University of Latvia

Faculty of Biology

Ina Baļķe

Doctoral Thesis

Ryegrass mottle virus genome organization and characterization of its encoded proteins

Promotion to the degree of Doctor of Biology

Molecular Biology

Supervisor: Dr.biol. Andris Zeltiņš

Consultant: Dr.biol. Kaspars Tārs

Rīga, 2010

The Promotional Paper was prepared at Latvian Biomedical research and study centre during 2005 - 2009. This study was supported by a grant from the Science Council of Latvia, European Regional Development Fond (ERDF), European Social Found (ESF), and FEBS.

KOPSAVILKUMS

Sobemovīrusu ģints pārstāvji ir vienpavediena plus RNS vīrusi, kas inficē augus. To genomu veido viena RNS 4000 – 4500 nukleotīdus gara molekula, kura sastāv no četrām atklātajām translācijas fāzēm (ORF). Genomiskā RNS 5' galā satur kovalenti piesaistītu ar vīrusa genomu saistīto proteīnu (VPg), bet 3' gals nesatur poli(A) sekvenci. Kas attiecas uz identificēto translācijas fāžu izvietojumu, tad 5' galā ir novietots ORF1, tam seko ORF2a un ORF2b, bet 3' galā atrodas ORF3. Pirmie trīs ORF kodē nestrukturālos proteīnus, kas tiek translēti no genomiskās RNS vai nu kā atsevišķs proteīns, vai kā poliproteīni, kas tiek sašķelti, izmantojot vīrusa proteāzi. Vienīgais struktūras proteīns ir apvalka proteīns (CP), ko kodē genoma beigās esošais ORF, bet tas tiek translēts no subgenomiskās RNS.
Sākotnēji sobemovīrusi tika iedalīti divās grupās atkarībā no to ORF organizācijas atšķirībām genoma centrā. Tika uzskatīts, ka ir SCPMV-tips, kur ir liels centrālais ORF2 ar tajā pilnībā iekļāvušos mazu ORF3, un CfMV-tips, kur ORF2 ir sadalīts divos atsevišķos ORF2a un ORF2b un nav mazā ORF3.
Sobemovīrusu ikosahedriskajiem kapsīdiem ir T=3 simetrija, ko veido 180 CP subvienības.
Poliproteīni satur vairākus proteīnu domēnus – 3C serīna-tipa proteāzi, VPg, 2a C-gala domēnu (Pn) un RNS atkarīgo RNS polimerāzi (RdRp). Informācija par šo proteīnu funkcijām un īpašībām ir niecīga. Ir zināms, ka P1, iespējams, nodrošina vīrusa transportu no šūnas uz šūnu, kā arī sistēmisku infekciju. P1 arī darbojas kā pēctranslācijas gēnu klusinātājs. Vienīgie no sobemovīrusu proteīniem, kam ir zināma 3D struktūra, ir CP un proteāze. VPg darbojas līdzīgi kā „cepurītes" struktūra eikariotu mRNS translācijā, ir konstatēts, ka tas kalpo kā proteāzes aktivātors, kā arī mijiedarbojas ar augu translācijas iniciācijas faktoriem. Sobemovīrusu VPg strukturāli pieder pie dabīgi nestrukturētiem proteīniem. Informācija par Pn ir nepilnīga. Ir zināms, ka tas, iespējams, darbojas kā RdRp translācijas regulators un tam ir ATFāzes aktivitāte un RNS saistošas īpašības. SeMV RdRp ir raksturota tikai *in vitro*, iespējams, tās darbībai nemaz nav nepieciešams VPg kā tika uzskatīts iepriekš.
Kā pirmais šī darba uzdevums bija izveidot pilna garuma daudzziedu airenes vīruss (RGMoV) kDNS no inficēta augu materiāla. Šī darba gaitā tika pārsekvenēts viss RGMoV genoms. Sekvenču salīdzinošā analīze uzrādīja vairākas nozīmīgas atšķirības. Konstatētā nukleotīdu insercija un delēcija proteāzes domēnā, izveidoja jaunu 52 AA garu sekvenci, kas saturēja katalītiskā un substrāta saistošo centru AA atlikumus, kas netika konstatēti iepriekš. Jaunās sekvences salīdzinošā analīze uzrādīja no 38.5% līdz 50.9% līdzību ar citu sobemovīrusus proteāzēm. Jaunās RGMoV sekvences papildus analīzē tika konstatēts, ka RdRp tiek translēta caur – 1 ribosomlā rāmja nobīdi. Šie rezultāti ļauj klasificēt RGMoV genoma organizāciju kā CfMV-tipa, nevis SCPM-tipa, kā tika uzskatīts iepriekš. Šiem datiem ir taksonomiska nozīme un tie papildina koncepciju, ka visiem sobemovīrusiem ir CfMV-tipa genoma organizācija.
Otrs uzdevums bija nestrukturālo proteīnu iegūšana preparatīvos daudzumos, un to strukturāla un funkcionāla raksturošana. Proteīnu kodējošās sekvences no RGMoV kDNS tika klonētas uz pET-, pACYC- un pColdI bāzētos ekspresijas vektoros un ekspresēti vairākos *E.coli* celmos. Rezultātu analīze uzrādīja, ka proteāzes un RdRp gadījumā ekspresijai vispiemērotākā ekspresijas kombinācija bija pColdI vektors un WK6 celms. VPg un P16 tika ekspresēti, izmantojot pACYC- un pET-atvasinājumu vektorus un attīrīti ar afīno His-tag hromatogrāfiju. Neskatoties uz dažādo kristalizācijas apstākļu izmēģināšanu, VPg un P16 kristāli netika iegūti, iespējams, tas ir saistīts ar nestrukturēto AA segmentu klātbūtni proteīnu struktūrās. Kā uzrādīja vairākas bioinformātikas analīzes metodes, tad viss VPg domēns un P16 C-beigu gals satur nestrukturētas AA struktūras, kas var traucēt kristālu veidošanos.

