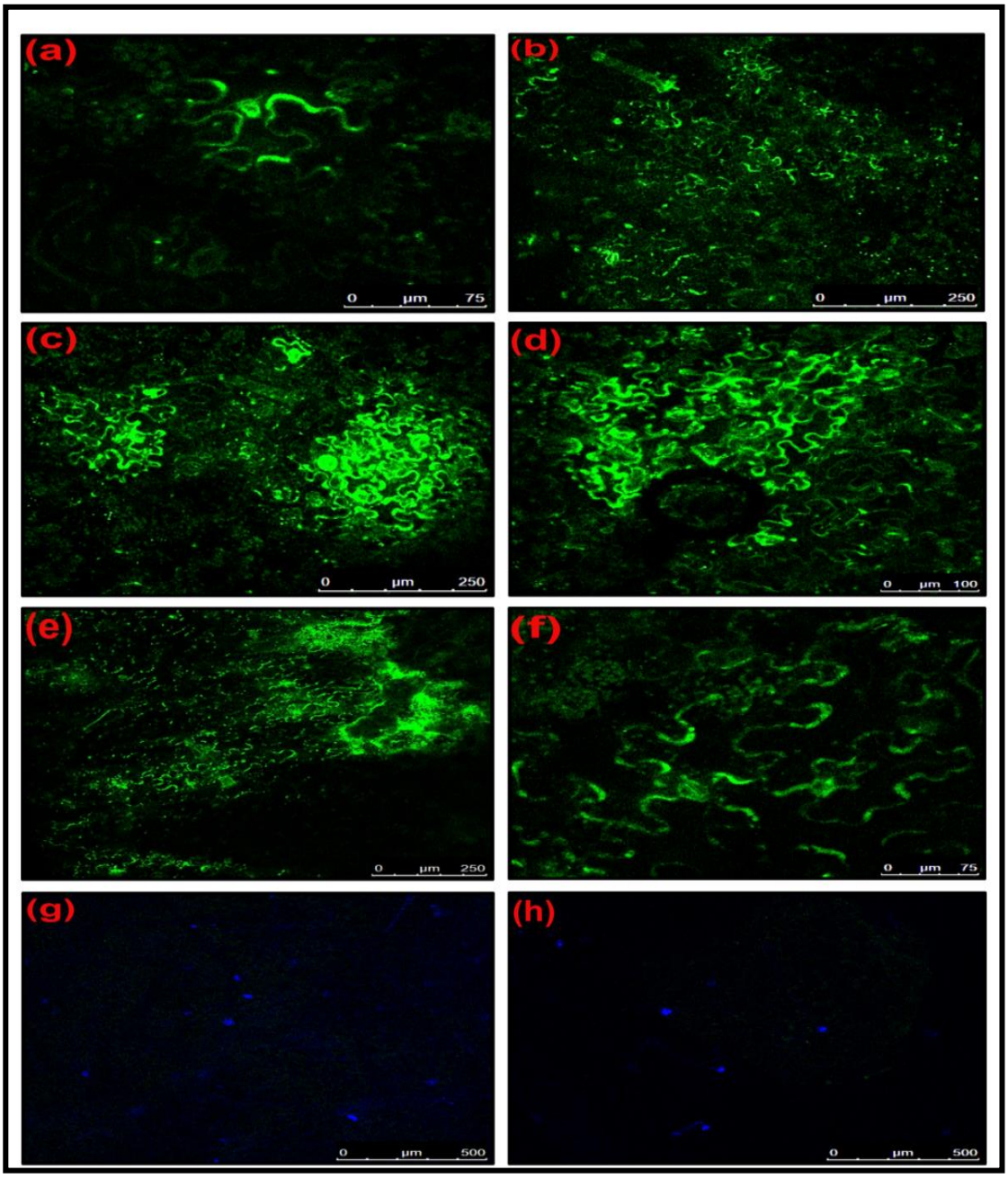


Fig. 4. 30a *In planta* cellular expression of GFP by DG-2 construct of CGMMV at different temporal scale when agro-infiltrated into the CGMMV infected *N. benthamiana* plant. Expression of GFP in 2 days post infiltration (DPI) of DG-2(a), 4DPI(b), 6DPI(c), 8DPI(d), 10DPI(e), 12DPI(f) was examined in Confocal microscopy. It indicates the critical time for higher expression of GPF as well as other foreign proteins by this construct lies within 6-10DPI span. Here, expression of GFP was not visualized in the only CGMMV infected plant (g) and only DG2 infiltrated plant tissue (h)

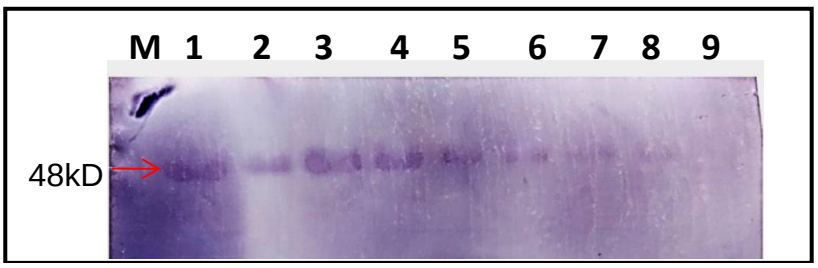


Fig 4.30b. The Western blot assay of the crude plant protein showing the expression of GFP and CP fused protein (48kDa) of CGMMV using a Polyclonal anti-GFP antibody. The change in the expression level was visualized on a different time–scales. The expression of GFP increases from 2DPI (1) to 6DPI(3) after that gradually decreases but remains in detectable level up to 15DPI(7). The GFP expression level can be compared with pCambia1302 (8). Here M indicates protein marker and 9 represents the crude protein from a healthy plant.

Western blot assay

The western blot assay of the extracted total crude protein also depicts the similar expression pattern of GFP. The *in planta* expression of GFP was detected using GFP polyclonal antibody; The GFP expression was confirmed starting from 2DPI and extended up to 15 DPI (**Table 4.11.)**. The highest expression of GFP was visualized during 6-8 DPI time (**Fig 4.30b)**; afterward, the GFP expression level declined slowly and remained up to the detectable level using a GFP-specific antibody. Here, the GFP expression using pCambia 1302 was used as the control. pCambia 1302 was infiltrated tissues and harvested at 5DPI, and also exhibit the expression of GFP. It was noteworthy to clarify the GPF expression level of DG2 was significantly higher than the pCambia 1302 plasmid construct, and also signified its translational ability to express the foreign protein.

Table 4.11. Trans-replication of DG2 and its GFP expression ability in the infiltrated leaves of CGMMV infected *N. benthamiana*

Treatment	Detection of DG-2		
	No of plant showing RT-PCR positive for DG2 construct/ Total no of plant tested (%)	**No of plant showing GFP expression/Total no of plant tested in Confocal microscopy (%)**	**Duration of GFP expression (DPI)**
BP4 + DG2	12/12(100)	6/12(50)	2-12
BP7 + DG2	12/12(100)	8/12(66.7)	2-15
BP4	0/6(0)	0/6(0)	-
BP7	0/6(0)	0/6(0)	-
DG2	0/6(0)	0/6(0)	-
Healthy Control	0/6(0)	0/6(0)	-

[1]The experiment was replicated twice. The DG-2 was agro- infiltrated into the infected *N. benthamiana* after 7DPI (days post infiltration) of the CGMMV symptomatic (BP4) and asymptomatic (BP7) clones. Healthy plants were infiltrated with only MES buffer solution. The RT-PCR was performed after 7DPI of DG-2 delivery. Confocal microscopy of the samples was performed at 2 days interval.

The valorization of plant viruses as the suitable expression vector for the in planta production of foreign proteins is becoming popular with the aim of low-cost manufacturing of vaccines and therapeutics. So far, a large number of plant viruses have been exploited for the construction of different types of viral vectors, Cucumber green mottle mosaic virus (CGMMV) is one of them. The mild symptomatology, higher replication rate, expression of the huge amount of coat protein (CP) and movement protein (MP), wide host range, ability to reach all plant parts, etc., make them a very good choice for the construction of virus genome-based vector. It was extensively exploited for the heterologous protein expression (Jailani *et al*., 2017) as well as the silencing of the host gene (Liu *et al*., 2020). Initially, the full-length genome was targeted for the insertion of the target genes either by fusing with the functional reading frame of movement and/or coat protein (insertion approach) or replacing the dispensable part of the viral gene (CP) (substitution approach). Sometimes, an additional gene promoter was also included for the efficient expression of foreign genes. Such, gene insertion and gene substitution vectors represent the first generation virus vectors based on their full-length genome.

Notably, the unstable and lower expression of the foreign protein was the biggest issue in gene insertion strategy due to the over-loading in the virus genome and their poor packaging. However, the expression of foreign protein is much higher in the substitution strategy, but unfortunately for a short window of time, with the restricted cell-to-cell and systemic movement of the virus also. Moreover, frequent events of homologous recombination force the reversion of wild type virus through of loss inserted foreign sequences (Willemsen and Zwart, 2019). Although many full-genome based virus vectors are available for the expression of foreign proteins, their instability in expression, limited cargo-capacity, lack of suitable restriction sites, difficulty in expressing multiple foreign genes, the problem in handling (>15Kb) encourage us to search for new alternatives.

To circumvent such hurdles, the development of the ‘deconstructed’ viral vector systems is found to be most successful (Gleba *et al*., 2014). In this strategy, the virus genome is redesigned in such a way that it will carry only the essential genomic element, necessary for the replication and translation, but other unwanted gene elements, required for the symptom development, virus packaging, movement, vector transmission, etc. are deleted. Previously, genome deconstruction was done for some monopartite RNA viruses, *viz*., potato virus X

(PVX) (Komarova *et al*., 2006), and TMV (Lindbo, 2007). The deconstructed virus genome acts as the dependent replicon that can efficiently replicate, transcribe, and translate with the help of wild type virus, once co-infect the plants; thus exploited as the suitable vector for production of plant biologics consistently at the sufficient level and express speed. The transient expression with deconstructed viral vectors allows the production of up to 5 milligrams of mAb per gram of fresh leaf weight within 2 weeks, in contrast to transgenic plants that require several months to years for generating the transgenics (Chen and Davis, 2016). Such a deconstructed genome of TMV was created by different scientists (Shivprasad *et al*., 1999), but its rational utilization for the higher expression of recombinant foreign proteins was directed by Marillonnet *et al.*, (2004). They designed a bipartite modular system in which the 5' module contains the *AtACT2* (actin) promoter with tobamovirus replicase and movement protein, and the 3' module contains the gene of interest (in place of CP gene), and *nos* terminator. This hybrid TMV-based deconstructed-vector module, magnifection (trademarked as magnICON®, Icon Genetics) is highly efficient in yielding 4-5 g/kg of fresh weight tissue (FWT) of recombinant protein (Gleba *et al*., 2005) and 4.8 g/kg FWT of full immunoglobulin G (IgG) in less than 2 weeks (Bendandi *et al*., 2010).

In similar to the magnICON, dRT-V, we are aiming for CGMMV also to make it an efficient vector system. This conversion of the virus from a pathogenic entity to a biological toolkit requires extensive genome modification that relies on the identification of cis-regulatory elements, subgenomic promotes, etc, and their subsequent engineering for efficient expression; so that a higher amount of protein can be expressed within the limited time. But, unfortunately, neither such information is available for CGMMV, nor the functionality of the deconstructed genome has been explored. Thus, the development of a deconstructed virus vector system based on the CGMMV genome is a real challenge. To achieve that, we have attempted to identify some key regulatory elements necessary for replication and translation *via in silico* approach and to create deconstructed genomes of CGMMV for their functionality assay. The results are organized and discussed in the following way.

5.1. *In silico* prediction of *cis*-regulatory elements within CGMMV genome

5.1.1. *In silico* prediction of *cis*-regulatory elements for genome replication

A virus genome is composed of some cis and trans-regulatory sequences. Of them, trans-regulatory sequences contain the part of viral replicase and movement protein which is

responsible for replication and cell-to-cell movement and that are not obligatory for virus replication; thus can be deleted in the deconstructed virus genome. Whereas, the cis-acting RNA elements are also located within viral genomes and play essential roles in most, if not all, of the fundamental viral infection cycles, including translation and replication (Chen *et al.*, 2017; Liu *et al*., 2009; Pathak *et al*., 2011; Newburn and White, 2015; Simon and Gehrke, 2009). It contains the primary sequences, secondary and tertiary structures, that are required for the binding of replicase, long-range RNA-RNA interactions, and they typically function by interacting with viral or host proteins. Therefore, identification of this cis-regulatory part is essential for its active replication of the deconstructed genome of the virus and efficient transient expressions of a foreign gene in plants (Klimyuk *et al.*, 2014; Gleba *et al*., 2007, 2014; Mortimer *et al.*, 2015).

Initially, whole genome-based searching of regulatory elements in the Rfam database depicts only 6289-6319 nt region of the 3'UTR of CGMMV as the canonical structured region containing pseudoknots. But 3'UTR of tobamovirus is reported to be folded into three consecutive pseudoknots followed by tRNA-like structures at the 3' proximal ends. Further, any information regarding the regulatory elements in the 5' terminal end was missing. To decipher such regulatory elements located in the 5' and 3' terminal part of the CGMMV genome, this study was carried out, and their potential interaction with the helicase and RdRp domain of replicase enzyme was also predicted through a computational approach to identify certain critical amino acid motif sequence interacting with specific nucleotides. For that, the comparative sequence analysis of the 5' terminal end (1-95nt) and 3' UTR (6249-6424nt) was carried out and various conserved regions along with many repeat sequences were identified. Further, the secondary structure of these parts was also elucidated and some conserved structural elements, other than canonical structure as reported in the Rfam database was identified. Like other tobamoviruses, 5'-untranslated region (UTR) of CGMMV contains the typical G-deficient omega ladder containing $(AAAC)_3$-$(AAC)_5$ rich repeats sequence, spanning between 10-55nt. This region is highly conserved across the 105 reported global isolates of CGMMV. A similar structure in 5'UTR of TMV is reported to have important regulatory functions (Gallie and Kado, 1989; Agalarov *et al.*, 2011). This region was identified as the binding site for some host proteins like heat-shock proteins (Wells *et al*., 1998), and elongation factors (eIF4F) (Gallie, 2002). Moreover, it is the site for the internal entry of small ribosomal subunits necessary for translational initiation (Mundry *et al.*, 1991). Although information regarding its interaction with viral replicase enzyme was missing, deletion analysis showed

that a major deletion of the sequences from 9-47 nt and 25-71 nt completely obliterate the replication of TMV, while the small deletions (~10 nt) deletions throughout this 5' leader diminish the virus replication up to a detectable level (Takamatsu *et al*., 1991). Thus, it is quite obvious that this 9-47 nt region contains the (CAAA)2-(CAA)5 repeat sequence that might be crucial for the binding of replicase. Later, Kawamura-Nagaya *et al*., (2014) proposed the possibility of binding of multiple replicase molecules to few CAA repeat units. In our study also, the binding of helicase and RdRp to the multiple nucleotides, specifically at the juncture of (AAAC) & (AAC) repeats support these previous findings.

Just downstream to the omega ladder, an SL-2 structure is predicted in 56-76nt of 5'terminal and found to be highly conserved across the isolates of CGMMV. Although it contains the AUG start codon for the synthesis of different components of the replicase enzyme, its functional significance in genome replication is yet to be established. Previously, the binding of TMV replicase enzyme to the 70-nucleotide region located at 3-94 nt of the 5'-proximal end was reported during cotranslational replication of genomic RNA (Kawamura-Nagaya *et al*., 2014). Thus, specific binding of the replicase enzyme at the SL2 might be necessary for cotranslational replication in tobamoviruses. In this way, these cis-acting elements in the 5′ terminal end are supposed to play a crucial role in the negative-strand RNA synthesis and regulation of translation in CGMMV.