RGMoV virionu kristalizācija tika izvirzīta kā trešais uzdevums. Par kristalizācijas materiālu kalpoja no inficētiem augiem izolēts un attīrīts RGMoV. Kristālu struktūra tika noteikta pie 2.9 Å. Iegūtie dati uzrādīja, ka RGMoV ir tā pati T=3 ikosahedrālā simetrija kā pārējiem kristalizētajiem sobemovīrusiem. CP piemīt kanoniskā β-„sendviča" struktūra. Atšķirībā no citiem sobemovīrusiem nelielas strukturālas atšķirības tika konstatētas GH un FG cilpās, kur iztrūka E1 un C spirāles. GH cilpas samazināšanās CP ietekmēja relatīvās subvienību pozīcijas. FG cilpas piedalās β-plākšņu stabilizēšanā. Kontaktu trūkums starp FG cilpu unβ-plāksnēm tiek kompensēts ar papildus kontaktiem starp N-gala AA segemntiem. Šo izmaiņu rezultātā RGMoV diametrs ir par 3% mazāks kā citiem sobemovīrusiem. Arī RGMoV iekšējais tilpums samazinās par 7%, kas ir nozīmīgs nukleīnskābes iepakošanai. Nukleīnskābes blīvums aptuveni ir 480 mg/ml, kas ir līdzīgs kā pārējiem sobemovīrusiem un tabakas nekrozes vīrusam. Kalcija jonu saistošo saistu tuvumā netika novērota kalcija jonu klātbūtne. Tam par iemeslu var būt kristalizācijas bufera zemais pH (3.0), kurā kalciju saistošās grupas nav lādētas.

ABSTRACT

Members of the genus Sobemovirus are single-stranded plus RNA viruses that infect plants. The genome consists of one RNA molecule in size 4000 – 4500 nucleotides, which contains four open reading frames (ORFs). Genomic RNA (genRNA) at its 5'-end contains covalently attached a virus-encoded protein (VPg) and 3'-end does not have poly(A) tail. As for localization of ORFs the ORF1 is located at the 5 ' end, followed by ORF2a and ORF2b, but the ORF3 is located at the 3 ' end. The first three ORFs encode non-structural proteins. Translation of ORF1 results in P1 protein, whereas ORF2a and ORF2b are translated from genRNA in a form of polyproteins, which are processed by virus protease. The only structural protein, coat protein (CP) is encoded by the last ORF3 and is translated from subgenomic RNA (sgRNA). Sobemoviruses has been originally divided into two groups depending on their differences in genome organization: SCPMV-type contains big central ORF2 with nested small ORF, and CfMV-type - ORF2 is sparated in two overlapping parts, ORF2a and ORF2b and the nested ORF is missing.

Sobemovirus icosahedral capsids have T=3 symmetry and are built up from 180 CP subunits. Polyproteins contain several identified protein domains – 3C serine like protease, VPg, C-terminal part of 2a (Pn), and RNA dependent RNA polymerase (RdRp). There are incomplete information about non-structural protein functions and properties. It is known that P1 can be involved in the transport of the virus from cell to a cell and in systemic infection. P1 protein also acts as gene post-translational silencer. Protease and coat protein are the only proteins of sobemoviruses with known three-dimensional (3D) structures. VPg can act as the "cap" structure in translation of eukaryotic mRNA, and is shown to activate the protease, as well as interacts with plant host transcription factors. Structurally sobemoviral VPg's belong to naturally disordered proteins. Information about polyprotein 2a C-terminal domain is poor. It is supposed to control the RdRp expression, and possess the ATPase and nucleic acid binding activity. RdRp of SeMV is characterized only in vitro; possibly, it does not need VPg for its activity, as it was considered previously.

The first goal of this work was to construct full length cDNA of RGMoV from infected plant material. In the course of this work the complete genome of Ryegrass mottle virus (RGMoV) was sequenced. Sequence analysis revealed several important differences if compared with previously published. Identified nucleotide insertion and deletion in protease domain resulted in a new 52 AA sequence, containing AA residues for catalytic and substrate binding sites,

which were not found previously. Sequence analysis of the newly identified AA sequence revealed 38.5% to 50.9% similarities to other sobemovirus proteases. Additionally, the analysis of the new RGMoV genome sequence suggested the translation of RdRP via -1 ribosomal frameshifting mechanism. These results allow to classify RGMoV genome organization as CfMV-like, but not SCPMV-like, as reported earlier. These data contribute to the taxonomic concept that all sobemoviruses have the CfMV-like genome organization.
The second goal was obtaining of non-structural proteins in preparative quantities and characterizing them structurally and functionally. Protein coding sequences from RGMoV cDNA were cloned in pET-, pACYC-, and pColdI-based plasmid vectors and expressed in several E.coli host cells. As revealed experimental data for protease and RdRp, most suitable combination was pColdI-derived plasmid vector and WK6 strain. VPg and P16 were expressed from pACYC- or pET-derived vectors and purified by affinity His-tag chromatography. Despite of numerous conditions tested all crystallization attempts of VPg and P16 were unsuccessful, possibly due to non-structured amino acid segments in structures of both proteins. As predicted by different bioinformatic methods, the entire VPg domain and P16 C-terminal end have putatively disordered structure, which prevent formation of crystals.
The third goal was crystallization of RGMoV. As crystallization material was used from plants isolated and purified RGMoV. The crystal structure has been determined at 2.9 Å. As revealed obtained data RGMoV posses T=3 icosahedral symmetry similar to other crystallized sobemoviruses. The coat protein has a canonical jellyroll β-sandwich fold. The structural differences from other sobemoviruses were found in GH loop and FG loop, where E1 and C helices were missing. The reduced size of the GH loop in CP results in the change of the relative subunit position comparison to other sobemoviruses. The FG loop participates in stabilization of β-annulus. The loss of interaction between FG loop and the β-annulus has been compensated for by additional interactions between the N-terminal arms. This difference resulted in 3% smaller diameter of RGMoV than for other sobemoviruses. This changes results in 7% smaller internal volume of capsid than the other representatives of this group, which is important for nucleic acid incapsidation. Nucleic acid density is approximately 480 mg/ml, which is similar to the rest of the sobemoviruses and tobacco necrosis virus. There was no density identified for calcium ions in the proximity of the conserved residues involved in calcium binding. The reason could be the low pH (3.0) of the crystallization buffer in which the groups interacting with the calcium ions are not charged.

CONTENTS

ABBREVIATIONS

AA – amino acid
AMV - *Alfalfa mosaic virus*
bp – base pair
BMYV - *Beet mild yellowing virus*
BSA - Bovine serum albumin
BSSV - *Blueberry shoestring virus*
BYDV - *Barley yellow dwarf virus*
BWYV - *Beet western yellows virus*
CABYV - *Cucurbit aphid-borne yellows virus*
CPMV - *Cowpea mosaic virus*
CYDV-RPV - *Cereal yellow dwarf virus-RPV*
cDNA – complimentary DNA
CfMV – *Cocksfoot mottle virus*
CfMV-NO - Norwegian isolate of CfMV
CfMV-RU - Russian isolate of CfMV
CP – coat protein
diRNA – defective interfiering RNA
DNA – desoxiribonucleinacid
dsDNA – dable stranded DNA
EGFP – enhanced green fluorescent protein
eIF4E - translation initiation factor 4E
GLV - *Giardia lamblia virus*
HAst-1 - *Human astrovirus serotype-1*
HIV - *Human immunodeficiency virus*
HIV- 1 - *Human immunodeficiency virus type 1*
HTLV-II - *Human T-cell leukemia virus type II*
ICTV - International Committee on Taxonomy of Viruses
IFs - initiation factors
IPTG – isopropyl-β-D-thiogalactopyranoside
IRES – internal ribosome binding site
kb - kilobase
kDa – kilodalton
LTSV - *Lucerne transient streak virus*
MBV - *Mushroom bacilliform virus*
MP - movement proteins
mRNA – messenger RNA
NLS - nuclear localization signals
NMR - nuclear magnetic resonance
Nt – nucleotide
ORF – open reading frame
PAGE – polyacrylamide gel
PEMV- 1 - *Pea enation mosaic virus 1*
PLRV - *Potato leafroll virus*
Pn – C-terminal part of polyprotein 2a
Poly(A) - polyadenylic acid
PPV - *Plum pox potyvirus*
−1 PRF - −1 programmed ribosomal frameshifting
PTGS - posttranscriptional gene silencing
P1 – sobemovirus ORF1 coded protein
P8 – SeMV ORF2a C-terminal coded protein domain
P10 – SeMV ORF2a C-terminal coded protein domain
P16 – RGMoV encoded ORF2a protein with molecular mass 16.6 kDa

P2a – polyprotein coded by ORF2a
P18 – SeMV ORF2a coded protein
P27 – CfMV ORF2a C-terminal coded protein
RCNMV - *Red clover necrotic mosaic virus*
RdRp – RNA dependent RNA polymerase
RGMoV – *Ryegrass mottle virus*
RFS – ribosomal frameshifting signal
RNA – ribonucleic acid
RNPC - ribonucleoprotein complex
RT-PCR – reverse transcriptase polymerase chain reaction
RT - reverse transcription
RuCMV - *Rubus chlorotic mottle virus*
RYMV – *Rice yellow mottle virus*
RYMV-CI - Ivory Coast isolate of RYMV
RYMV-Nig - Nigerian isolate of RYMV
SARS - *Severe acute respiratory syndrome*
satRNA - satellite RNA
SCMoV - *Subterranean clover mottle virus*
SeMV - *Sesbania mosaic virus*
SBMV – *Southern bean mosaic virus*
SBMV-ALM – Almerian isolate of SBMV
SBMV-Mor – Morocco isolate of SBMV
SBMV-Ark - Arkansas isolate of SBMV
SCPMV - *Southern cowpea mosaic virus*
SDS – Sodium dodecyl sulfate
SDS-PAGE - Sodium dodecyl sulfate polyacrylamide gel
sgRNA – subgenomic RNA
siRNAs - small interfering RNAs
SIV - *Simian immunodeficiency virus*
SNMoV - *Solanum nodiflorum mottle virus*
SoMV - *Sowbane mosaic virus*
ssRNA – single-stranded RNA
ss – single stranded
UTR – untranslated region
TNV - *Tobacco necrosis virus*
tRNA – transport RNA
TRoV - *Turnip rosette virus*
TuMV - *Turnip mosaic virus*
TVMV - *Tobacco vein mottling virus*
VPg – viral protein, genome-linked
VTMoV - *Velvet tobacco mottle virus*
WGE - wheat germ extract
wt – wild type

Amino acids may be described by one or three letters.