Similar to the 5' terminal end, searching for regulatory elements in the 176 nt long 3'UTR of CGMMV is also performed and a complex structure carrying three upstream pseudoknots (located between 6266-6319 nt) followed a tRNA like structure (6320-6384) and another downstream pseudoknot (6385-6424) was predicted based on their secondary structure. Typically, ~200-nt long 3-UTR of tobamoviruses contains three consecutive pseudoknot structures (~70 nts long) followed by ~100-nt tRNA-like structures which is highly essential for replication and translation (Goelet *et al*., 1982; Ohno *et al*., 1984; Shirokikh *et al*., 2010; Agalarov *et al*., 2011). But substantial variation within this pseudoknot region of tobamoviruses is also reported. Generally, TMV and ToMV have prototypical three pseudoknots within the 5' region of the 200 nt long 3'UTR; but some tobamoviruses show variations in pseudoknot numbers (van Belkum *et al*., 1985) as nine upstream pseudoknots were observed in the 412 nts long 3'UTR of Odontoglossum ringspot virus (ORSV) (Gultyaev *et al*., 1994), and six upstream pseudoknots were predicted in the 357 nts long 3' UTR of *N. glauca* infecting U5 strain of Tobacco mild green mosaic virus (TMGMV) (Shivprasad *et al*.,

1999; Bodaghi *et al*., 2000). Despite such variation, these secondary structures provide the replication signals in a cis-preferential manner.

To further analyze the functional significance of pseudoknots and t-RNA like structure in CGMMV, molecular docking is performed and various critical nucleotide sequences were identified within these pseudoknot regions. Mostly, the helicase prefers the unwinded single strands of RNA, whereas the binding of RdRp is random but in a sequence-specific manner. The binding of RdRp to the $(UGG)_2$ repeat sequence of PK1 stem region indicates some functional significance of this sequence in interaction with replicase. Further strong binding of RdRp to the stem (GUGC) and terminal-loop (AUA) region of PK-2 reflects the structural and functional similarity to the clamped adenine motif (AUA terminal loop) localized in the stem-loop C (SLC) region of TLS in 3′-terminal regions of BMV RNAs (Kim *et al*., 2000; Kim and Kao, 2001). This SLC is supposed to be a key binding element for BMV replicase enzyme (Chapman and Kao, 1999; Sivakumaran *et al*., 2000). This finding is corroborated with our prediction also. Although based on deletion mapping, these two PKs located at the 5'-proximal end of TMV 3'UTR are reported to be dispensable for viral replication, but necessary to reach the replication rate to the wild-type level (Takamatsu *et al*., 1991). Probably, these pseudoknot structures, PK1 and PK2 are required for efficient binding of replicase enzyme and recognition of the tRNA-like structure at 3'-terminal end. Contrastingly, it was presumed that the PK3 at 3'-terminal region of TMV 3'UTR is dispensable for minus-strand synthesis during replication (Takamatsu *et al*., 1991). Based on the point mutation study, they presumed that a nucleotide stretch UAAAUCGAA in the terminal loop of PK3 in TMV might be recognized by the trans-acting factors. Strikingly, the structurally similar pseudoknot3 like double helix with a terminal hairpin loop composing the same nucleotide stretch (UAAAUCG) is predicted in CGMMV. The specific binding of helicase to this nucleotide stretch in the hairpin loop strongly supports the previous hypothesis of recognition by replicase enzyme. Furthermore, the high degree of conservation in the nucleotide stretch with other tobamoviruses (Leathers *et al*., 1993) enlightens the functional significance of this element.

Other than these three consecutive PKs, CGMMV bears a tRNA-like structure followed by a complex pseudo-knot structure with two double-helical stems that is analogous to the TMV 3′-UTR. According to our prediction, RdRp prefers the 5'-strand of the double helix stem of TLS arm resembling the t-RNA anticodon branch; whereas helicase showed a strong affinity for internal single-stranded RNA bulge. The sequence conserveness of both this internal bulge and arm is very limited, but its structural similarity with other tobamoviruses like TMV

indicates the essentiality of the structural configuration of this element rather than its sequence. Mutation and deletion analysis of this arm and internal bulge harrowing the structural conformation largely abolish the negative-strand synthesis (Osman *et al*., 2000). Thus, the practical significance of these structural elements interacting with the RdRp and helicase can be advocated.

In the complex PK 4, the two double helix stem (S1 & S2) were interacting strongly with helicase and RdRp. Both these stems are rich in G-C base pairs. Similar GC enriched two double-helix regions were predicted in TMV also (Felden *et al*., 1996). Later disruption of this G-C base-pairing was reported to reduce replication efficiency greatly (Osman *et al*., 2000), as it affects the structural conformation of the double helix structures. Further restoration of these structures through reverse mutation reinstates the replication rate as wild type (Osman *et al*., 2000). Thus functional relevance of this pseudoknot PK4 depends upon the structural stability of both the double helix stem (S1 & S2) of the pseudoknot. The high affinity of binding RdRp and helicase to these double helix stems indicates the possibilities of localizing RNA promoter in this region for negative-strand synthesis.

5.1.2. *In silico* prediction of *cis*-regulatory elements for genome translation

Identification of viral promoters for replication and subgenomic RNA production is one such step, as these promoters possess some key regulatory elements like enhancers, spacer sequences that are indispensable for virus replication, and protein translation. Hitherto, various *in vivo* and in vitro approaches were adopted for the identification and characterization of these elements; all of them are cost-intensive and time-consuming. Thus, alternative strategy-based bioinformatics tools become a choice for the robust prediction of these regulatory elements. These tools can accurately identify the conserved/census elements within a genome; thus, holds a great promise to bridge the gap (Nain *et al.*, 2011). Previously, different scientists have used the computational approach for genome-wide prediction and characterization of promoter elements in various plants (Kumari and Ware, 2013; Koramutla *et al*., 2016) as well as in viruses (Marks *et al*., 2006; Gupta and Ranjan, 2017). But they were restricted to the identification of DNA promoters and its associated elements. Here, we have adopted a similar approach to predict the RNA promoter elements for sgRNA production in CGMMV. Earlier, the borderline of CP-SGP was mapped for CGMMV (Liu *et al*., 2020), CFFMV (Rhee *et al*., 2016), and TMV (Grdzelishvili *et al.*, 2000), where multiple deletions of upstream and downstream sequences in respect to TSS was performed to delineate the SGP margins. The

same strategy is also followed for internal SGP mapping in the case of other RNA viruses, *viz*., BMV (Wierzchoslawski *et al*., 2006), CMV (Chen *et al*., 2000), etc. All such site-specific deletions are decided based on the presence of knots, pseudoknots, and stem-loops in the RNA secondary structures of these viruses. These structural elements and sequences in the promoter region are responsible for RdRp binding. Interesting, viral RdRp of BMV and TMV have the overall structural similarity with other polymerases, *viz*., DNA- dependent DNA polymerase (DdRp), DNA-dependent RNA polymerase (DdRp), and reverse transcriptase (RdDp) (O'Reilly and Kao, 1998) and composed of palm, thumb and finger sub-domains; the palm sub-domain contains the catalytic core which is embedded with five motifs(A-E) and the finger subdomain is housed with motif F and G (Venkataraman *et al.*, 2018) which are essential for binding to RNA and subsequent RNA polymerization. Like other polymerases, in the search for the promoter, the RNA polymerase preliminary binds to some cis-elements (structural shape and sequence), located in the template and non-template strand of the promoter, which helps them to unwind the strands to create an open promoter, followed by subsequent specific recognition of the template strand for the synthesis of RNA (Feklistov, 2013). But, the promoter recognition events of viral RNA polymerase (RdRp) are missing. Thus, it can be assumed that the viral RdRp, especially for positive-stranded RNA viruses requires the double helix form (dsRNA) containing both positive and negative strand of the genome, for the recognition of promoter to start the replication of a new positive-strand and the transcription of subgenomic RNA, rather than only on the newly synthesized negative strand. The *in vivo* colonization of TMV either in the replicative form (RF) or in replicative intermediates' (RI) during replication supports our hypothesis (Young and Zaitlin, 1986). The negative strands never remain free; and always engaged in the synthesis of multiple copies of either gRNAs or sgRNAs, in a time-dependent manner. The existence of free progeny viral genome and sgRNAs in huge copy number (Watanabe *et al*., 1987) indicates the same. Practically, a single viral polymerase of tobamoviruses can recognize different types of promoters necessary for genomic RNA replication and sgRNA production. Despite their sequence dissimilarity, few common structural features still exist (Haasnoot *et al.*, 2002; Olsthoorn *et al*., 2004), which acts as the signal for RdRp binding at 5' promoter and SGP. The promoter recognizing the ability of RdRp is exploited here to delineate the SGP elements. Understanding these regulatory elements/sequences is the key to portray the promoter for sgRNAs synthesis and their recognition by core RdRp domain of replicase enzyme is used as the yardstick for the SGP mapping.

Conventionally, tobamoviruses produce two sgRNAs, and the first MP-sgRNA promoter is overlapped in the 3' terminal of ORF2 encoding 183-kDa RdRp protein and 5' terminal of ORF4 encoding MP, whereas the second CP-sgRNA lies in between the 3' terminal of MP coding sequences (MP-CDS) and 5' terminal of CP-CDS (Watanabe *et al*., 1987; Grdzelishvili *et al*., 2000). Sometimes, these ORFs are overlapping (Dorokhov *et al*., 2017) or may be separated by different sizes of intron/spacer (Dorokhov *et al*., 2017). This diversification in the genome organization of tobamoviruses leads to the variation in subgenomic promoter length. Thus, based on sequence information, it is quite difficult to map their actual length. These SGPs are placed anywhere within the genome and enriched with various regulatory elements. Thus, mapping based on deletion mutagenesis of sequences around the transcription start site of sgRNA would not be enough for accurate prediction of promoter length. Thus, robust bioinformatics tools are used to support the deletion-based mapping process. In the present study, initially, the RNA secondary structure of the MP-SGP and CP-SGP region of six cucurbit infecting tobamoviruses was generated to analyze the identical stem-loops (SL) and other regulatory elements. The existence of trans-acting binding sites was also determined and various regulatory elements, *viz*., CAAT-box, TATA-box are identified and found to be distributed in a noncanonical manner. Nevertheless, some degree of similarities persists among the cucurbit infecting tobamovirus species; that is sufficient enough to decipher the key regulatory elements within SGP regions of CGMMV. Further, scanning of a strong binding affinity of RdRp throughout the SGP region helps to predict the actual promoter length.

According to our prediction, a zone spanning from -64 nt to +39nt in respect to MP-TSS (+1) is the core active promoter for the expression of MP protein; while the fully active promoter probably distributed within -136 nt to + 110 nt regions. Previously, the SGP region of MP was mapped in TMV and was identified at -95 to +40nt of MP-TSS with the -35 to +10nt as the core promoter (Grdzelishvili *et al*., 2000). Akin to TMV, the MP-SGP of CGMMV is composed of multiple SL and bulge-like structures which are essential for the RdRp binding as evident from our finding too. The site-directed deletion and substitution mutagenesis assay revealed the structural significance of SL1 for MP sgRNA synthesis in TMV, rather than its sequence (Grdzelishvili *et al*., 2000). Unfortunately, such information is missing for the cucurbit infecting tobamoviruses. Despite significant sequence dissimilarity in sgRNA promoters, the considerable structural homology persists among CGMMV, CMoV, and WGMMV. Our finding will be helpful to stipulate the SGP length of members of this subgroup.

On the contrary, we have predicted certain key regulatory elements within -114 nt to +144nt region of CP-SGP. These critical regulatory elements are necessary for the binding of RdRp and transcription of CP sgRNA. Our finding is corroborated with the recent report of Liu *et al.* (2020) showing -110 to +175nt relative to CP-TSS as the putative full promoter for CP-sg RNA production in CGMMV. In our *in silico* prediction, the critical length of CP-SGP is positioned very close to their wet lab observation. We visualized the strong binding affinity of RdRp at -60 to +21 nt zone, with respect to CP-TSS (+1). Interestingly, a strong affinity of RdRp binding was also evident at the SL structures located at 97-114 nt downstream zone, which signifies our previous report which showed notable expression of eGFP up to +105nt (Jailani *et al*., 2017). It indicates that key regulatory elements are residing within +114 nt zone with respect to CP-TSS.

5.2. Development of deconstructed genomes of CGMMV and their functional analysis in presence of wild type 'helper' CGMMV

5.2.1. Development of deconstructed genome-1 (DG-1) of CGMMV and its functional analysis

The production of defective genomes during the virus pathogenesis is a natural phenomenon. These are spontaneously generated during the replication of a virus, especially those containing RNA genome (Perrault, 1981) and when its concentration reaches high titer (López, 2014). It is hypothesized that the premature detachment of RdRp from its negative-strand template during replication (the breakpoint) and further reattachment to the same template leads to the synthesis of defective genomes (Lazzarini *et al.*, 1981; López, 2014; Bosma *et al*., 2019). Although they are defective for self-replication and pathogenesis as lacking the substantial parts of their parental viral genome, still they carry some key regulatory genomic elements that are indispensable for replication. Therefore, these truncated forms can only replicate with the assistance of the wild viruses, which acts as the helper by providing the replicase enzymes. They may either interfere with the replication of the parental virus (Huang and Baltimore, 1970) or not; thus referred to as the Defective Viral Genomes (DVGs) (Tapia *et al*., 2013; Rezelj *et al.*, 2018). They have a significant functional impact on the pathogenesis and evolution of the wild type virus. Naturally, their occurrence is reported during the late infection events of many human and animal viruses, viz., dengue virus (Aaskov *et al*., 2006), Ebola virus (Calain *et al*., 1999), hepatitis C virus (Noppornpanth *et al*., 2007), influenza A virus (Saira *et al.*, 2013; Vasilijevic *et al*., 2017), mumps virus (Santak *et al*., 2015), poliovirus (Mclaren and

Holland, 1974), respiratory syncytial virus (Treuhaft *et al*., 1982), Sendai virus (Strahle *et al.*, 2006), Sindbis virus (Fuller and Marcus, 1980), etc. Similar subviral agents are also evident (Vignuzzi and López, 2019) in case many plant viruses like *Bromovirus* (Pogany *et al.*, 1995), *Carmovirus* (Li *et al*., 1989), *Flexivirus* (Teycheney *et al*., 2005), *Tombusvirus* (Hillman *et al*., 1987), etc. that are synthesized *de novo* (Knorr *et al*., 1991). But such documentation of naturally occurring DVGs is lacking in the case of the majority of *Tobamoviruses* (Lewandowski and Dawson, 1998). It raises the question about the efficacy of CGMMV to support the trans-replication of subgenomic virus particles. Thus, the different artificial forms of dRNAs are generated from their genome to study their functional significance.

Previously, the numbers of DVGs are artificially generated from the full-length RNA genome of TMV through the deletion of internal genome sequences (Lewandowski and Dawson, 1998; Knapp *et al.*, 2001, 2005) that posses the cis-regulatory elements for replication and movements. Based on this information on TMV, we also try to develop a deconstructed genome named DG-1 derived from the CGMMV genome and its functionality was tested. DG-1 construct of CGMMV lost its ability of self-replication and in planta movement due to the large internal deletion of the genome containing the helicase, RNA dependent RNA polymerase domain of replicase enzyme and the significant part of movement protein. Thus, DG-1 replicon was undetected once infiltrated alone, neither in the infiltrated leaf nor in the systemic leaf of *N. benthamiana*, as previously reported by Ogawa *et al*., (1991) that large deletion of genes encoding methyltransferase, helicase, and RdRp domains of replicase generate replication-defective mutant in TMV. This DG-1 can only replicate and move systemically within *N. benthamiana* with the help of wild type CGMMV. Here, wild type CGMMV provides the replicase enzyme in *trans* to support the replication of DG-1. This was also the evident earlier experimentation of Raffo and Dawson (1991) in TMV; they reported that the artificially created subgenomic replicons, named KL that retain the 1-256nt of 5’ terminal and 5237-6395nt of 3' terminal ends, and KLL that contain only the 1-258 nt of 5’ terminal ends and 4900-6395nt of 3' terminal ends of TMV genome can replicate and spread systemically within tobacco plants, once co-inoculated with wild type TMV. It establishes our hypothesis that the wild type CGMMV can assist the trans-replication of subviral particles. But, the DG-1 replicon (dRNA) was not detected in the infiltrated leaves of *N. benthamiana* once co-infiltrated with wild type CGMMV (i.e., symptomatic BP4 and asymptomatic BP7); this is not at par with the previous finding of Raffo and Dawson (1991) and Knapp *et al.* (2005) in TMV where trans-replication

of dRNAs is reported when co-inoculated along with its parental TMV either in the tobacco protoplast or in the leaves of *N. benthamiana* plant. This difference might be correlated with the methods of delivery in the plant system. Previously, in vitro transcripts of dRNA genomes were used for the inoculation of protoplast; after 1-2day of the growth in protoplast, the cell lysate containing the active viral culture was directly inoculated into the plant leaves. Thus, the rapid multiplication of dRNAs along with its helper virus was evident. But, here, during co agro-infiltration of the cDNA construct of both DG-1 and 'helper' CGMMV, the wild type virus does not support the dRNA transcripts of DG-1 to amplify due to the lack of sufficient amount of replicase enzymes *in planta*. At the initial stage of infection by CGMMV, it produces replicase enzyme for the replication of its genome to establish its infection and colonization. It requires time to replicate its genome up to a critical level, at which it will support the replication of other dependent genomes the replication of dRNA immediately after infection. Identification of that critical time is necessary for achieving the sufficient amplification of DG-1. It may require further refinement in the delivery system also.