INTRODUCTION

Plant viruses cause significant losses in world agriculture in tropical and subtropical areas, as well as in the temperate zones. In addition, modern agricultural techniques, such as monoculture, have contributed to significant increase of emerging plant viruses, particularly the geminiviruses, closteroviruses and tospoviruses. For example, plant viruses of Begomovirus group cause millions of tonnes of crop losses worldwide. Although viruses are simply biological objects, which are studied more than 100 years, their penetration and reproduction in host plant cells, and plant-defense mechanisms against virus infections are still not completely understood. These facts demonstrate the importance of plant virus research leading to the introduction of new agricultural technologies, which allow enhancing the output of agricultural production. Virus-resistant transgenic plants are probably the best examples of successful exploitation of basic virology studies in the agriculture.
The other important area for potential use of plant viruses is biotechnology application. First of all, as shown in studies of recent years, plant viruses can be converted into gene vectors, able to replicate and produce recombinant proteins in plants. With the help of these vectors such proteins of medical interest as enzymes, vaccines, antibodies and other active components can be synthesized. To construct efficient virus-based vector corresponding genomes and their encoded proteins have to be characterized functionally and structurally. Additionally, plant viruses can serve as a model in understanding the life-cycle of animal and human viruses.
Sobemovirus genomes are relatively well characterized, but their encoded protein structures and, in particular their role and interactions in the infection process remain to be cleared up. To characterize the proteins, it is necessary to clone the corresponding genes, find out appropriate expression system, purify the proteins in large-scale quantities, find out their properties and, after successful crystallization experiments, to solve their 3D structure.

The aim of this work was to re-evaluate the genome organization *Ryegrass Mottle Virus* and characterize virus-encoded proteins biochemically and structurally.

Accordingly to the aim of the work several work tasks were made:

1. Construct full-sized RGMoV cDNA from virus RNA isolated from infected plant material, re-sequence the RGMoV genome;

2. Clone, express and purify all RGMoV-encoded proteins and their domains from bacterial expression system;

3. Characterize the RGMoV proteins using bioinformatical and biochemical methods;

4. Crystallize and solve 3D structures of RGMoV proteins.

REVIEW OF THE LITERATURE

Description of *Sobemovirus* genus

Sobemoviruses are single stranded (ss) messenger-sense RNA (ssRNA) plant viruses with monopartite genome; they are named after first detected specie ***So****uthern* ***be****an* ***mo****saic* ***virus*** (SBMV). As separate genus it was recognized only in 1995 by the International Committee on Taxonomy of Viruses (ICTV) (Hull, 1995). Befor it was identified as a *Southern bean mosaic virus* group consisting of single-component-RNA beetle transmitted plant viruses (Walters, 1969). Later there were classified according to protein subunit molecular weight, capsid stabilization, sedimentation coefficient, and distribution of particles within the cell (Hull, 1977a). As the last step before identification as genus was acteptance of the name *Sobemovirus* by ICTV (Matthews, 1982).

In 1982 those viruses were accepted by the International Committee on Taxonomy of Viruses (ICTV) under the name *Sobemovirus* (Matthews, 1982). In 1995 the ICTV recognized the group as an unassigned genus, *Sobemovirus* (Hull, 1995). Today in this genus are 14 definitive species, four tentative species, and several not jet classified species (http://www.ncbi.nlm.nih.gov/ICTVdb/Ictv/fs_sobem.htm).

Virus particles of genus *Sobemovirus* are about 30 nm in diameter (Hull, 1995). The virions consist of a single coat protein (CP) of approximate size 30 kDa. A capsid is arranged in to icosahedral T=3 quasi-symmetry by 180 CP molecules. It 60 icosahedral asymmetric units are made of three chemically identical copies of the CP - A, B and C differ only bay conformation (Rossman et al., 1989). In each virion is encapsided one genomic RNA (genRNA) (4 to 4.5 kb in size), and one subgenomic RNA (sgRNA) molecule (Tamm and Truve, 2000b).

Together with genRNA and sgRNA, some members of *Sobemovirus* group encapsidate a viroid-like satellite RNA (satRNA). Encapsidation of satRNA is depending on a helper virus for replication. The viroid-like satRNAs has been reported for *Lucerne transient streak virus* (LTSV), *Rice yellow mottle virus* (RYMV), *Subterranean clover mottle virus* (SCMoV), *Solanum nodiflorum mottle virus* (SNMoV), and *Velvet tobacco mottle virus* (VTMoV) (Francki *et al.*, 1983; Sehgal *et al.*, 1993; Collins *et al.*, 1998)(Gould and Hatta, 1981; Randles et al., 1981; Francki et al., 1986)(Tien-Po *et al.*, 1981; Jones and Mayo, 1983; Jones and Mayo, 1984). The sizes of satRNAs range from 220 to 390 nucleotides (Nt) (Haseloff and Symons, 1982; Keese *et al.*, 1983; AbouHaidar and Paliwal, 1988; Davies *et al.*, 1990; Collins *et al.*, 1998). RYMV satRNA with 220 nt is the smallest naturally occurring viroid-like RNA known today (Collins *et al.*, 1998). Additionally, it is reported that sobemoviruses can encapside defective interfering RNAs (diRNA). Such diRNA are reported only case of *Cocksfoot mottle virus* (CfMV) (Mäkinen *et al.*, 2000).

Sobemovirus at 5’ end of genRNA contains covalently linked uridinylated genome-linked protein (VPg), and 3’ end lacks poly(A) tail (Tamm and Truve, 2000b).

The 5'-terminal half of the sobemovirus genome is similar to the poleroviruses and enamoviruses with the successive functional domains of a serine protease-like, a VPg and a RdRp. Phylogenetical analysis of the sobemoviruses polymerase sequences demonstrated relation to the poleroviruses and enamoviruses. But the CP analysis of the sobemoviruses - encoded by the 3'terminal half of the genome – revealed sequences and structural similarities with the genus Necrovirus of the family *Tombusviridae*. *Mushroom bacilliform virus* (MBV), the unique species of the family *Barnaviridae*, has a genomic organization similar to that of the sobemoviruses and showed sequence identities with the polymerase and the CP. Genomic organization of animal viruses of the family *Astroviridae* also shows similarities to that of

sobemoviruses, but the sequence identities are more remote (http://www.ncbi.nlm.nih.gov/ICTVdb/Ictv/fs_sobem.htm).

Biological properties of *Sobemoviruses*

Sobemoviruses infection from plant to plant is transmitted mostly by mechanical injuries. In study to find the transmission agent of RYMV infection it was demonstrated that highest and widest virus transmission causes such vertebrates as donkies, cattle and grass rats, and not insects as suggested earlier,. Infection was also transmitted by wind-mediated leaf contacts and via soil (Sarra, 2005). In SCMoV case was established that its infection can be transmitted by grazing and trampling by livestock, crushing under vehicle wheels and by mowing for hay production, i.e. contact-transmission (Jones, 2004). Similar situation is with CfMV, which is transmitted by folder harvesters (Rognli *et al.*, 1995).

Based on investigations of vector transmission of *Sobemovirus* was identified that insects take part in virus transmission to the next plant. As most common transmitters were detected leave eating beetles, where transmission is occurred in a semi-persistent manner. Sobemoviruses could be transmitted by aphids, leaf miners, leafhoppers, or myriads, too. There are reports that some members of sobemoviruses are seed-transmissible (Table 1).

Sobemovirus members CfMV, RYMV, SNMoV, SBMV, *Southern cowpea mosaic virus* (SCPMV), and *Turnip rosette virus* (TRoV) are transmitted by beetles. *Blueberry shoestring virus* (BSSV) is transmitted by aphids, and VTMoV is transmitted by myriads. *Sowbane mosaic virus* (SoMV) is transmitted by leafminers and leafhoppers (Tamm and Truve, 2000b).

The natural host range of each virus species is relatively narrow. Sobemoviruses infect plant species from approximately 15 different families, int. al., dicotyledonous and monocotyledonous. SBMV was isolated in Louisiana and California and reported as the first Sobemovirus member (Zaumeyer and Harter, 1943). Later discoveries reported that sobemoviruses were distributed all over the world, infecting plants from Scandinavia to New Zealand and throughout tropical Africa (Tamm and Truve, 2000b).

Table 1. Viruses of the genus *Sobemovirus* and their biological properties (Sõmera, 2010).

Virus	Abbr.	Distribution	Natural host	Insect vector	Transmission	
					Mechanical	Seed
Definitive species						
Blueberry shoestring virus	BSSV	USA (Maine, Michigan, New Jersey, Oregon, Virginia, Washington), Canada (New Brunswick, Nova Scotia, Ontario, Quebec)	*Vaccinium corymbosum, V. angustifolium*	*Illinoia pepperi* (aphid)	Yes	No
Cocksfoot mottle virus	CfMV	Europe (Denmark, France, Germany, Norway, Russia, UK), New Zealand, Japan	*Chionochloa rubra, Dactylis glomerata, Festuca novae-zelandiae, Poa anceps, P. cita, Triticum aestivum*	*Lema melanopus, L. lichensis* (beetles)	Yes	No
Lucerne transient streak virus	LTSV	Australia (Victoria, Tasmania), New Zealand, Canada	*Medicago sativa*	ND	Yes	No
Rice yellow mottle virus	RYMV	Africa (Benin, Burkina Faso, Cameroon, Chad, Cote d'Ivoire, Gambia, Ghana, Guinea, Guinea Bissau, Kenya, Liberia, Madagascar, Mali, Malawi, Mauritania, Mozambique, Niger, Nigeria, Rwanda, Senegal, Sierra Leone, Tanzania, Togo, Uganda)	*Echinochloa colona, Ischameum rugosum, Oryza barthii, O. glaberrima, O. longistaminata, O. sativa, Panicum repens*	*Chaetocnema* spp., *Dicladispa gestroi, Dactylispa spp., Monomolepta spp., Oulema dunbrodiensis, Sesselia pusilla, Trichispa sericea* (beetles), *Conocephalus merumontanus* (grasshopper)	Yes	No
Ryegrass mottle virus	RGMoV	Japan, Germany	*Dactylis glomerata, Lolium multiflorum*	ND	Yes	No
Sesbania mosaic virus	SeMV	India (Andra Pradesh)	*Sesbania grandiflora*	ND	Yes	ND
Solanum nodiflorum mottle virus	SNMoV	Australia (Queensland, New South Wales)	*Solanum nodiflorum, S. nitidibaccatum, S. nigrum*	*Epilachna sparsa, E. doryca australica, E. guttatopustulata, Psylliodes* sp. (beetles), *Cyrtopeltis nicotianae* (mirid)	Yes	No
Southern bean mosaic virus	SBMV	USA (Arkansas, California, Louisiana, Maryland), Central and	*Phaseolus vulgaris*	*Ceratoma trifurcata, Epilachna variestis,*	Yes	Yes