Practically, the naturally occurring dRNAs possess certain cis-acting elements that are necessary for efficient replication and movement with the help of the parental virus. This principle is used here to generate the DG-1 based on the previously reported information on TMV dRNAs. The DG-1 containing the 1-834nt of N- terminal part and 5271 to 6424 nt of the C-terminal part of the genome, is designed based on the genome map of deconstructed RNA (ΔHINC151) of TMV (Knapp *et al*., 2005) that retained 1-841nt (i.e., 258 aa) of the N- terminal and 5182-6395 nt of the C-terminal of its genome. The evidence of *in planta* trans-replication and systemic movement of DG-1 with help of wild type CGMMV is in accordance with the ΔHINC151 of TMV. Interestingly, ΔHINC151 of TMV is also reported to be accumulated in the inoculated leaves of *N. benthamiana*, about 4-5 folds higher than its helper TMV. But in our case, DG-1 is found to replicate at a maximum 4.5 fold over the actin; that is about 7.7 times lower than that of wild type CGMMV. After the cellular accumulation of DG-1 within infected tissue, DG-1 can spread systemically up to the 10th leaf or even more (as per our observation) in *N. benthamiana via* vascular bundles, which is not the evidence in the case of ΔHINC151 of TMV that was reported to move within the specialized non-vascular tissue; thus restrict their systemic movement (Knapp *et al.*, 2005). Whereas, the DG-1 replicon can systemically move long distances throughout the plant, but its uniform distribution in all local tissue is limited. It signifies that the 258aa of MT domain of CGMMV is necessary for long-distance movement of dRNA, but have a negative impact on its higher accumulation. This

finding is corroborated with the previous report of (Knapp *et al.*, 2001, 2005) in TMV, where they elucidated the functional significance of the N-terminal MT domain-containing (1-258aa) in the virus movement (Knapp *et al*., 2005).Further, this MT domain also has a crucial role in the formation of methylated cap that enhances the stability of viral genomic RNA. Interestingly, this 5' RNA capping mechanism is conserved among the diverse members of Alfavirus like superfamily; despite their limited animo acid identify (Meng and Lee, 2017). A comparative analysis of the MT domain of various plant infecting depicts some conserved amino acid residues, viz., histidine (H81) aspertate (D134), arginine (R137), tyrosine (Y271) located at its N-terminal part and form a catalytic tetrad that might have a significant role in the cap-formation. Disruption of any one of them greatly hamper the methyltransferase and AdoMet-dependent guanylyltransferase activity of this enzyme that affect the RNA capping (Kong *et al*., 1999); and results in the subsequent instability of the new synthesized RNA as proved in semliki forest virus (Ahola *et al*., 1997) and bamboo mosaic virus (Huang *et al*., 2004). The existence of a partial N-terminal domain in the DG-1 of CGMMV may also impair the RNA stability. Thus, despite its replication, the limited stability of DG-1 replicon affects it's *in planta* higher accumulation. Further, this DG-1 has other limitations, like a single gene can be expressed at a time under the control of the CP promoter. Thus, further redesigning of a new deconstructed genome was carried out.

The functional validation of DG-1 is proved through the silencing of phytoene desaturase (PDS) gene of *N. benthamiana*. Effective silencing of *NbPDS* signifies the replication and systemic movement of the DG-1 construct carrying the *NbPDS* gene. This is the first report of the silencing of the host gene by using CGMMV based binary system containing BP4 and DG(PDS)-1 replicon. Previously, CGMMV full-length genome-based VIGS vector was designed for the silencing of endogenous host gene in cucurbits like cucumber, melon, etc. (Liu *et al.*, 2020). Despite of their claim, their system did not work efficiently in *N. benthamiana.* At the same time, CGMMV full-length genome-based vector system was successfully established for the gene function assay in *N. benthamiana* by Jailani *et al*., (2020). Here, sense and antisense strategy were followed for gene silencing and observed the efficacy of silencing is higher when the target gene was inserted in the sense orientation. Based on this finding, we have adopted only the sense -strategy for the insertion of *NbPDS* gene and successfully proved the silencing of the target host gene using a deconstructed replicon. Similar, post-transcriptional gene silencing in *N. benthamiana* was also proved by

using a Grapevine virus A minireplicon (Brumin *et al*., 2009). This finding suggests the potentiality of the use of deconstructed replicon for plant functional genomic study.

5.2.2. Development of deconstructed genome-2 (DG-2) of CGMMV and its functional analysis

To avoid the shortfalls of DG-1, the DG-2 was created by incorporating 1173 nt (371 aa) of the MT domain with aim of achieving higher replication. Additionally, some multiple cloning sites were incorporated under the control of subgenomic promoters; so that multiple foreign genes can be expressed simultaneously. When the DG-2 construct was infiltrated into the CGMMV infected *N. benthamiana*, the DG-2 was trans-replicated with the help of wild type CGMMV. The RT-PCR based detection was used to prove, but it was not detected within the systemic leaf. This restricted movement of the deconstructed genome was previously reported in TMV (Knapp *et al*., 2001). They reported two categories of dRNAs, i.e., one that can't move were replicated by wild type virus, whereas others that can move but not replicated by its wild type. Here also, the RT-PCR based detection proved that the replication rate of DG-2 was gradually increased, but remain restricted within infiltrated tissue. This result indicated that the C'-terminal part of the MT domain (i.e., 258-371aa) was involved in the higher accumulation of dRNAs and it might be due to the proper capping mechanism. Further, it can be assumed that this region may be involved in the restricted movement of dRNAs, also. Similar evidence was reported by Knapp *et al*. (2005) and showed a higher accumulation but restricted movement of the deconstructed RNA genome, ΔCla151 that possesses the 425aa of MT domain. Previously, the role of a nonconserved region located between the MT domain and helicase domain in the cell to cell movement was reported and mutation of certain nucleotides of 248aa to 538aa region hampered this function in TMV (Hirashima and Watanabe, 2003). From the confocal microscopic analysis of DG-2 in the infiltrated tissue, it was clear that the localization of DG-2 was scattered within the lamina. Thus, it could be predicted that the cell-to-cell movement of DG-2 might be restricted due to the extensive modification of the MT domain.

Further, a qPCR assay of the DG-2 replicon indicated the progression in its replication rate. Thus, the load of DG-2 replicon was increased over time (within 15 DPI) from 0.067 to 0.098ng/g tissue. The quantification of the GFP expression within infiltrated tissue also signifies that trans-replication of DG-2 is assisted by the wild type CGMMV. Other than its trans-replication, DG-2 also has a better translational ability. To prove so, the GFP protein was inserted within the coding frame of CP and was expressed successfully within infiltrated tissue. The expression of GFP was visualized over time in confocal microscopy. It was observed that

the GFP expression was started from 2nd DPI and reaches at a high level in between 6th-10th DPI, thereafter gradually decreases; but remains expressed up to 15DPI. The finding is also supported by the western blot assay. The amount of GFP translation in the form of CP fused protein (48Kd) was predicted by using the Polyclonal anti-GFP antibody. The pattern of GFP expression was similar to confocal visualization. From these findings, it was clear that DG-2 can be used to express a foreign protein in plant tissues. Thus, DG-2 can be utilized for establishing a bipartite vector system using the CGMMV genome. A similar bipartite system based on TMV and its dRNA was developed by Knapp *et al.* (2005). Later, this system was explored as the novel two-component vector system (dRT-V) for the high-level expression of multiple therapeutic proteins including a human monoclonal antibody in plants (Roy *et al*., 2010). This new design based on the deconstructed genome of TMV was proved to be a successful alternative. In this system, a defective/deconstructed RNA (dRNA) of TMV was created by major deletion of replicase and movement protein and its functionality (trans-replication and movement) was predicted using a helper TMV construct. Interestingly, two foreign genes were inserted, one each in the helper construct and deleted construct, and used to express simultaneously in the same cell. This non-competitive bipartite module is named as defective RNA (dRNA)-based TMV vector (dRT-V). It is used to express mABs IgG against the antigen of *Bacillus anthracis* in *N. benthamiana* @ 120mg/kg FWT. In the same way, our bipartite system can also be exploited for the expression of multiple proteins in the plant.

Cucumber green mottle mosaic virus (CGMMV), a cucurbit infecting tobamovirus (family *Virgaviridae*) contains a monopartite, single-stranded, positive-sense RNA genome of ~6.4 kb. It produces very mild, mottle, mosaic symptoms on a large number of the cucurbitaceous crops. Despite that, it replicates at a very high rate and produced a huge quantity of proteins, especially coat protein. Thus, CGMMV becomes an attractive choice for the construction of a full-length genome-based virus-based vector system, so that, it can be utilized as an efficient toolkit for the heterologous expression of foreign proteins in plant. However, their limited cargo capacity, lack of suitable restriction sites, and difficulty in loading and expression of multiple heterologous proteins simultaneously limit their application. This can be overcome through the introduction of a shorter replicon along with the wild type CGMMV genome to design a bipartite expression system. To achieve this, primarily we have to elucidate the functionality of the artificially deconstructed shorter replicon of CGMMV; as any replicative sub-viral or defective viral genomes is neither present in the naturally infecting CGMMV, nor the functional significance of the artificially deconstructed genome has been analyzed so far. Our hypothesis is (1) the deconstruction of the CGMMV genome will lead to the development of shorter replicons that will replicate *in planta* with the help of the wild type CGMMV; (2) this short replicon can be used as the two component vector system that can be easily manipulated for loading a large and multiple foreign genes. In order to test the above hypothesis, a research programme was formulated with the title, "**Functional analysis of deconstructed genome of cucumber green mottle mosaic viru**s". Initially, *in silico* analysis of CGMMV genome sequence was performed to identify the *cis*-regulatory elements located at the 5' and 3' terminal ends and in the subgenomic promoter regions of the genome that are necessary for genome replication and translation. Based on the *cis*-regulatory elements, two deconstructed genomes (*viz.*, DG-1 & DG-2) of CGMMV were created. The functionalities of DG-1 and DG-2 were analysed by introducing them in the CGMMV infected *Nicotiana benthamiana* plant through agroinfiltration. The trans-replicability and systemic movement of DG-1 and its ability to silence foreign genes were testified by using various techniques like RT-PCR, qPCR, Northern blotting. Similarly, the translation ability of DG-2 replicon was visualized by using confocal microscopy, and further, the expression of foreign protein was verified through western blotting. The salient findings of this research was highlighted below:

1. The RNA structural analysis of the CGMMV genome helped identify numerous secondary elements distributed within the genome; these structures were the only indicators of the complex genome, but not enough for specifying their functional significance in relation to the replication and translation of the CGMMV genome.

2. To decipher their role, some conserved elements localized in the 5' and 3' terminal ends of the CGMMV genome were targeted for the binding of helicase and RNA dependent RNA polymerase (RdRp) domain of the replicase enzyme. Of them, three conserved pseudoknots (PK1, PK2, PK3) along with a complex of tRNA structure and PK4 were identified as the key elements for the binding of replicase enzyme for the synthesis negative strand. Further, omega lagger (SL1) and stem-loop (SL-2) were predicted to be the potential site for the binding of helicase and RNA dependent RNA polymerase (RdRp) domain at the 5' terminal of the CGMMV genome that were necessary for the synthesis of genomic (+) stand using negative-strand as the template.

3. Our computational analysis of the subgenomic promoter region of ORF3 and ORF4 encoding the movement protein (MP) and coat protein (CP), respectively, predicted some crucial structural elements that were requisite for the binding of RdRp. The binding of RdRp with these elements indicated that the core active promoter of CP was located within -60 nucleotides (nt) to +114 nt region, whereas, the full active promoter was from -114 nt to +144 nt with respect to the translation start site (TLSS). Similarly, for MP protein expression, the core promoter was predicted within -64 nt to + 39 nt regions relative to TLSS (+1), and further the full promoter was distributed within -36 nt in the upstream and +110 nt in the downstream with respect to TLSS.

4. Based on the *in silico* predictions of *cis*-regulatory elements, two deconstructed genomes were created. The deconstruction genome of CGMMV was designed based on the identification and retention of the minimal *cis*-regulatory elements which are compulsory for the efficient replication and translation of the deconstructed genomes in trans-complementation with its wild type genome.

5. The deconstructed genome-1 (DG-1) was designed based on the deconstructed RNA based vector of TMV. It was composed of 834 nt (1-834nt) in the 5' terminal end containing the 258 amino acids (aa) of methyltransferase (MT) domain along with 5'UTR and 1154 nt (5271-6424 nt) in the 3' terminal end of the CGMMV genome, containing CP subgenomic promoter and CP coding frame along with 3'UTR.

6. The DG-1 (~2.0 Kb) was developed by deleting the 835 nt-5270 nt from the CGMMV genome carrying the helicase, RdRp domain as well as MP coding sequences. Overlapping extension PCR technique was employed for the deletion of the large genome segment from BP4, the infectious cDNA clone of CGMMV.

7. Another deconstructed genome (DG-2) was designed by incorporating three separate genomic segments of the CGMMV genome. It was composed of 1173 nt (1-1173 nt) of the 5' terminal end representing the 371 aa of MT domain along with 5'UTR, 181 nt (4900-5080 nt) of the putative MP-SGP region, and 813 nt (5612 -6424 nt) of the 3' terminal of the genome containing CP coding frame and CP-subgenomic promoter (SGP) and the 3' UTR. The internal fusion of multiple restriction sites was introduced for facilitating the insertion of multiple foreign genes and further genome modification.

8. The trans-replication of DG-1 differed significantly with reference to the time of infection of wild type CGMMV. DG-1 could not replicate when agroinfiltrated together with BP-4, the wild type CGMMV in *N. benthamiana* or cucumber. The wild type CGMMV may not support the replication of DG-1 immediately after infection as probably, a critical level of replication of the wild type CGMMV is required to support the dependent genomic component, DG-1. Interestingly, DG-1 could successfully replicate when infiltrated in the preinfected *N. benthamiana* with CGMMV.

9. RT-PCR based detection assay was optimized to detect DG-1 in plant in the presence of the wild type CGMMV. DG-1 replicon was detected within the infiltrated tissue as well as in the systemic leaves of *N. benthamiana* by RT-PCR when DG-1 was infiltrated in *N. benthamiana* plants, which were inoculated with the wild type CGMMV 7 days ago. Long-distance systemic movement of DG-1 and its distributionin to all the leaves of the CGMMV infected plant indicated a sustainable co-existence of DG-1 with the wild type CGMMV in the plant system. This results suggest the DG-1 + CGMMV is a potential bipartite vector system.

10. Furthermore, *in planta* functional analysis of the DG-1 was performed to validate the newly developed bipartite vector system. The DG-1 was successfully deployed for the silencing of the host gene. For this, a construct was developed by incorporating a segment of PDS gene isolated from *N. benthamiana* (*NbPDS*) in the DG-1. DG-1 containing PDS induced photo-bleaching of leaves of *N. benthamiana* infected with CGMMV and this was further confirmed by RT-PCR and qRT-PCR and Northern blotting. The results suggested that DG-1 + CGMMV as a virus induced gene silencing vector.

11. Another deconstructed genome, DG-2 was synthesised and then a construct was prepared by placing it in a binary vector containing double 35S promoter and a NOS terminator with the aim of achieving a higher replication rate. The DG-2 replicated at a much higher rate but remained restricted within the infiltrated tissues. Its trans-replication and restricted movement within infiltrated tissue were confirmed by RT-PCR. Further, the qPCR analysis showed an increasing load of DG-2 over time from 0.067 to 0.098 ng/g tissue.