		South America (Brazil, Colombia, Mexico), Africa (Ivory Coast, Morocco), Europe (France, Lithuania, Spain), Asia (Iran)		*Diabrotica undecimpunctata howardii* (beetles)		
Southern cowpea mosaic virus	SCPMV	USA (Wisconsin), Africa (Botswana, Ghana, Kenya, Nigeria, Senegal, Togo), Asia (India, Pakistan)	*Cassia toria, Pisum sativum, Vigna unguiculata*	*Ootheca mutabilis, Madurasia obscurella* (beetles)	Yes	Yes
Sowbane mosaic virus	SoMV	USA (California), Canada, Central and South America, Europe (Bulgaria, Czechia/Slovakia, Croatia, France, Greece, Hungary, Italy, Moldavia, Netherlands), Japan, Australia (Queensland, New South Wales, Victoria, Tasmania)	*Chenopodium* spp., *Atriplex subrecta, Spinacia oleracea, Vitis* sp., *Prunus domestica, Alisma plantago-aquatica, Danae racemosa*	*Myzus persicae* (aphid), *Liriomyza langei* (leafminer), *Circulifer tenellus* (leafhopper), *Cyrtopeltis nicotianae*, *Halticus citri* (mirids)	Yes	Yes
Subterranean clover mottle virus	SCMoV	Australia (New South Wales, South Australia, Tasmania, Victoria, Western Australia)	*Trifolium subterraneum, Trifolium* spp.	Not found	Yes	Yes
Turnip rosette virus	TRoV	UK (England, Scotland), Switzerland	*Brassica campestris, B. nigra*	*Phyllotreta nemorium* (beetle)	Yes	ND
Velvet tobacco mottle virus	VTMoV	Australia (Northern Territory, Queensland, South Australia)	*Nicotiana velutina*	*Cyrtopeltis nicotianae* (mirid), *Epilachna* spp. (beetle)	Yes	No
Tentative species						
Cocksfoot mild mosaic virus	CMMV	Europe (Czechia/Slovakia, Denmark, France, Germany, Norway, UK), Canada (Ontario), New Zealand	*Agrostis stolonifera, Bromus mollis, B. diandrus, Cynosurus cristatus, Dactylis glomerata, Festuca pratensis, Lolium perenne, L. perenne* x *L. multiflorum, Phleum pratense, Poa trivialis, Triticum aestivum*	*Myzus persicae* (aphid), *Lema melanopus* (beetle)	Yes	No
Cynosurus mottle virus	CnMoV	Europe (Germany, UK, Ireland), New Zealand	*Agrostis tenuis, A. stolonifera, Cynosurus cristatus, Lolium perenne, L. perenne,* x *L. multiflorum*	*Lema melanopus* (beetle), *Rhopalosiphum padi* (aphid)	Yes	ND
Ginger chlorotic fleck virus	GCFV	Africa (Mauritius), Asia (India, Malaysia, Thailand)	*Zingiber officinale*	Not found	Yes	ND

Rottboellia yellow mottle virus	RoMoV	Nigeria	*Rottboellia cochinchinensis*	ND	Yes	ND
Species not yet recognized by ICTV						
Calopo yellow mosaic virus	CAYMV	Colombia	*Calopogonium mucunoides*	*Diabrotica balteata* (beetle)	Yes	Yes
Imperata yellow mottle virus	IYMV	Burkina Faso	*Imperata cylindrica, Zea mays*	ND	Yes	No
Papaya lethal yellowing virus	PLYV	Brasil (Pernambuco, Bahia, Rio Grande do Norte, Ceará, Paraíba)	*Carica papaya*	Not found	Yes	No
Rubus chlorotic mottle virus	RuCMV	UK (Scotland)	*Rubus* sp	ND	Yes	ND
Snake melon asteroid mosaic virus	SMAMV	Sudan	*Cucumis melo* var *flexuosus*	Not found	Yes	Yes

Genome organization

The complete nucleotide sequences of 10 sobemoviruses have been reported. The *Sobemovirus* genome is composed of four ORFs, and most of the predicted ORFs overlap (Fig. 1).

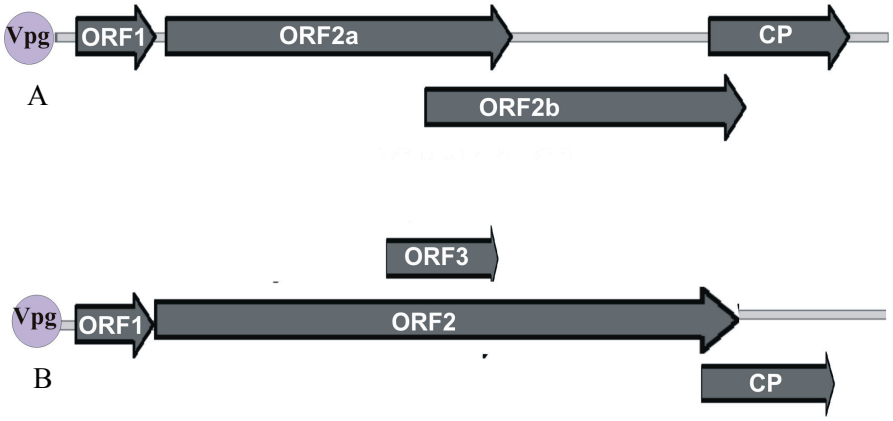

Figure 1. Schematic representation of Sobemovirus genomic organizations based on RGMoV. A - CfMV-like; B - SCPMV-like.

Sobemoviruses contain a small ORF1 encoding P1, at the 5' end of the genome and a 3'-proximal ORF3 which encodes the viral CP. Early analysis of LTSV genome suggested that it was exception (Jeffries *et al.*, 1995) because of two small ORFs instead of one at the 5' end - ORF1a and ORF1b. Later sequence analysis revealed that it was sequencing error and LTSV, like other sobemoviruses, contains a single ORF1.