12. The translational ability of DG-2 was proved through the expression of GFP within *N. benthamiana.* The expression of GFP from 2-15DPI time-span was visualized through confocal microscopy. The level of expression of GFP was higher within 6-10 DPI that was further confirmed by western blotting. The *in planta* expression of GFP through DG-2 in assistance with wild type CGMMV indicated the suitability of the bipartite vector system for the foreign protein expression in plant.

In conclusion, two novel bi-component CGMMV-based vector systems have been developed in this study. In the first system, the DG-1 replicon was capable of replicating *in trans* with the assistance of its helper CGMMV. Further, it could move from cell to cell and then through the vascular system to reac the upper leaves of *N. benthamiana*. The results showed that DG-1 along with helper CGMMV represented a new bipartite genome system that was functionally similar to the bipartite plant viruses. As the DG-1 system could silence *N. benthamiana* PDS gene resulting the photo bleaching phenotype, it was useful as a virus induced gene silencing vector. In the second system, DG-2 construct could replicate at a much higher rate, but remains localized within the infiltrated tissues. Is translational ability of foreign protein was proved through the expression of GFP in *N. benthamiana.* The study resulted in successful development of novel bipartite vector systems using CGMMV genome that could be utilised as an *in planta* gene silencing as well as expression vector system.

Abstract (In English)

Functional analysis of deconstructed genome of cucumber green mottle mosaic virus

Cucumber green mottle mosaic virus (CGMMV), a cucurbit infecting tobamovirus (family *Virgaviridae*) contains a monopartite, single-stranded, positive-sense RNA genome of 6.4 kb. CGMMV is an attractive choice for the construction of a full-length genome-based virus-based vector system, so that, it can be utilized as an efficient biological toolkits for the plant functional genomics study as well as for the heterologous expression of foreign proteins in the plant. However, their limited cargo capacity, lack of suitable restriction sites, unstable expression of foreign genes, and difficulty in the simultaneous expression of multiple heterologous proteins limits their application. To circumvent such issues, it was attempted to create a bipartite expression system where a short genomic component was designed in a manner that it was dependent on wild type CGMMV for its replication, but was able to translate proteins from its backbone. The shorter genome component is expected to provide ease of handling while insertion of new or multiple foreign genes. Further, the shorter genome containing the foreign gene can be directly inoculated to a CGMMV infected plant for the expression of foreign proteins. In this study, *in silico* prediction of putative *cis*-regulatory elements were carried out to identify the key regulatory elements indispensable for CGMMV genome replication and translation. This information was utilized to develop two artificially deconstructed genomes of CGMMV and their *in planta* functionality assay was performed. Initially, the deconstructed genome-1 (DG-1) was created by a major deletion of the internal sequences, i.e., 835 to 5270 nucleotides (nt) including the helicase and RNA dependent RNA polymerase domain of replicase enzyme and part of movement protein from the full-length CGMMV cDNA infectious clone using overlapping extension PCR technique. Due to the loss of a significant part of replicase enzyme, this DG-1 was unable to self-replicate, but able to trans-replicate with the help of wild type CGMMV. Both symptomatic and asymptomatic infectious clone of CGMMV was used as the helper virus. A PCR assay was optimized to detect the DG-1 simultaneously with wild type helper CGMMV. The agro-infiltration of DG-1 in the CGMMV infected *N. benthamiana* resulted in the trans-replication and systemic movement of DG-1replicon. After the trans-replication within infiltrated tissue, it reached in the first systemic leaves at 2-3 DPI, and in second systemic leaf at 3 DPI; subsequently spread to other systemic leaves of *N. benthamiana* as observed by testing with RT-PCR using the plant leaf samples of

different time and different position of leaves. Further to validate its functionality, the partial mRNA sequence (~227 nt long) of the phyton desaturase (*NbPDS*) gene of *N. benthamiana* was inserted within DG-1 construct in sense orientation and when delivered into the plant, it resulted in the photobleaching of the leaves. The presence of the DG-1 construct in the PDS silenced plant was also confirmed by RT-PCR, qRT-PCR, and Northern blotting. The result confirmed the ability of DG-1to silence a host gene. Therefore, DG-1 and CGMMV together formed a 'bipartite system' of virus-induced gene silencing (VIGS) vector. Additionally, another deconstructed genome (DG-2) was developed based on *in silico* analysis of the CGMMV genome. Based on computational analysis, the subgenomic promoter (SGP) for the movement protein and coat protein were identified within -136 nt to +110 nt and the -114 nt to +144 nt region, with respect to MP-translation start site (TLSS) and CP-TLSS (+1), respectively; and the binding ability of RNA dependent RNA polymerase with the regulatory elements of the SGP region was used as the yardstick. The DG-2 was constructed by affixing 1173nt from 5' terminal (1-1173nt), 181 nt from MP-SGP region (4900-5080nt), and 813 nt from 3' terminal (5612-6424nt) sequence containing CP-SGP, CP-ORF, and 3'UTR. The DG-2 was 2.1kb in length where multiple cloning sites were incorporated in five different locations for convenient editing and loading of foreign genes. The design of DG-2 was conceived as an improved version of DG-1. When DG-2 was agro-infiltrated within CGMMV infected *N. benthamiana*, the DG-2 trans-replicated with the help of wild type CGMMV. Both, the symptomatic and asymptomatic CGMMV were equally efficient to support the trans-replication of DG-2. Although it was trans-replicated within infiltrated tissue and load of DG-2 replicon increased from 0.067 to 0.098ng/g tissue over the time of 2-15DPI, DG-2 remained localized within the infiltrated leaves only. Its systemic movement was restricted. Its trans-replication and restricted movement within infiltrated tissue were also confirmed by RT-PCR and confocal microscopy. Furthermore, its translational ability was also validated through the expression of GFP within *N. benthamiana*. The GFP was expressed throughout 2-15 DPI, but the level of expression was visualized higher between 6-10DPI, which was later validated with western blotting. The present study resulted in the first successful demonstration of a bipartite vector system based on the CGMMV genome that could be utilized for the silencing of plant genes as well as the expression of foreign proteins in *Nicotiana benthamiana.*

खीरा हरा मटमैला मोज़ेक वायरस विनिर्माणित जीनोम का कार्यात्मक विश्लेषण

सार

खीरा हरा मटमैला मोज़ेक वायरस (सीजीएमएमवी), खीरा परिवार को संक्रमित करने वाला एक टोबामोवायरस (परिवार वीरगविरिडे) है, जिसका आनुवांशिक पदार्थ 6.4 केबी का एक मोनोपार्टाइट, एकल-फंसे हुए, सकारात्मक-भावना वाले आरएनए का है। सीजीएमएमवी एक पूर्ण-लंबाई जीनोम-आधारित वायरस-आधारित वेक्टर प्रणाली के निर्माण के लिए एक आकर्षक विकल्प है, ताकि, इसका उपयोग पौधे के कार्यात्मक जीनोमिक्स अध्ययन के साथ-साथ विदेशी प्रोटीन की विषम अभिव्यक्ति के लिए एक कुशल जैविक उपकरण-किट के रूप में किया जा सके। हालांकि, उनकी सीमित मालवाहक क्षमता, उपयुक्त प्रतिबंध साइटों की कमी, विदेशी जीन की अस्थिर अभिव्यक्ति और कई विषम प्रोटीनों की एक साथ अभिव्यक्ति में कठिनाई उनके आवेदन को सीमित करती है। इस तरह के मुद्दों को दरकिनार करने के लिए, यह द्विदलीय अभिव्यक्ति प्रणाली बनाने का प्रयास किया गया था, जहां एक छोटे जीनोमिक घटक को इस तरह से डिजाइन किया गया था कि यह अपनी प्रतिकृति के लिए जंगली प्रकार सीजीएमएमवी पर निर्भर था, लेकिन इसकी रीढ़ से प्रोटीन का अनुवाद करने में सक्षम था। छोटे जीनोम घटक से नए या कई विदेशी जीनों को सम्मिलित करते समय में आसानी प्रदान करने की उम्मीद की जाती है। इसके अलावा, विदेशी जीन वाले छोटे जीनोम को विदेशी प्रोटीन की अभिव्यक्ति के लिए सीधे सीजीएमएमवी संक्रमित पौधे में डाला जा सकता है। इस अध्ययन में, सीजीएमएमवी जीनोम प्रतिकृति और अनुवाद के लिए अपरिहार्य रूप से प्रमुख नियामक तत्वों की पहचान करने के लिए पुटीय सिस-विनियामक तत्वों की सिलिको भविष्यवाणी में किया गया था। यह जानकारी सीजीएमएमवी के दो कृत्रिम रूप से विघटित जीनोम विकसित करने के लिए उपयोग की गई थी और पादप तंत्र कार्यक्षमता में उनका प्रदर्शन किया गया था। प्रारंभ में, विघटित जीनोम -1 (डीजी -1) आंतरिक अनुक्रमों के एक प्रमुख विलोपन द्वारा निर्मित किया गया था, अर्थात, हेलिकेज़ और आरएनए पर निर्भर आरएनए पोलीमरेज़ डोमेन के प्रतिकृत एंजाइम और आंदोलन प्रोटीन के हिस्से सहित आरएनए पोलीमरेज़ डोमेन सहित 835 से 5270 न्यूक्लियोटाइड्स (एनटी) जोकि अतिव्यापी विस्तार पीसीआर तकनीक का उपयोग करके पूर्ण लंबाई सीजीएमएमवी सीडीएनए संक्रामक क्लोन सेलिय गया था। प्रतिकृति एंजाइम का एक महत्वपूर्ण हिस्सा खो जाने के कारण, यह डीजी-1 स्वयं-प्रतिकृति करने में असमर्थ था, लेकिन जंगली प्रकार सीजीएमएमवी की मदद से ट्रांस-प्रतिकृति करने में सक्षम था। सीजीएमएमवी के रोगसूचक और स्पर्शोन्मुख संक्रामक क्लोन दोनों को हेल्पर वायरस के रूप में इस्तेमाल किया गया था। एक पीसीआर परख जंगली प्रकार सहायक सीजीएमएमवी के साथ एक साथ डीजी -1 का पता लगाने के लिए अनुकूलित किया गया था। सीजीएमएमवी संक्रमित *एन बेंथामियाना* में डीजी-1 के एग्रो-घुसपैठ के परिणामस्वरूप डीजी-1 प्रति के ट्रांस-प्रतिकृति और प्रणालीगत गति का परिणाम हुआ। घुसपैठ किए गए ऊतक के भीतर ट्रांस-प्रतिकृति के बाद, यह 2-3 डीपीआई में पहली प्रणालीगत पत्तियों में पहुंचा, और 3 डीपीआई में दूसरी प्रणालीगत पत्ती में; बाद में *एन बेंथामियाना* के अन्य प्रणालीगत पत्तियों में फैल गया जैसा कि विभिन्न समय और पत्तियों की अलग-अलग स्थिति के पौधे की पत्ती के नमूनों का उपयोग करके आरटी-पीसीआर के साथ परीक्षण करके देखा गया। इसकी

कार्यक्षमता को मान्य करने के लिए, *एन. बेंथामियाना* के फाइटोन डिसचूरेट (Nbपीडीएस) जीन का आंशिक mRNA अनुक्रम (~227 nt लंबा) डीजी-1 निर्माण के भीतर मे भावना अभिविन्यास में डाला गया और जब पौधे में वितरित किया गया, तो इसका परिणाम पत्तियां के फोटोब्लैचिंग के रूप में सामने आया। पीडीएस-खामोशीत पौधे मे डीजी-1 निर्माण की उपस्थिति की पुष्टि RT-PCR, qRT-PCR और उत्तरी सोख्ता द्वारा भी की गई थी। परिणाम ने डीजी-1 के द्वारा मेजबान जीन को मौन एक की क्षमता की पुष्टि की। इसलिए, डीजी-1 और सीजीएमएमवी ने मिलकर वायरस-प्रेरित जीन साइलेंसिंग (VIGS) सदिश का एक 'द्विदलीय प्रणाली' बनाया। इसके अतिरिक्त, सीजीएमएमवी जीनोम के सिलिको विश्लेषण के आधार पर एक और विघटित जीनोम (डीजी-2) विकसित किया गया था। कम्प्यूटेशनल विश्लेषण के आधार पर, गति प्रोटीन और कोट प्रोटीन के लिए उप-जीनोमिक प्रमोटर (एसजीपी) की पहचान -136 nt से +110 nt और -114nt से +144 nt क्षेत्र तक, क्रमशः; एमपी-ट्रांसलेशन स्टार्ट साइट और सीपी-टीएलएसएस (+1) (TLSS) के संबंध में की गई थी, और एसजीपी क्षेत्र के नियामक तत्वों के साथ आरएनए पर निर्भर आरएनए पोलीमरेज़ की बाध्यकारी क्षमता का उपयोग मापदंड के रूप में किया गया था। डीजी-2 का निर्माण 5' टर्मिनल से 1173nt (1-1173nt), MP-SGP क्षेत्र से 180 nt (4900-5080nt) और 3' टर्मिनल से 812nt (5612-6424nt) सीपीपी-एसजीपी युक्त अनुक्रम से 812 nt का निर्माण किया गया था जिसमे सीपी-ओआरएफ, और 3'यूटीआर उपस्थित थे। डीजी-2 की लंबाई 2.1kb थी जहां विदेशी जीनों के सुविधाजनक संपादन और लोडिंग के लिए पांच अलग-अलग स्थानों में कई क्लोनिंग साइटों को शामिल किया गया था। डीजी-2 के डिजाइन की कल्पना डीजी-1 के बेहतर संस्करण के रूप में की गई थी। जब डीजी-2 को सीजीएमएमवी संक्रमित *एन. बेंथामियाना* के भीतर घुसपैठ किया गया था, तो डीजी-2 ने जंगली प्रकार सीजीएमएमवी की मदद से ट्रांस-प्रतिकृति की। दोनों, रोगसूचक और स्पर्शोन्मुख सीजीएमएमवी डीजी-2 की ट्रांस-प्रतिकृति का समर्थन करने के लिए समान रूप से कुशल थे। हालाँकि यह घुसपैठ किए गए ऊतक के भीतर ट्रांस-प्रतिकृति था और 2-15 DPI के समय में डीजी-2 प्रतिकृति का भार 0.067 से बढ़कर 0.098ng / g ऊतक हो गया, डीजी-2 केवल घुसपैठ वाली पत्तियों के भीतर स्थानीयकृत रहा। इसका प्रणालीगत गति प्रतिबंधित था। घुसपैठ ऊतक के भीतर इसकी ट्रांस-प्रतिकृति और प्रतिबंधित गति भी आरटी-पीसीआर और कोनफोकल सूक्ष्मदर्शी द्वारा पुष्टि की गई थी। इसके अलावा, इसकी अनुवाद क्षमता भी *एन. बेंथामियाना* के भीतर जीएफ़पी की अभिव्यक्ति के माध्यम से मान्य की गई थी। जीएफ़पी लगातार 2-15 DPI के दौरान अभिव्यक्तित पाया गया, लेकिन अभिव्यक्ति का स्तर 6-10DPI के बीच उच्चतर पायी गई थी, जिसे बाद में वेस्टर्न ब्लोटिंग के साथ मान्य किया गया था। वर्तमान अध्ययन के परिणामस्वरूप सीजीएमएमवी जीनोम पर आधारित एक द्विदलीय वेक्टर प्रणाली का पहला सफल प्रदर्शन किया गया, जिसका उपयोग पौधों के जीनों के साथ-साथ *एन. बेंथामियाना* में विदेशी प्रोटीन की अभिव्यक्ति के लिए किया जा सकता है।

Aaskov, J., Buzacott, K., Thu, H. M., Lowry, K. and Holmes, E. C. (2006). Long-term transmission of defective RNA viruses in humans and Aedes mosquitoes. *Science*. 311: 236-238.

Agalarov, S. C., Sogorin, E. A., Shirokikh, N. E. and Spirin, A. S. (2011). Insight into the structural organization of the omega leader of TMV RNA: The role of various regions of the sequence in the formation of a compact structure of the omega RNA. *Biochemical and Biophysical Reserach Communications*. 404: 250-253.

Ahlquist, P., Bujarski, J. J., Kaesberg, P. and Hall, T. C. (1984). Localization of the replicase recognition site within brome mosaic virus by hybrid-arrested RNA synthesis. *Plant Molecular Biology*. 3: 37-44.

Ahola, T., Laakkonen, P., Vihinen, H. and Kaariainen L. (1997). Critical residues of Semliki Forest virus RNA capping enzyme involved in methyltransferase and guanylyltransferase-like activities. *Journal of Virology*. 71: 392-397.

Ainsworth, G. C. (1935). Mosaic disease of cucumber. *Annals of Applied Biology*. 22: 55-67.

Ali, M. E., Waliullah, S. and Nishiguchi, M. (2016). Molecular analysis of attenuated strain of *Cucumber green mottle mosaic virus* using in vitro infectious cDNA clone: pathogenicity and suppression of RNA silencing. *Journal of Plant Biochemistry and Biotechnology*. 25: 79-86.

Andrews, R. J., Roche, J. and Moss, W. N. (2018). ScanFold: an approach for genome-wide discovery of local RNA structural elements-applications to Zika virus and HIV. *Peer J*. 6: e6136. Doi:10.7717/peerj.6136.

Antignus, Y., Wang, Y., Pearlsman, M., Lachman, O., Lavi, N. and Gal-On, A. (2001). Biological and molecular characterization of a new cucurbit-infecting tobamovirus. *Phytopathology*. 91: 565-571.

Bachan, S. and Dinesh-Kumar, S. P. (2012). Tobacco rattle virus (TRV)-based virus-induced gene silencing. *Methods in Molecular Biology*. 894: 83-92.

Balogun, O. S., Xu, L., Teraoka, T. and Hosokawa, D. (2002). Effects of single and double infections with Potato virus X and Tobacco mosaic on disease development, plant growth and virus accumulation in tomato. *Fitopatologia Brasileira*. 27: 241-248.

Baumstark, T. **and** Ahlquist, P. (2001). The brome mosaic virus RNA3 intergenic replication enhancer folds to mimic a tRNA TpsiC-stem loop and is modified *in vivo*. *RNA (New York, N.Y.).* 7: 1652-1670.

Bendandi, M., Marillonnet, S., Kandzia, R., Thieme, F., Nickstadt, A., Herz, S., Frode, R., Inoges, S., Lopez-Diaz De Cerio, A., Soria, E., Villanueva, H., Vancanneyt, G., McCormick, A., Tuse, D., Lenz, J., Butler-Ransohoff, J. E., Klimyuk, V. and Gleba, Y. (2010). Rapid, high-yield production in plants of individualized idiotype vaccines for non-Hodgkin's lymphoma. *Annals of Oncology*. 21: 2420-2427.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*. 28: 235-242.

Biesiada, M., Purzycka, K. J., Szachniuk, M., Blazewicz, J. and Adamiak, R. W. (2016). Automated RNA 3D Structure Prediction with RNAComposer. In: RNA Structure Determination. Methods in Molecular Biology. D. Turner and D. Mathews, (eds). Humana Press, New York, NY. pp 199-215.

Birnboim, H. C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*. 7: 1513-1523.

Boccard, F. and Baulcombe, D. (1993). Mutational analysis of cw-acting sequences and gene function in RNA3 of cucumber mosaic virus. *Virology*. 193: 563-578.

Bodaghi, S., Ngaon, A. Y. M. and Dodds, J. A. (2000). Heterogeneity in the 39-terminal untranslated region of tobacco mild green mosaic tobamoviruses from Nicotiana glauca resulting in variants with three or six pseudoknots. *Journal of General Virology*. 81: 577-586.

Bosma,T. J., Karagiannis, K., Santana-Quintero, L., Ilyushina, N., Zagorodnyaya, T., Petrovskaya, S., Laassri, M., Donnelly, R. P., Rubin, S., Simonyan, V. and Sauder, C. J. (2019). Identification and quantification of defective virus genomes in high throughput sequencing data using DVG-profiler, a novel post-sequence alignment processing algorithm. *PLoS ONE*. 14: e0216944.

Boyer, J. C. and Haenni, A. L. (1994). Infectious transcripts and cDNA clones of RNA viruses. *Virology*. 198: 415-426.

Brumin, M., Stukalov, S., Haviv, S., Muruganantham, M., Moskovitz, Y., Batuman, O., Fenigstein, A. and Mawassi, M. (2009). Post-transcriptional gene silencing and virus resistance in *Nicotiana benthamiana* expressing a Grapevine virus A minireplicon. *Transgenic Research*. 18: 331-345.

Burch-Smith, T. M., Anderson, J. C., Martin, G. B. and Dinesh-Kumar, S. P. (2004). Applications and advantages of virus-induced gene silencing for gene function studies in plants. *The Plant Journal*. 39: 734–746.

Calain, P., Monroe, M. C. and Nichol, S. T. (1999). Ebola virus defective interfering particles and persistent infection. *Virology*. 262: 114-128.

Callaway, A., Giesman-Cookmeyer, D., Gillock, E. T., Sit, T. L. and Lommel, S. A. (2001). The multifunctional capsid proteins of plant RNA viruses. *Annual Review of Phytopathology*. 39: 419–460.

Celix, A., Luis Arteaga, M. and Rodriguez Cerezo, E. (1996). First report of cucumber green mottle mosaic tobamovirus infecting greenhouse-grown cucumber in Spain. *Plant Disease*. 80: 1303.

Chandrika, R., Rabindran, S., Lewandowski, D. J., Manjunath, K. L. and Dawson, W. O. (2000). Full-length tobacco mosaic virus RNAs and defective RNAs have different 3' replication signals. *Virology*. 273:198-209.

Chapman, M. R. and Kao, C. C. (1999). A minimal RNA promoter for minus-strand RNA synthesis by the brome mosaic virus polymerase complex. *Journal of Molecular Biology*. 286: 709-720.

Chapman, S., Kavanagh, T. and Baulcombe, D. (1992). Potato virus X as a vector for gene expression in plants. *The Plant Journal*. 2: 549–557.

Chen, H. Y., Lin, S. M., Chen, Q., Zhao, W. J., Liao, F. R., Chen, H. J. and Zhu, S. F. (2009). Complete genome sequence of a watermelon isolate of *Cucumber green mottle mosaic virus* in northern China. *Bing Du Xue Bao.* 25: 68-72. (In Chinese)

Chen, I. H., Huang, Y. W. and Tsai, C. H. (2017). The functional roles of the *Cis*-acting elements in *Bamboo mosaic virus* RNA Genome. *Frontiers in Microbiology.* 8: 645.

Chen, J., Noueiry, A. O. and Ahlquist, P. (2001). Brome mosaic virus Protein 1a recruits viral RNA2 to RNA replication through a 5′ proximal RNA2 signal. *Journal of Virology*. 75: 3207–3219.

Chen, M. H., Roossinck, M. J. and Kao, C. C. (2000). Efficient and specific initiation of subgenomic RNA synthesis by cucumber mosaic virus replicase in vitro requires an upstream RNA stem-loop1. *Journal of Virology*. 74: 11201–11209.

Chen, Q. and Davis, K. R. (2016). The potential of plants as a system for the development and production of human biologics. *F1000Research*. 5: F1000 Faculty Rev-912. Doi: 10.12688/f1000research.8010.1.

Chen, V. B., Arendall, W. B. 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica, Section D: Biological crystallography*. 66: 12-21.

Choi, B., Kwon, S. J., Kim, M. H., Choe, S., Kwak, H. R., Kim, M. K., Jung, C. and Seo, J. K. (2019). A Plant Virus-Based Vector System for Gene Function Studies in Pepper. *Plant Physiology*. 181: 867–880.

Chujo, T., Ishibashi, K., Miyashita S. and Ishikawa M. (2015). Functions of the 5′-and 3′-untranslated regions of tobamovirus RNA. *Virus Research*. 206: 82–89.

Coates, N. and McCarthy, M. (2015). Queensland losing the biosecurity battle as Cucumber green mottle mosaic virus detected on north Queensland watermelon farm. ABC Rural. https://www.abc.net.au/news/rural/2015-04-21/major-melon-virus-detected-on-north-queensland-farm-biosecurity/6407984.

Colovos, C. and Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Science*. 2: 1511-1519.

Darty, K., Denise, A. and Ponty, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics.* 25: 1974-1975.

Ding, X. S., Liu, J., Cheng, N. H., Folimonov, A., Hou, Y. M., Bao, Y., Katagi, C., Carter, S. A. and Nelson, R. S. (2004). The tobacco mosaic virus 126-kDa protein associated with

virus replication and movement suppresses RNA silencing. *Molecular Plant-Microbe Interaction*. 17: 583-592.

Ding, X. S., Mannas, S. W., Bishop, B. A., Rao, X., Lecoultre, M., Kwon, S. and Nelson, R. S. (2018). An improved *Brome mosaic virus* silencing vector: Greater insert stability and more extensive VIGS. *Plant Physiology*. 176: 496–510.

Dohi, K., Nishikiori, M., Tamai, A., Ishikawa, M., Meshi, T. and Mori, M. (2006). Inducible virus mediated expression of a foreign protein in suspension cultured plant cells. *Archives of Virology*. 151: 1075–1084.

Dombrovsky, A., Tran-Nguyen, L. T. T. and Jones, R. A. C. (2017). Cucumber green mottle mosaic virus: Rapidly increasing global distribution, etiology, epidemiology, and management. *Annual Review of Phytopathology*. 55: 231-256.

Donson, J., Kearney, C. M., Hilf, M. E. and Dawson, W. O. (1991). Systemic expression of a bacterial gene by a tobacco mosaic virusbased vector. *Proceedings of the National Academy of Sciences of the United States of America*. 88: 7204-7208.

Dorokhov, Y. L., Sheshukova, E. V. and Komarova, T. V. (2017). Tobamovirus 3'-terminal gene overlap may be a mechanism for within-host fitness improvement. *Frontiers in Microbiology*. 8: 851.

Dreher, T. W. (2009). Role of tRNA-like structures in controlling plant virus replication. *Virus Research*. 139: 217–229.

Dreher, T. W. and Miller, W. A. (2006). Translational control in positive strand RNA plant viruses. *Virology*. 344: 185–197.

Dunigan, D. and Zaitlin, M. (1990). Capping of Tobacco Mosaic Virus RNA: Analysis of viral-coded guanylyltransferase-like activity. *Journal of Biological Chemistry*. 265: 7779-7786.

Faivre-Rampant, O., Gilroy, E. M., Hrubikova, K., Hein, I., Millam, S., Loake, G. J, Birch, P., Taylor, M. and Lacomme, C. (2004). Potato virus X-induced gene silencing in leaves and tubers of potato. *Plant Physiology*. 134**:** 1308-1316.

Feklistov A. (2013). RNA polymerase: in search of promoters. *Annals of the New York Academy of Sciences*. 1293: 25-32.

Felden, B., Florentz, C., Giege, R. and Westhof, E. (1996). A central pseudoknotted three-way junction imposes tRNA-like mimicry and the orientation of three 5' upstream pseudoknots in the 3' terminus of tobacco mosaic virus RNA. *RNA (New York, N.Y.)*. 2: 201-212.

French, R. and Ahlquist, P. (1987). Intercistronic as well as terminal sequences are required for efficient amplification of brome mosaic virus RNA3. *Journal of Virology*. 61: 1457-1465.

Frick, D. N. and Lam, A. M. (2006). Understanding helicases as a means of virus control. *Current Pharmaceutical Design*. 12: 1315-1338.

Fukuda, M., Ohno, T., Okada, Y., Otsuki, Y. and Takebe, I. (1978). Kinetics of biphasic reconstitution of Tobacco mosaic virus *in vitro*. *Proceedings of the National Academy of Sciences of the United States of America*. 75: 1727-1730.

Fuller, F. J. and Marcus, P. I. (1980). Interferon induction by viruses. Sindbis virus: defective-interfering particles temperature-sensitive for interferon induction. *Journal of General Virology*. 48: 391–394.

Gallie, D. R. (2002). The 5'-leader of tobacco mosaic virus promotes translation through enhanced recruitment of eIF4F. *Nucleic Acids Reserach*. 30: 3401-3411.

Gallie, D. R. and Kado, C. I. (1989). A translational enhancer derived from tobacco mosaic virus is functionally equivalent to a Shine-Dalgarno sequence. *Proceedings of the National Academy of Sciences of the United States of America*. 86: 129-132.

Gallie, D. R. and Kobayashi, M. (1994). The role of the 3′-untranslated region of non-polyadenylated plant viral mRNAs in regulating translational efficiency. *Gene*. 142: 159-165.

Gleba, Y., Klimyuk, V. and Marillonnet, S. (2005). Magnifection - a new platform for expressing recombinant vaccines in plants. *Vaccine*. 23: 2042-2048.

Gleba, Y., Klimyuk, V. and Marillonnet, S. (2007). Viral vectors for the expression of proteins in plants. *Current Opinion in Plant Biology*. 18: 134-141.

Gleba, Y., Marillonnet, S. and Klimyuk, V. (2004). Engineering viral expression vectors for plants: the ‘full virus’ and the ‘deconstructed virus’ strategies. *Current Opinion in Plant Biology*. 7: 182-188.

Gleba, Y. Y., Tuse, D. and Giritch, A. (2014). Plant viral vectors for delivery by *Agrobacterium*. *Current Topics in Microbiology and Immunology*. 375: 155-192.

Goelet, P., Lomonosoff, G. P., Butler, P. J. G., Akam, M. E., Gait, M. J. and Karn, J. (1982). Nucleotide sequence of tobacco mosaic virus RNA. *Proceedings of the National Academy of Sciences of the United States of America*. 79: 5818-5822.

Gopinath K., Dragnea, B. and Kao, C. C. (2005). Interaction between brome mosaic virus proteins and RNAs: Effects on RNA replication, protein expression, and RNA stability. *Journal of Virology*. 79: 14222-14234.

Grdzelishvili, V. Z., Chapman, S. N., Dawson, W. O. and Lewandowski, D. J. (2000). Mapping of the Tobacco mosaic virus movement protein and coat protein subgenomic RNA promoters *in vivo*. *Virology*. 275: 177-192.

Grimsley, N., Hohn, T., Davies, J. W. and Hohn, B. (1987). *Agrobacterium* mediated delivery of infectious maize streak virus into maize plants. *Nature*. 325: 177-179.

Gruber A. R., Findeiß, S., Washietl, S., Hofacker, I. L. and Stadler, P. F. (2010). RNAz 2.0: Improved noncoding RNA detection. *Pacific Symposium on Biocomputing.* 15: 69-79.

Gultyaev, A. P., van Batenburg, E. and Pleij, C. W. A. (1994). Similarities between the secondary structure of satellite tobacco mosaic virus and tobamovirus RNAs. *Journal of General Virology*. 75: 2851-2856.

Gupta, D. and Ranjan, R. (2017). *In silico* comparative analysis of promoters derived from plant pararetroviruses. *Virus disease*. 28: 416-421.

Haasnoot, P. C., Olsthoorn, R. C. and Bol, J. F. (2002).The Brome mosaic virus subgenomic promoter hairpin is structurally similar to the iron-responsive element and functionally equivalent to the minus-strand core promoter stem-loop C. *RNA*. 8: 110-122.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*. 41: 95-98.

He, G., Zhang, Z., Sathanantham, P., Diaz, A. and Wang, X. (2020). Brome Mosaic Virus (*Bromoviridae*). *Reference Module in Life Sciences*. B978-0-12-809633-8.21294-6. Doi: 10.1016/B978-0-12-809633-8.21294-6.

Hiatt, A. and Pauly, M. (2006). Monoclonal antibodies from plants: a new speed record. *Proceedings of the National Academy of Sciences of the United States of America*. 103: 14645–14646.

Hillman, B. I., Carrington, J. C. and Morris, T. J. (1987). A defective interfering RNA that contains a mosaic of a plant–virus genome. *Cell*. 51: 427-433.

Hirashima, K. and Watanabe, Y. (2003). RNA helicase domain of tobamovirus replicase executes cell-to-cell movement possibly through collaboration with its nonconserved region. *Journal of Virology*. 77: 12357–12362.

Hofmann, M. A., Sethna, P. B. and Brian, D. A. (1990). Bovine coronavirus mRNA replication continues throughout persistent infection in cell culture. *Journal of Virology*. 64: 4108–4114.

Huang, A. S. and Baltimore, D. (1970). Defective viral particles and viral disease processes. *Nature*. 226: 325–327.