At the beginning based on the organizational differences in the central part of the genome (encoding the virus polyprotein), the sobemoviruses were subdivided into SCPMV-like and CfMV-like viruses. The polyprotein of SCPMV is encoded by the large continuous ORF2. The genome of SCPMV also contains an internal coding region, ORF3, situated in the -1 RFS within ORF2. Instead, CfMV has two overlapping ORFs, ORF2a and ORF2b, and the polyproteins are expressed through a -1 RFS mechanism. As in LTDV case genome of SBMV (Othman and Hull, 1995) revealed different organization. It lacked the small ORF3 characteristic of SCPMV and SBMV-Ark, and the three remaining ORFs do not overlapped. In the other case SBMV-Ark contained necessary four putative overlapping ORFs as in SCPMV case but there was no similarities in amino acid sequences putative proteins (Lee and Anderson, 1998). Those differences were resulted due to mutations or sequencing errors in SBMV (Lee and Anderson, 1998).

10 sequenced *Sobemoviral* genomes available in the GenBank was grouped by genome organization as follows: SCPMV-like - SCPMV (Wu *et al.*, 1987), SBMV (Othman and Hull, 1995; Lee and Anderson, 1998), *Sesbania mosaic virus* (SeMV) (Lokesh *et al.*, 2001), *Ryegrass mottle virus* (RGMoV) (Zhang *et al.*, 2001b) and LTSV (Jeffries *et al.*, 1995), in other CfMV-like - CfMV (Makinen *et al.*, 1995; Ryabov *et al.*, 1996; Zhang and Toriyama, 2001), RYMV (Fargette *et al.*, 2004), SCMoV (Dwyer *et al.*, 2003) and TRoV (Callaway *et al.*, 2002). RYMV has been suggested to have the SCPMV-like genomic organization (Yassi *et al.*, 1994). In 2004, Fargette and his colleagues (Fargette *et al.*, 2004) sequenced 14 new natural isolates of RYMV that had the CfMV-like genomic organization. The error in genome organization was caused by extra nucleotide T in position 2244, which was published in earlier sequences of RYMV-Nig and

RYMV-CI (Yassi *et al.*, 1994; Pinto and Baulcombe, 1995). The same situation was observed in to novel SBMV isolate partial sequences (Verhoeven *et al.*, 2003; Segundo *et al.*, 2004). The alignment of these isolates SBMV-Almerian (SBMV-ALM) and SBMV-Morocco (SBMV-Mor) to previously published SBMV-Ark revealed GG instead of GGG at position 2176–2178 the removal of extra nucleotide resembled in CfMV-like genome organization. These findings raised the question about genomic organization of all SCPMV-like sobemoviruses. Resequencing of SBMV-Col revealed GG instead of GGG at position 2176–2178 like in SBMV-ALM and SBMV-Mor when compared with the SBMV-Ark sequence (Lee and Anderson, 1998). In new SCPMV-WI sequence G2232 was missing when compared with the published SCPMV sequence (Wu *et al.*, 1987). In LTSV-Can case extra nucleotides GC after position 2299 were identified in comparison of LTSV sequence available from GenBank. Re-sequencing of RGMoV revealed the GGGGGG instead of GGGGGGG at position 2203–2209 when compared with the RGMoV sequence published before (Zhang *et al.*, 2001b). Re-sequencing of SeMV revealed the presence of C instead of CC at position 2177– 2178 in the previously published sequence (Lokesh *et al.*, 2006). A CfMV-like genomic organization is common to all *Sobemoviruses* (Meier and Truve, 2007). During our publication revision appeared (Meier and Truve, 2007) publication, we discovered the same sequence error which resembled as RdRP ORF separation from ORF2, disappearance of small ORF3 with unknown protein, and RGMoV genome organization changed from SCPMV- to CfMV-like.

The CfMV-type genome organization schema is similar to the MBV (genus *Barnavirus*, family *Barnaviridae*) (Revill *et al.*, 1994). This similarity is based on the presence of proximal ORFs at 5' and 3', as well as two overlapping ORFs for polyprotein expression. CfMV genome ORFs arrangement in the 5' half of the genome is similar to the genus *Polerovirus* (formerly known as subgroup II luteoviruses — *Potato leafroll virus* (PLRV; (van der Wilk *et al.*, 1989), *Beet western yellows virus* (BWYV; (Veidt *et al.*, 1988), *Beet mild yellowing virus* (BMYV; (Guilley *et al.*, 1995), *Cereal yellow dwarf virus-RPV* (CYDV-RPV, formerly BYDV-RPV; (Vincent *et al.*, 1991)), and *Cucurbit aphid-borne yellows virus* (CABYV; (Guilley *et al.*, 1994) and the genus *Enamovirus* (*Pea enation mosaic virus 1* (PEMV- 1; formerly RNA1 of PEMV; (Demler and de Zoeten, 1991) in the family *Luteoviridae*. The 5' gene cluster of these viruses contains a small ORF0 and overlapping ORFs 1 and 2, and the polyprotein is expressed as a translational frameshift fusion of the ORF1 and 2 products (Tamm and Truve, 2000b).