Huang, Y. L., Han, Y. T., Chang, Y. T., Hsu, Y. H. and Meng, M. (2004). Critical residues for GTP methylation and formation of the covalent m7GMP-enzyme intermediate in the capping enzyme domain of bamboo mosaic virus. *Journal of Virology*. 78: 1271-1280.

Hyodo, K., Nagai, H. and Okuno, T. (2017). Dual function of a cis-acting RNA element that acts as a replication enhancer and a translation repressor in a plant positive-stranded RNA virus. *Virology*. 512: 74-82.

Igarashi, A., Yamagata, K., Sugai, T., Takahashi, Y., Sugawara, E., Tamura, A., Yaegashi, H., Yamagishi, N., Takahashi, T., Isogai, M., Takahashi, H. and Yoshikawa, N. (2009). Apple latent spherical virus vectors for reliable and effective virus-induced gene silencing among a broad range of plants including tobacco, tomato, Arabidopsis thaliana, cucurbits, and legumes. *Virology*. 386: 407- 416.

Inouye, T., Inouye, N., Asatani, M. and Mitsuhata, K. (1967). Studies on *Cucumber green mottle mosaic virus* in Japan. *Ber Ohara Inst Landw Biol*. 14: 49–69.

Ishibashi, K. and Ishikawa, M. (2016). Replication of Tobamovirus RNA. *Annual Review of Phytopathology*. 54: 55-78.

Jackson, R. J. (2000). Comparative view of initiation site selection mechanisms, In: Translational Control of Gene Expression, N. Sonenberg, J. W. B. Hershey and M. B.

Mathews, (eds). Cold Spring Harbor, New York , Cold Spring Harbor Laboratory. pp 127–184.

Jailani, A. A. K., Chattopadhyay, A., Washington, O. S., Roy, A., Mukherjee, S. K., Mishra, N. S. and Bikash, M. (2020). Overlapping primer extension PCR based cloning techniques engineer a cucumber green mottle mosaic virus genome-based VIGS vector. VIROCON: International Conference on “Evolution of Viruses and Viral Diseases”. Indian Virological Society. p.112.

Jailani, A. A. K., Solanki, V., Roy, A., Sivasudha, T. and Mandal, B. (2017). A CGMMV genome-replicon vector with partial sequences of coat protein gene efficiently expresses GFP in *Nicotiana benthamiana*. *Virus Research*. 233: 77-85.

Kawamura-Nagaya, K., Ishibashi, K., Huang, Y. P., Miyashita, S. and Ishikawa, M. (2014). Replication protein of tobacco mosaic virus cotranslationally binds the 5' untranslated region of genomic RNA to enable viral replication. *Proceedings of the National Academy of Sciences of the United States of America*. 111: e1620-1628.

Keane, L. (2015). Katherine property given CGMMV green light. https://www.katherinetimes.com.au/story/2891105/katherine-property-given-cgmmv-green-light/

Kehoe, M. A., Jones, R. A. and Coutts, B. A. (2017). First complete genome sequence of *cucumber green mottle mosaic virus* isolated from Australia. *Genome Announcements*. 5: e00036-17.

Kibbe, W. A. (2007). OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research*. 35(Web Server issue): W43-46.

Kim, C. H. and Kao, C. (2001). A mutant viral RNA promoter with an altered conformation retains efficient recognition by a viral RNA replicase through a solution-exposed adenine. *RNA*. 7: 1476-1485.

Kim, C. H., Kao, C. and Tinoco, I. Jr. (2000). RNA motifs that determine specificity between a viral replicase and its promoter. *Nature Structural & Molecular Biology*. 7: 415-423.

Klimyuk, V., Pogue, G., Herz, S., Butler, J. and Haydon, H. (2014). Production of recombinant antigens and antibodies in *Nicotiana benthamiana* using 'magnifection' technology: GMP-

compliant facilities for small- and large-scale manufacturing. *Current Topics in Microbiology and Immunology*. 375: 127-154.

Knapp, E., Danyluk, G. M., Achor, D. and Lewandowski, D. J. (2005). A bipartite Tobacco mosaic virus-defective RNA (dRNA) system to study the role of the N-terminal methyl transferase domain in cell-to-cell movement of dRNAs. *Virology*. 341: 47-58.

Knapp, E., Dawson, W. O. and Lewandowski, D. J. (2001). Conundrum of the lack of defective RNAs (dRNAs) associated with tobamovirus Infections: dRNAs that can move are not replicated by the wild-type virus; dRNAs that are replicated by the wild-type virus do not move. *Journal of Virology*. 75: 5518-5525.

Knorr, D. A., Mullin, R. H., Hearne, P. Q. and Morris, T. J. (1991). De novo generation of defective interfering RNAs of tomato bushy stunt virus by high multiplicity passage. *Virology*. 181: 193-202.

Koev, G., Mohan, B. R. and Miller, W. A. (1999). Primary and secondary structural elements required for synthesis of barley yellow dwarf virus subgenomic RNA1. *Journal of Virology*. 73: 2876-2885.

Kollipara, S. K. (2016). Infectivity of *Cucumber green mottle mosaic virus* and application of the viral genome as a gene silencing vector. Ph.D. thesis, Indian Agricultural Research Institute, New Delhi.

Komarova, T. V., Skulachev, M. V., Zvereva, A. S., Schwartz, A. M., Dorokhov, Y. L. and Atabekov, J. G. (2006). New viral vector for efficient production of target proteins in plants. *Biochemistry (Moscow)*. 71: 846-850.

Komuro, Y., Tochichara, H., Fukatsu, R., Nagai, Y. and Yoneyama, S. (1971). *Cucumber green mottle mosaic virus* (watermelon strain) in watermelon and its bearing on deterioration of watermelon fruit known as "Konnyaku" disease. *Annals of the Phytopathological Society of Japan*. 37: 34-42.

Kong, F., Sivakumaran, K. and Kao, C. (1999). The N-terminal half of the brome mosaic virus 1a protein has RNA capping-associated activities: specificity for GTP and S-adenosylmethionine. *Virology*. 259: 200-210.

Koramutla, M. K., Bhatt, D., Negi, M., Venkatachalam, P., Jain, P. K. and Bhattacharya, R. (2016). Strength, stability, and cis-motifs of in silico identified phloem-specific promoters in Brassica juncea (L.). *Frontiers in Plant Science*. 7: 457. Doi: 10.3389/fpls.2016.00457.

Kubota, K., Tsuda, S., Tamai, A. and Meshi, T. (2003). Tomato mosaic virus replication protein suppresses virus-targeted posttranscriptional gene silencing. *Journal of Virology*. 77: 11016-11026.

Kumar, S., Stecher, G. and Tamura K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*. 33: 1870-1874.

Kumari, S. and Ware, D. (2013). Genome-wide computational prediction and analysis of core promoter elements across plant monocots and dicots. *PLoS ONE*. 8: e79011. Doi:10.1371/journal.pone.0079011

Laakkonen, P., Hyvonen, M., Peranen, J. and Kaariainen, L. (1994). Expression of Semliki Forest virus nsP1-specific methyltransferase in insect cells and in Escherichia coli. *Journal of Virology*. 68: 7418–7425.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 227: 680–685.

Larsen, J. S. and Curtis, W. R. (2012). RNA viral vectors for improved *Agrobacterium*-mediated transient expression of heterologous proteins in *Nicotiana benthamiana* cell suspensions and hairy roots. *BMC Biotechnology*. 12: 21.

Lazareva, E., Lezzhov, A., Vassetzky, N., Solovyev, A. and Morozov, S. (2015). Acquisition of full-length viral helicase domains by insect retrotransposon-encoded polypeptides. *Frontiers in Microbiology*. 6: 1447.

Lazzarini, R. A., Keene, J. D. and Schubert, M. (1981).The origins of defective interfering particles of the negative-strand RNA viruses. *Cell*. 26: 145–154.

Leathers, V., Tanguay, R., Kobayashi, M. and Gallie, D. R. (1993). A phylogenetically conserved sequence within viral 3' untranslated RNA pseudoknots regulates translation. *Molecular and Cellular Biology*. 13: 5331–5347.

Lecoq, H. and Desibiez, C. (2012). Viruses of cucurbit crops in the Mediterranean region: an everchanging picture, *Advances in Virus Research*. 84: 67–126.

Lescot, M., Dehais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouze, P. and Rombauts, S. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research*. 30: 325–327.

Leuzinger, K., Dent, M., Lai, H., Zhou, X. and Chen, Q. (2013). Efficient agroinfiltration of plants for high-level transient expression of recombinant proteins. *Journal of Visualized Experiments*. 77: 50521.

Lewandowski, D. J. and Dawson, W. O. (1998). Deletion of internal sequences results in tobacco mosaic virus defective RNAs that accumulate to high levels without interfering with replication of the helper virus. *Virology*. 251: 427-437.

Li, D. and Aaskov, J. (2014). Sub-genomic RNA of defective interfering (D.I.) dengue viral particles is replicated in the same manner as full length genomes. *Virology.* 46: 248–255.

Li, L., Wang, M., Chen, Y., Hu, T., Yang, Y., Zhang, Y., Bi, G., Wang, W., Liu, E., Han, J., Lu, T. and Su, D. (2020). Structure of the enterovirus D68 RNA-dependent RNA polymerase in complex with NADPH implicates an inhibitor binding site in the RNA template tunnel. *Journal of Structural Biology*. 211: 107510.

Li, R., Zheng, Y., Fei, Z. and Ling, K. S. (2015). First complete genome sequence of an emerging cucumber green mottle mosaic virus isolate in North America. *Genome Announcements*. 3: e00452-15. Doi:10.1128/genomeA.00452-15.

Li, X. H., Heaton, L. A., Morris,T. J. and Simon, A. E. (1989).Turnip crinkle virus defective interfering RNAs intensify viral symptoms and are generated de novo. *Proceedings of the National Academy of Sciences of the United States of America*. 86: 9173- 9177.

Lim, S. P., Noble, C. G., Seh, C. C., Soh, T. S., El –Sahili, A., Chan, G. K., Lescar, J., Arora, R., Benson, T., Nilar, S., Manjunatha, U., Wan, K. F., Dong, H., Xie, X., Shi, P. Y. and Yokokawa, F. (2016). Potent Allosteric Dengue Virus NS5 Polymerase Inhibitors: Mechanism of Action and Resistance Profiling. *Plos Pathogens.* 12: e1005737.

Lin, B., Parrish, C. R., Murray, J. M. and Wright, P. J. (1994). Localization of a neutralizing epitope on the envelope protein of dengue virus type 2. *Virology*. 202: 885-890.

Lindbo, J. A. (2007). TRBO: a high-efficiency tobacco mosaic virus RNA-based overexpression vector. *Plant Physiology*. 145: 1232-1240.

Liu, M., Liang, Z., Aranda, M. A., Hong, N., Liu L., Kang, B. and Gu, Q. (2020). A cucumber green mottle mosaic virus vector for virus-induced gene silencing in cucurbit plants. *Plant Methods*. 16: 9.

Liu, Y., Wang, Y., Wang, X. and Zhou, G. (2009). Molecular characterization and distribution of Cucumber green mottle mosaic virus in China. Journal of Phytopathology. 157: 393-399.

Lopez, C. B. (2014). Defective Viral Genomes: Critical Danger Signals of Viral Infections. *Journal of Virology*. 88: 8720-8723.

Lough, T. J., Lee, R. H., Emerson, S. J., Forster, R. L. and Lucas, W. J. (2006). Functional analysis of the 5′untranslated region of potexvirus RNA reveals a role in viral replication and cell-to-cell movement. *Virology.* 351: 455–465.

Lovell, S. C., Davis, I. W., Arendall, W. B. 3rd., de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. and Richardson, D. C. (2003). Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins.* 50: 437-450.

Lu, R., Martin-Hernandez, A. M., Peart, J. R., Malcuit, I. and Baulcombe, D. C. (2003). Virus-induced gene silencing in plants. *Methods*. 30: 296-303.

Lukasiak, P., Antczak, M., Ratajczak, T., Bujnicki, J. M., Szachniuk, M., Adamiak, R. W., Popenda, M. and Blazewicz, J. (2013). RNAlyzer - novel approach for quality analysis of RNA structural models. *Nucleic Acids Research*. 41: 5978-5990.

Maia, I. G., Haenni, A. and Bernardi, F. (1996). Potyviral HC-Pro: a multifunctional protein. *Journal of General Virology*. 77: 1335-1341.

Mandal, S., Mandal, B., Haq, Q. M. R. and Varma, A. (2008). Properties, Diagnosis and Management of *Cucumber green mottle mosaic virus*. *Plant Viruses.* 2: 25-34.

Marillonnet, S., Giritch, A., Gils, M., Kandzia, R., Klimyuk, V. and Gleba, Y. (2004). In planta engineering of viral RNA replicons: Efficient assembly by recombination of DNA modules delivered by Agrobacterium. *Proceedings of the National Academy of Sciences of the United States of America*. 101: 6852–6857.

Marks, H., Ren, X. Y., Sandbrink, H., van Hulten, M. C. and Vlak, J. M. (2006). *In silico* identification of putative promoter motifs of White Spot Syndrome Virus. *BMC Bioinformatics* . 7**:** 309. Doi: 10.1186/1471-2105-7-309.

Marsh, L. E., Dreher, T. W. and Hall, T. C. (1988). Mutational anlaysis of the core promoter and modulator sequence of the brome mosaic virus RNA-3 subgenomic promoter. *Nucleic Acids Reserach*. 16: 981-995.

Martinez-Salas, E., Ramos, R., Lafuente, E. and de Quinto S. L. (2001). Functional interactions in internal translation initiation directed by viral and cellular IRES elements. *Journal of General Virology*. 82: 973-984.

Mashiach, E., Schneidman-Duhovny, D., Andrusier, N., Nussinov, R. and Wolfson, H. J. (2008). FireDock: a web server for fast interaction refinement in molecular docking. *Nucleic Acids Research*. 36(Web Server issue): W229-232.

Mathews, D. H., Sabina, J., Zuker, M. and Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology*. 288: 911-940.

Matsuda, D. and Dreher, T. W. (2004). The tRNA-like structure of Turnip yellow mosaic virus RNA in a-3′-translational enhancer. *Virology*. 321: 36-46.

Mclaren, L. C. and Holland, J. J. (1974). Defective interfering particles from poliovirus vaccine and vaccine reference strains. *Virology*. 60: 579–583.

Meng, M. and Lee, C. C. (2017). Function and Structural Organization of the Replication Protein of Bamboo mosaic virus. *Frontiers in Microbiology*. 8: 522.

Merits, A., Kettunen, R., Makinen, K., Lampio, A., Auvinen, P., Kaariainen, L. and Ahola, T. (1999). Virus-specific capping of tobacco mosaic virus RNA: methylation of GTP prior to formation of covalent complex p126-m^7GMP. *FEBS Letters*. 455: 45-48.

Miller, W. A., Dinesh-Kumar, S. P. and Paul, C. P. (1995). Luteovirus gene expression. *Critical Reviews in Plant Sciences*. 14: 179–211.

Miller, W.A. and Koev, G. (2000). Synthesis of subgenomic RNAs by positive-strand RNA viruses. *Virology*. 273: 1-8.

Mortimer, C. L., Dugdale, B. and Dale, J. L. (2015). Updates in inducible transgene expression using viral vectors: from transient to stable expression. *Current Opinion in Plant Biology*. 32: 85-92.

Mundry, K. W., Watkins, P. A., Ashfield, T., Plaskitt, K. A., Eisele-Walter, S. and Wilson, T. M. (1991). Complete uncoating of the 5' leader sequence of tobacco mosaic virus RNA occurs rapidly and is required to initiate cotranslational virus disassembly in vitro. *Journal of General Virology*. 72: 769–777.

Nagamatsu, A., Masuta, C., Senda, M., Matsuura, H., Kasai, A., Hong, J. S., Kitamura, K., Abe, J. and Kanazawa, A. (2007). Functional analysis of soybean genes involved in flavonoid biosynthesis by virus-induced gene silencing. *Plant Biotechnology Journal*. 5: 778-90.

Nain, V., Sahi S. and Ananda Kumar, P. (2011). In silico identification of regulatory elements in promoters, In: Computational biology and applied bioinformatics. H.S. Lopes and L.M. Cruz, (eds). Intech Open. Doi: 10.5772/22230.

Newburn, L. R. and White, K. A. (2015). *Cis*-acting RNA elements in positive-strand RNA plant virus genomes. *Virology*. 479–480: 434-443.

Nishikiori, M., Sugiyama, S., Xiang, H., Niiyama, M., Ishibashi, K., Inoue, T., Ishikawa, M., Matsumura, H. and Katoh, E. (2012). Crystal structure of the superfamily 1 helicase from Tomato mosaic virus. *Journal of Virology*. 86: 7565-7576.

Noppornpanth, S., Smits, S. L., Lien, T. X., Poovorawan, Y., Osterhaus, A. D. and Haagmans, B. L. (2007). Characterization of hepatitis C virus deletion mutants circulating in chronically infected patients. *Journal of Virology*. 81: 12496-12503.

Ogawa, T., Watanabe, Y. and Okada, Y. (1991). Trans complementation of virus-encoded replicase components of tobacco mosaic virus. *Virology*. 185: 580-584.

Ohno, T., Aoyagi, M., Yamanashi, Y., Saito, H., Ikawa, S., Meshi, T. and Okada, Y. (1984). Nucleotide sequence of the tobacco mosaic virus (tomato strain) genome and comparison with the common strain genome. *Journal of Biochemistry*. 96: 1915-1923.

Olsthoorn, R. C., Haasnoot, P. C. and Bol, J. F. (2004). Similarities and differences between the subgenomic and minus-strand promoters of an RNA plant virus. *Journal of Virology*. 78: 4048–4053.

Ooi, A., Tan, S., Mohamed, R., Rahman, N. A. and Othman, R. Y. (2006).The full-length clone of cucumber green mottle mosaic virus and its application as an expression system for Hepatitis B surface antigen. *Journal of Biotechnology*. 121: 471-481.

O'Reilly, E. K. and Kao, C. C. (1998). Analysis of RNA-dependent RNA polymerase structure and function as guided by known polymerase structures and computer predictions of secondary structure. *Virology*. 252: 287-303.

Osman, T. A., Hemenway, C. L. and Buck, K. W. (2000). Role of the 3' tRNA-like structure in tobacco mosaic virus minus-strand RNA synthesis by the viral RNA-dependent RNA polymerase *in vitro*. *Journal of Virology*. 74: 11671-11680.

Padmanabhan, M. S., Goregaoker, S. P., Golem, S., Shiferaw, H. and Culver, J. N. (2005). Interaction of the tobacco mosaic virus replicase protein with the Aux/IAA protein PAP1/IAA26 is associated with disease development. *Journal of Virology* .79: 2549-2558.

Pathak, K. B., Pogany, J. and Nagy, P. D. (2011). Non-template functions of the viral RNA in plant RNA virus replication. *Current Opinion in Virology.* 1: 332-338.

Perrault, J. (1981). Origin and replication of defective interfering particles. *Current Topics in Microbiology and Immunology*. 93: 151-207.

Petrova, E. K., Nikitin, N. A., Protopopova, A. D., Arkhipenko, M. V., Yaminsky, I. V., Karpova, O. V. and Atabekob, J. G. (2013). The role of the 5′-cap structure in viral ribonucleoproteins assembly from potato virus X coat protein and RNAs. *Biochimie*. 95: 2415-2422.

Pogany, J., Romero, J., Huang, Q., Sgro, J. Y., Shang, H. and Bujarski, J. J. (1995). De novo generation of defective interfering-like RNAs in broad bean mottle bromovirus. *Virology*. 212: 574-586.

Popenda, M., Szachniuk, M., Antczak, M., Purzycka, K. J., Lukasiak, P., Bartol, N., Blazewicz, J. and Adamiak, R. W. (2012). Automated 3D structure composition for large RNAs. *Nucleic acids Research*. 40: e112.

Raffo, A. J. and Dawson, W. O. (1991).Construction of tobacco mosaic virus subgenomic replicons that are replicated and spread systemically in tobacco plants. *Virology*. 184: 277-289.

Ramegowda, V., Mysore, K. S. and Senthil-Kumar, M. (2014). Virus-induced gene silencing is a versatile tool for unraveling the functional relevance of multiple abiotic-stress-responsive genes in crop plants. *Frontiers in Plant Science*. 5: 323.

Reingold, V., Lachman, O., Belausov, E., Koren, A., Mor, N. and Dombrovsky, A. (2016). Epidemiological study of *Cucumber green mottle mosaic virus* in green houses enables reduction of disease damage in cucurbit production. *Annals of Applied Biology.* 168: 29-40.

Reingold, V., Lachman, O., Koren, A. and Dombrovsky, A. (2013). First report of *Cucumber green mottle mosaic virus* (CGMMV) symptoms in watermelon used for the discrimination of non-marketable fruits in Israeli commercial fields. *New Disease Reports*. 28: 11. Doi:10.5197/j.2044-0588.2013.028.011.

Rezelj, V. V., Levi, L. I. and Vignuzzi, M. (2018). The defective component of viral populations. *Current Opinion in Virology*. 33: 74-80.

Rhee, S. J., Jang, Y. J. and Lee, G. P. (2016). Identification of the subgenomic promoter of the coat protein gene of cucumber fruit mottle mosaic virus and development of a heterologous expression vector. *Archives of Virology*. 161: 1527-1538.

Robertson, D. (2004). VIGS vectors for gene silencing: Many targets, many tools. *Annual Review of Plant Biology*. 55: 495–519.

Roossinck, M. J. (1997). Mechanisms of plant virus evolution. *Annual Review of Phytopathology* . 35: 191-209.

Rossmann, M. G. (2013). Structure of viruses: a short history. *Quarterly Reviews of Biophysics*. 46: 133-180.

Roy , G., Weisburg, S., Rabindran, S. and Yusibov, V. (2010). A novel two-component Tobacco mosaic virus-based vector system for high-level expression of multiple therapeutic proteins including a human monoclonal antibody in plants. *Virology*. 405: 93-99.

Roy, G., Weisburg, S., Foy, K., Rabindran, S., Mett, V. and Yusibov, V. (2011). Co-expression of multiple target proteins in plants from a tobacco mosaic virus vector using a combination of homologous and heterologous subgenomic promoters. Archives of Virology. 156: 2057-2061.

Sainsbury, F. and Lomonossoff, G. P. (2008). Extremely high-level and rapid transient protein production in plants without the use of viral replication. *Plant Physiology*. 148: 1212–1218.

Sainsbury, F., Lavoie, P. O., D'Aoust, M. A., Vezina, L. P. and Lomonossoff, G. P. (2008). Expression of multiple proteins using full-length and deleted versions of cowpea mosaic virus RNA-2. *Plant Biotechnology Journal*. 6: 82–92.

Saira, K., Lin, X., DePasse, J. V., Halpin, R., Twaddle, A., Stockwell, T., Angus, B., Cozzi-Lepri, A., Delfino, M., Dugan, V., Dwyer, D. E., Freiberg, M., Horban, A., Losso, M., Lynfield, R., Wentworth, D. N., Holmes, E. C., Davey, R., Wentworth, D. E., Ghedin, E., INSIGHT FLU002 and FLU003Study Group (2013). Sequence analysis of in vivo defective interfering-like RNA of influenza A H1N1 pandemic virus. *Journal of Virology*. 87: 8064–8074.

Sambrook, A. and Russel, D.W. (2001). Molecular Cloning: A Laboratory Manual.: 3rd ed. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY.

Santak, M., Markusic, M., Balija, M. L., Kopac, S. K., Jug, R., Orvell, C., Tomac, J. and Forcic, D. (2015). Accumulation of defective interfering viral particles in only a few passages in Vero cells attenuates mumps virus neurovirulence. *Microbes and Infection*.17: 228-236.

Sawicki, S. G. and Sawicki, D. L. (1990). Coronavirus transcription: subgenomic mouse hepatitis virus replicative intermediates function in RNA synthesis. *Journal of Virology*. 64: 1050–1056.

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. and Wolfson, H. J. (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Research*. 33: W363-367.

Sethna, P. B., Hung, S. L. and Brian, D. A. (1989). Coronavirus subgenomic minus-strand RNA and the potential for mRNA replicons. *Proceedings of the National Academy of Sciences of the United States of America*. 86: 5626–5630.

Shao, Y., Zhu, H. L., Tian, H. Q., Wang, X. G. and Lin, X. J. (2008). Virus –induced gene silencing in plant species. *Russian Journal of Plant Physiology*. 55: 168-174.

Shepherd, R. J. (1989). Biochemistry of DNA plant viruses. *In:* The Biochemistry of Plants. A. Marcus, (ed). Academic Press, New York. pp. 536–616.

Shirokikh, N. E., Agalarov, S. Ch. and Spirin, A. S. (2010). Chemical and enzymatic probing of spatial structure of the omega leader of tobacco mosaic virus RNA. *Biochemistry (Moscow)*. 75: 405-411.

Shivprasad, S., Pogue, G. P., Lewandowski, D. J., Hidalgo, L., Donson, J., Grill, L. K. and Dawson, W. O. (1999). Heterologous sequences greatly affect foreign gene expression in tobacco mosaic virus-based vectors. *Virology*. 255: 312-323.

Simon, A. E. and Gehrke, L. (2009). RNA conformational changes in the life cycles of RNA viruses, viroids, and virus-associated RNAs. *Biochimica et Biophysica Acta*. 1789: 571-583.

Simon-Buela, L. and Garcia-Arenal, F. (1999). Virus particles of cucumber green mottle mosaic tobamovirus move systemically in the phloem of infected cucumber plants. *Molecular Plant-Microbe Interactions*. 12: 112-118.

Sivakumaran, K., Bao, Y., Roossinck, M. J. and Kao, C. C. (2000**).** Recognition of the core RNA promoter for minus-strand RNA synthesis by the replicases of Brome mosaic virus and Cucumber mosaic virus. *Journal of Virology*. 74: 10323-10331.

Sivakumaran, K., Hema, M. and Kao, C. C. (2003). Brome mosaic virus RNA syntheses *in vitro* and in barley protoplasts. *Journal of Virology*. 77**:** 5703-5711.

Sivakumaran, K., Kim, C. H., Tayon, R. Jr. and Kao, C. C. (1999). RNA sequence and secondary structural determinants in a minimal viral promoter that directs replicase recognition and initiation of genomic plus-strand RNA synthesis. *Journal of Molecular Biology*. 294**:** 667-682.

Sperschneider, J. and Datta, A. (2010). DotKnot: pseudoknot prediction using the probability dot plot under a refined energy model. *Nucleic Acids Research*. 38: e103.

Steckert, J. J. and Schuster, T. M. (1982). Sequence specificity of trinucleoside diphosphate binding to polymerized tobacco mosaic virus protein. *Nature*. 299: 32-36.

Strahle, L., Garcin, D. and Kolakofsky, D. (2006). Sendai virus defective-interfering genomes and the activation of interferon-beta. *Virology*. 351: 101-111.

Takamatsu, N., Watanabe, Y., Iwasaki, T., Shiba, T., Meshi, T. and Okada, Y. (1991). Deletion analysis of the 5' untranslated leader sequence of tobacco mosaic virus RNA. *Journal of Virology*. 65: 1619-1622.

Tapia, K., Kim, W. K., Sun, Y., Mercado-Lopez, X., Dunay, E., Wise, M., Adu, M. and Lopez, C. B. (2013). Defective viral genomes arising in vivo provide critical danger signals for the triggering of lung antiviral immunity. *PLoS Pathogens*. 9: e1003703. Doi: 10.1371/journal.ppat.1003703.

Teycheney, P. Y., Marais, A., Svanella-Dumas, L., Dulucq, M. J. and Candresse, T. (2005). Molecular characterization of banana virus X (BVX), a novel member of the Flexiviridae family. *Archives of Virology*. 150: 1715–1727.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*. 22 : 4673-4680.

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

Tran, H. H., Chen, B., Chen, H., Menassa, R., Hao, X., Bernards, M., Huner, N. P. A. and Wang, A. (2019). Development of a cucumber green mottle mosaic virus-based expression vector for the production in cucumber of neutralizing epitopes against a devastating animal virus. *Journal of Virological Methods*. 269: 18-25.

Treuhaft, M. W. and Beem, M. O. (1982). Defective interfering particles of respiratory syncytial virus. *Infection and Immunity*. 37: 439-444.

Turnage, M. A., Muangsan, N., Peele, C. G. and Robertson, D. (2002). Geminivirus-based vectors for gene silencing in Arabidopsis. *The Plant Journal*. 30: 107-114.

Turner, D. R., Joyce, L. E. and Butler, P. J. (1988). The tobacco mosaic virus assembly origin RNA. Functional characteristics defined by directed mutagenesis. *Journal of Molecular Biology*. 203: 531-547.

Tuszynska, I., Magnus, M., Jonak, K., Dawson, W. and Bujnicki, J. M. (2015). NPDock: a web server for protein–nucleic acid docking. *Nucleic Acids Research*. 43(W1): W425-W430.

Ugaki, M., Tomiyama, M., Kakutani, T., Hidaka, S., Kiguchi, T., Nagata, R., Sato, T., Motoyoshi , F. and Nishiguchi, M. (1991). The complete nucleotide sequence of

cucumber green mottle mosaic virus (SH strain) genomic RNA. *Journal of General Virology*. 72: 1487-1495.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R. and Leunissen, J. A. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research*. 35(Web Server issue):W71-7 4.

van Belkum, A., Abrahams, J. P., Pleij, C. W. A. and Bosch, L. (1985). Five pseudoknots are present at the 204 nucleotides long 39 noncoding region of tobacco mosaic virus RNA. *Nucleic Acids Research*. 13: 7673- 7686.

van der Kuyl, A. C., Langereis, K., Houwing, C. J., Jaspars, E. M. J. and Bol, J. F. (1990). *Cis*-acting elements involved in replication of alfalfa mosaic virus RNAs *in vitro*. *Virology*. 176: 346-354.

van der Vossen, E. A. G., Notenboom, T., and Bol, J. F. (1995). Characterization of sequences controlling the synthesis of alfalfa mosaic virus subgenomic RNA *in vivo*. *Virology*. 212: 667-672.

Varveri, C., Vassilakos, N. and Bem, F. (2002). Characterization and detection of *Cucumber green mottle mosaic virus* in Greece. *Phytoparasitica.* 30: 493-501.

Vasilijevic, J., Zamarreno, N., Oliveros, J. C., Rodriguez-Frandsen, A., Gomez, G., Rodriguez, G., Perez-Ruiz, M., Rey, S., Barba, I., Pozo, F., Casas, I., Nieto, A., Falcon, A. (2017). Reduced accumulation of defective viral genomes contributes to severe outcome in influenza virus infected patients. *PLoS Pathogens*. 13: e1006650.

Venkataraman, S., Prasad, B. V. L. S. and Selvarajan, R. (2018). RNA Dependent RNA Polymerases: Insights from Structure, Function and Evolution. *Viruses*. 10: 76.

Vignuzzi, M. and Lopez, C. B. (2019). Defective viral genomes are key drivers of the virus-host interaction. *Nature Microbiology*. 4: 1075-1087.

Wang, X., Kelman, Z. and Culver, J. N. (2010). Helicase ATPase activity of the tobacco mosaic virus 126-kDa protein modulates replicase complex assembly. *Virology*. 402: 292-302.

Watanabe, Y., Meshi, T. and Okada, Y. (1987). Infection of tobacco protoplasts with in vitro transcribed tobacco mosaic virus RNA using an improved electroporation method. *FEBS Letters*. 219: 65-69.

Weigel, D. and Glazebrook, J. (2006). Transformation of agrobacterium using the freeze-thaw method. *Cold Spring Harbor Protocols*. 2006 (7). Doi: 10.1101/pdb.prot4666.

Wells, D. R., Tanguay, R. L., Le, H. and Gallie, D. R. (1998). HSP101 functions as a specific translational regulatory protein whose activity is regulated by nutrient status. *Genes & Development*. 12: 3236-3251.

Werner, S., Breus, O., Symonenko, Y., Marillonnet, S. and Gleba Y. (2011). High level recombinant protein expression in transgenic plants by using a double-inducible viral vector. *Proceedings of the National Academy of Sciences of the United States of America*. 108: 14061-14066.

Wiederstein, M. and Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*. 35(Web Server issue): W407–W410.

Wierzchoslawski, R. and Bujarski, J. J. (2006). Efficient in vitro system of homologous recombination in brome mosaic bromovirus. *Journal of Virology*. 80: 6182-6187.

Willemsen, A. and Zwart, M. P. (2019). On the stability of sequences inserted into viral genomes. *Virus Evolution*. 5: vez045. Doi: 10.1093/ve/vez045.

Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B. 3rd, Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S. and Richardson, D. C. (2018). MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science*. 27: 293-315.

Yan, Y., Zhang, D., Zhou, P., Li, B. and Huang, S. Y. (2017). HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Research*. 45(W1): W365–W373.

Yi, G. and Kao, C. C. (2008). *Cis*- and *trans*-acting functions of brome mosaic virus protein 1a in genomic RNA1 replication. *Journal of Virology*. 82: 3045–3053.

Yi, G., Letteney, E., Kim, C. H., and Kao, C. C. (2009). Brome mosaic virus capsid protein regulates accumulation of viral replication proteins by binding to the replicase assembly RNA element. *RNA (New York, N.Y.)*. 15: 615–626.

Young, N. D. and Zaitlin, M. (1986). An analysis of tobacco mosaic virus replicative structures synthesized in vitro. *Plant Molecular Biology*. 6: 455-65.

Yusibov, V., Rabindran, S., Commandeur, U., Twyman, R. M. and Fischer, R. (2006). The potential of plant virus vectors for vaccine production. *Drugs in R&D*. 7: 203-217.

Zavriev, S. K., Hickey, C. M., and Lommel, S. A. (1996). Mapping of the red clover necrotic mosaic virus subgenomic RNA. *Virology*. 216: 407-410.

Zenko, V. V., Ryabova, L. A., Spirin A. S., Rothnie, H. M., Hess, D., Browning, K. S. and Hohn, T. (2002). Eukaryotic elongation factor 1A interacts with the upstream pseudoknot domain in the 3' untranslated region of tobacco mosaic virus RNA. *Journal of Virology*. 76: 5678–5691.

Zhang, C. and Ghabrial, S. A. (2006). Development of Bean pod mottle virus-based vectors for stable protein expression and sequence-specific virus-induced gene silencing in soybean. *Virology*. 344: 401–411.

Zhang, Z., Liu, L., Wu, H., Liu, L., Kang, B., Peng, B. and Gu, Q. (2017). The 96th amino acid of the coat protein of cucumber green mottle mosaic virus affects virus infectivity. *Viruses*. 10: 6.

Zheng, H., Xiao, C., Han, K., Peng, J., Lin, L., Lu, Y., Xie, L., Wu, X., Xu, P., Li, G., Chen, J. and Yan, F. (2015). Development of an agroinoculation system for full-length and GFP-tagged cDNA clones of cucumber green mottle mosaic virus. *Archives of Virology*. 160: 2867-2872.

Zimmern, D. (1983). An extended secondary structure model for the TMV assembly origin, and its correlation with protection studies and an assembly defective mutant. *The EMBO Journal*. 2: 1901–1907.

Zok, T., Antczak, M., Zurkowski, M., Popenda, M., Blazewicz, J., Adamiak, R. W. and Szachniuk M. (2018). RNApdbee 2.0: multifunctional tool for RNA structure annotation. *Nucleic Acids Research*. 46(W1): W30-W35.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*. 31: 3406-3415.

Online link of webservers

https://www.ncbi.nlm.nih.gov/nuccore/DQ767631.2

https://rfam.xfam.org/

https://pfam.xfam.org/

http://mfold.rna.albany.edu/

https://github.com/moss-lab/ScanFold

https://dotknot.csse.uwa.edu.au/

http://rna.tbi.univie.ac.at/cgi-bin/RNAz/RNAz.cgi

http://www.rcsb.org/pdb/home/home.do

https://swissmodel.expasy.org/interactive/05JjhN/

http://mordred.bioc.cam.ac.uk/~rapper/rampage.php

http://nihserver.mbi.ucla.edu/ERRATv2/

https://prosa.services.came.sbg.ac.at/prosa.php

http://molprobity.biochem.duke.edu/

http://rnalyzer.cs.put.poznan.pl/

http://rnapdbee.cs.put.poznan.pl/

http://hdock.phys.hust.edu.cn/

http://genesilico.pl/NPDock

https://pymol.org/2/

https://discover.3ds.com/discovery-studio-visualizer-

https://projects.biotec.tu-dresden.de/plip-web/plip

www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver

https://proteins.plus/

https://www.ncbi.nlm.nih.gov/

http://bioinformatics.psb.ugent.be /webtools/plantcare/html/

http://www.rcsb.org/pdb/home/home.do

https://swissmodel.expasy.org/interactive/05JjhN/

http://mordred.bioc.cam.ac.uk/ ~rapper/rampage. Php

http://molprobity.biochem. duke.edu/

http://nihserver.mbi.ucla.edu/ ERRATv2/

https://prosa.services.came. sbg.ac.at/prosa.php

http://rnacomposer.cs.put.poznan.pl/

http://rnapdbee.cs.put.poznan.pl/

http://bioinfo3d.cs.tau.ac.il/PatchDock/patchdock.html

http://bioinfo3d.cs.tau.ac.il/FireDock/php.php

http://genesilico.pl/NPDock

https://pymol.org/2/

https://discover.3ds.com/discovery-studio-visualizer-

https://projects.biotec.tu-dresden.de/plip-web/plip

http://mfold.rna.albany.edu/

https://dotknot.csse.uwa.edu.au/cgibin/waiting.py?realpage=SI8VMd5AQc

http://www.ncbi.nih.gov/index.html

http://biotools.nubic.northwestern.edu/OligoCalc.html

http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/

APPENDIX-I

COMMON REAGENTS, BUFFERS AND MEDIA USED

1. ANTIBIOTICS

Ampicillin	Stock solution (100 mg/ml) of the antibiotic Ampicillin sodium salt, HIMEDIA Laboratories [CAS No. 69-52-3 MW= 371.39] was made in double distilled water, filter sterilized (through 0.22 micron filter) and distributed into 200 µl aliquots and stored at -20 ^{0}C. It was used at a concentration of 100 µg/ml.
Kanamycin	Stock Solution (50 mg/ml) of Kanamycin. 0.5 g of Kanamycin (Kanamycin monosulfate), HIMEDIA Laboratories [CAS No: 25389-94-0 MW = 582.58] was weighed and dissolved in 10 ml of sterile water. 0.22 µm syringe filter was prewetted by drawing through 5-10 ml of sterile H_2O and water was discarded. Kanamycin stock solution was sterilized through the prepared 0.22 µm syringe filter and stock may be kept at -20 °C for 1 year and utilised at a working concentration of 50 µg/ml.
Rifampicin	Stock solution (20 mg/ml) of Rifampicin, Biogene, USA [CAS # 13292-46-1] was prepared in methanol/DMSO and used at a final concentration of 20 µg/ml. (Storage at -20 °C)

*All antibiotics were filter sterilized (through 0.22 micron filter) and distributed into 500µl aliquots and stored at-20^{0}C

2. ELECTROPHORESISREAGENTS

50XTAE	Trisbase	:242 g
	Glacialacetic acid	:57.1ml
0.5 M EDTA (pH8.0)	:100ml Distilled water to 1Litre	
Loading dye	1% Bromophenol blue	:200µl
	Glycerol	200µl
10% SDS	60µl	
0.5M EDTA	50µl	
10X TAE	60µl	
Distilled water	30µl	

3. MEDIA

1. LuriaAgar medium	Bacto-tryptone	10.0 g
	Bacto-yeast extract	5.0 g
	NaCl	10.0g
	Agar	15.0g
	Deionized water	950
2. Luria Broth medium	Tryptone	:10.0 g
	Yeast extracts	:5.0 g
	NaCl	:5.0 g
	Deionized water	:950 ml

3. YEP medium

Yeast extract : 10 gm/l

Peptone : 10 gm/l

NaCl : 5 gm/l

The pH was adjusted to 7 with 2 N NaOH. The agar was added to 1.5 %.

4. AB medium

AB salts (20X, 100 ml)

NH_4Cl : 2 gm

$MgSO_4 . 7H_2O$: 0.6 gm

KCl : 0.3 gm

$CaCl_2 . 2H_2O$: 0.31 gm

$FeSO_4 . 7H_2O$: 0.005 gm

The solution was autoclaved after making up the volume to 100 ml.

5. AB buffer (20 X, 100 ml)

K_2HPO_4 : 0.6 gm

$NaH_2PO_4 2H_2O$: 2.6 gm

These were dissolved separately, the pH was adjusted to 7 to 7.2, and the volume was made up to 100 ml and was autoclaved.

4. PLASMID ISOLATION BUFFERS

1. Solution I

Tris-HCl (pH 8.0) : 1.0 M

EDTA (pH 8.0) : 0.5 M

Glucose : 0.1 M

Lysozyme : 2.0 mg

2. Solution II

NaOH : 2M

SDS : 10 %

3. Solution III

Sodium acetate (pH 5.2) : 3 M

Tris HCl pH 8.0 : 1 M

4. RNase solution: 10 mM Tris-HCl pH 7.5,15 mM NaCl

5. Proteinase K (20 mg/ml stock)

5. AGROINFILTERATION Inoculation Buffer

MES : 10mM (Filter sterilized)

MgCl2 : 10mM (Autoclaved)

Acetosyringone : 150µM

6.OTHER BUFFERS/ REAGENTs

Phosphate buffer

Potassium dihydrogenphosphate(KH_2PO_4) 1.362 gm/l

Disodium hydrogen phosphatedihydrate ($Na_2HPO_4.2H_2O$) 1.781gm/l

51.0 ml $Na_2HPO_4.2H_2O$ solution mixed with 49.0 ml KH_2PO_4 solution gives 0.01 M phosphate buffer pH7.0.

10 X MOPS buffer (pH 7.0)

Chemical Component	Final concentration
MOPS	0.2 M
Sodium acetate •3H2O	80 mM/1 M
EDTA	10 mM/0.5 M
Diethyl pyrocarbonate (DEPC)-treated H2O	Adjust the volume upto 1 liter

The solution was sterilized through membrane filtration (a 0.45-µm Millipore filter), and stored it at room temperature with light protected condition. The buffer turns yellows with age or if it is exposed to light or autoclaved. Straw-colored buffer works well, but darker buffer does not recommended.

7. PREPARATION OF COMMONLY USED STOCKSOLUTION

Solution	Method of preparation
1M $CaCl_2$	54.0 g of $CaCl_2.2H_2O$ was dissolved in 200 ml of pure water (Mill-Q or equivalent). The solution was sterilised by passing through a 0.22 micron filter and stored in 1 ml aliquots at 4^0C.
0.5 M EDTA (pH 8.0)	186.1 g of ethylenediaminetetra aceticacid disodium salt.$2H_2O$ was added to 800 ml of distilled water, stirred vigorously on a magnetic stirrer, pH was adjusted to 8.0 with NaOH (20.0 g of NaOH pellets). Volume made upto 1 Lwith distilled water, dispensed into aliquots and sterilised byautoclaving.
Ethidium bromide (10 mg ml^{-1})	1.0 g of Ethidium bromide was added to 100 ml of distilled water and stirred on a magnetic stirrer for several hours to ensure that the dye has dissolved. The solution was transferred to a dark bottle and stored at room temperature.
Phenol:Chloroform: isoamyl alcohol	Buffer saturated phenol, chloroform and isoamyl alcohol were mixed in the ratio of 25:24: 1. The equilibrated mixture was stored under a layer of 0.01 M Tris-HCl (pH 7.6) at 4^0C in dark glass bottle.

Chloroform:isoamylalcohol Buffer saturated chloroform and isoamyl alcohol were mixed in the ratio of 24: 1.

IPTG (Isopropyl-β-D-thiogalactopyranoside)

A solution of IPTG was made by dissolving 2.0 g of IPTG in 8 ml of distilled water. Volume was made upto 10 ml with distilled water and sterilised by filtration through a 0.22 micron disposable filter. The solution was dispensed into 1 ml of aliquots and stored at-20^0C.

1M $MgCl_2$ 203.3 g of $MgCl_2.6H_2O$ was dissolved in 800 ml of distilled water. The volume was made upto 1L, dispensed into aliquots and sterilised by autoclaving.

3M Sodium acetate(pH4.8) 408.1 g of $NaOAc.3H_2O$ was dissolved in 800ml of distilled water. The pH was adjusted to 4.8 with glacial acetic acid. Volume made upto 1L with distilled water, dispensed into aliquots and sterilised by autoclaving.

5MNaCl 233.8 g of NaCl was dissolved in 800 ml of distilled water, volume made upto 1L with distilled water, dispensed into aliquots and sterilised by autoclaving.

10% SDS solution 100.0 g of electrophoresis grade **Sodium dodecyl sulphte (**SDS) was dissolved in 900 ml of distilled water, heated at 68 ^{0}C to assist dissolution and pH was adjusted to 7.2 by adding few drops of concentrated HCl. The volume was made up to 1L with distilled water, dispensed in aliquots.

10NNaOH 400 g of NaOH was dissolved in 800 ml of distilled water,and the volume was made up to 1L with distilledwater.

1M TrisHCl	121.1 g of Tris Base was dissolved in 800 ml of distilled water. pH was adjusted to the desired value by adding concentrated HCl (for pH 7.4, HCl 70 ml; for pH 8.0, HCl 42 ml). The solution was allowed to cool down to room temperature before making a final adjustment to the pH. The volume was made upto 1L with distilled water, dispensed into aliquots and sterilised by autoclaving.
X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside)	
	The stock solution was made by dissolving X-gal in dimethyl formamide to make a 20 mg/ml^{-1} solution and stored at-20^0C.
100mMMES	2.313 g of MES salt was dissolved in 100 ml of distilled water and pH was adjusted to 5.7with 10NNaOH

VII. Recombinant clone confirmation by Restriction digestion

The extracted plasmid construct was confirmed through restriction digestion. The reaction mixture contains 100-1000ng of plasmid DNA, 10X reaction buffer, 1µl of restriction enzyme(s) (MBI Fermatas. Inc) and DNase-free water.

Reaction Mixture

10X buffer	:2µl
Plasmid DNA (0.1-1.0µg)	:3µl
Restriction enzyme (1U)	:1µl
DNase-free water to make final volume	:20µl

The reaction mixture was incubated at 37°C for 2-3 hrs and insert size was assessed by agarose gel electrophoreses.

VIII. PCR protocol

Primary denaturation at 94˚C for 3mins; followed by 35 cycles of final denaturation at 94˚C for 30sec, annealing at 52-60˚C for 30 sec (depending on primer Tm) and extension at 72˚C for 30sec to 2mins (depending on the amplicon size); and then final extension at 72˚C for 10 mins.

IX. SDS-PAGE

Acrylamide bisacrylamide solution (30 %)

Acryl amide	: 29.2 %
N, N' methylene-bisacrylamide	: 0.8 %

Separating gel (10 %) for 10 ml

30 % acryl amide -bisacrylamide	: 3.3 ml
Tris HCl (pH 8.8 and1.5 M)	: 2.5 ml
SDS 10 %	: 0.1 ml

APS 10% : 0.1 ml

TEMED : 20 µl

Distilled H2O : 4.0 ml

Stacking gel (4 %)

30 % acrylamide -bisacrylamide : 1.33 ml

Tris HCl (pH 6.8 and 0.5 M) : 2.5 ml

SDS 10 % : 0.1 ml

APS : 0.1 ml

TEMED : 20 µl

Distilled H2O : 6.1 ml

Sample buffer (2X)

Tris HCl (pH 6.25) : 0.12 M

Bromophenol blue : 0.05 %

SDS : 4 %

Glycerol : 20 %

β-mercaptoethanol : 10 %

Running buffer (pH 8.3)

Glycine : 0.192 M

Tris HCl : 0. 025 M

SDS : 0.1 %

Staining solution

Coomassie Brilliant blue (R 250) : 0.25 %

Methanol : 50 %

Acetic acid : 7 %

Destaining solution

Methanol : 45 %

Acetic acid : 5 %

X. Western blotting (Towbin, et al., 1979).

Towbin's buffer (Transfer buffer)

Tris HCl	: 0.25 M
Glycine	: 0.192 M
Methanol	: 20 %
SDS	: 0.03 %

TBS (pH 7.5)

Tris HCl	: 0.1 M
NaCl	: 0.9 %

TTBS

Tween 20	: 0.1 % in TBS

Blocking solution

Gelatine	: 5 % in TBS

Diluting solution

Gelatin	: 3 % in TBS