Virus genome replication

All positive-sense RNA viruses could be divided in to two groups by their RNA synthesis initiation mechanisms. To the first group belong polymerases which carry out primer-independent (*de novo*) synthesis. The second-group polymerases perform primer-dependent initiation of synthesis (van Dijk *et al.*, 2004; Ferrer-Orta *et al.*, 2006). Majority of positive-sense RNA virus polymerases have to be classified as first-group enzymes, where the initiation is supported by a single nucleotide a 3 hydroxyl group. Some of the animal (*Picornaviridae*) and plant (*Comoviridae*, *Potyviridae*) virus families and a few species from *Luteoviridae* use viral protein VPg as a primer for initiating the reaction (Buck, 1996)). Andino *et al*. (1993) proposed that uridinylated VPg (VPg-pU-pU) serves as a primer for 3D replicase to initiate poliovirus RNA synthesis. Poliovirus replication involves the uridylylation of VPg by RdRp in the presence of an internal *cre* element of viral RNA. Uridylylated VPg acts as a primer and the complementary strand is elongated by RdRp (Morasco *et al.*, 2003; Murray and Barton, 2003).

The mechanisms of *Sobemovirus* replication are poorly understood. Until now all RNA multiplication concepts were based on poliovirus replication principle. All members of *Sobemovirus* group possess VPg at the genRNA 5'-end. Recent studies *in vitro* with SeMV recombinant RdRp showed that VPg is not necessary for genRNA replication by RdRP. This study demonstrated clearly that *Sobemovirus* RdRp could be classified as primer-independent polymerase (Govind and Savithri, 2010).
VPg is not the only component of many agents involved in positive RNA plant virus genome replication processes. From studies of other positive-sense RNA viruses it is known that crucial role plays virus genRNA untranslated regions (UTR) at 5'- and 3'-ends which contain specific secondary structures for virus RdRp template recognition and replication initiation. Until now there were no strict proofs of *Sobemovirus* 3'-end role in RNA replication. As Govind and Savithri (2010) demonstrated in *in vitro* experiments with SeMV genRNA, the deletion at 3'-UTR end caused the reduction of synthesized RNA by 80 %. It was shown that a stem–loop structure at the end of UTR is important for RdRP to synthesize genRNA. The 3′ UTR sequences of sobemoviruses are highly diverse and it is difficult to predict a common, conserved secondary structural element which may function in the initiation of RNA synthesis (Govind and Savithri, 2010).

Protein translation and processing mechanisms

The basic mechanism of translation is the precise decoding of the triplet-codon sequences from mRNA located in to one reading frame. The translational strategy of sobemoviruses is based on production of a large polyproteins from the genRNA and synthesis of the CP from a sgRNA (Matthews, 1991). Specific signals located into the mRNA sequences can cause deviations from basic translation rule. Viruses can use several translational mechanisms: translational hopping, stop-codon read-through and programmed ribosomal frameshifting (PRF) (Farabaugh, 1996; Gesteland and Atkins, 1996), for regulating the amount of proteins produced from their polyproteins. For positive-stranded RNA viruses the most common mechanism is -1 PRF. -1 PRF signals of large group of viruses are regulated by replication-associated proteins. For *Sobemovirus* CfMV it could be P27, as far known from experiments in heterologous hosts (Mäkeläinen, 2006). The efficiency of -1 PRF can be variable from 1% to 40% depending on virus (Brierley, 1995), and changes in the efficiency can inhibit virus assembly and replication (Dinman and Wickner, 1992; Hung *et al.*, 1998; Barry and Miller, 2002). There is little known about the *trans*-acting factors and the biophysical parameters affecting the -1 PRF efficiencies. Two necessary *cis*-acting signals are involved in slippage: a slippery heptamer X XXY YYZ and a downstream located element of RNA secondary structure (Jacks *et al.*, 1988). -1 PRF takes place after the accommodation step in the slippery sequence by simultaneous slippage of both tRNAs into the overlapping -1 frame XXX YYY (Jacks *et al.*, 1988; Dinman *et al.*, 1997). The sequence of the heptamer allows post-slippage base-pairing between the non-wobble bases of the tRNAs and the new -1 frame codons of the mRNA. Downstream RNA secondary structures (Giedroc *et al.*, 2000) force the ribosome to pause, and place the ribosomal A- and P-sites correctly over the slippery sequence (Kontos *et al.*, 2001). However, the pausing of the ribosome is not sufficient for -1 PRF to occur (Tu *et al.*, 1992); in fact, the duration of the halt does not necessarily correlate with the level of the -1 PRF observed (Kontos *et al.*, 2001). Crystallographic, molecular, biochemical and genetic studies suggest that a pseudoknot restricts the movement of the mRNA during the tRNA accommodation step of elongation by filling the entrance of the ribosomal mRNA tunnel (Plant *et al.*, 2003). This restriction can be eased either

by unwinding the pseudoknot, which allows the mRNA to move forward or by a slippage of the mRNA one nucleotide backwards. The parameters that affect efficiency of -1 PRF are the sequence of the slippery heptamer, the downstream secondary structure, and the length and sequence of the spacer between the two *cis*-acting signals. Efficiency of -1 PRF can affect up and downstream sequences like termination codons, those sequences can be located several kb away from it (Brierley, 1995; Honda and Nishimura, 1996; Lucchesi *et al.*, 2000; Kim *et al.*, 2001; Paul *et al.*, 2001; Dinman *et al.*, 2002; Dulude *et al.*, 2002). For *Barley yellow dwarf virus* (BYDV) a special sequence 30 UTR, 4 kb downstream from the slippage site, is vital for -1 PRF (Paul *et al.*, 2001; Barry and Miller, 2002). Similar situation was observed for *Human immunodeficiency virus* (HIV) it requires a more complex secondary structure than a simple stem– loop for optimal -1 PRF *in vivo* (Dinman *et al.*, 2002; Dulude *et al.*, 2002). These investigations suggest that -1 PRF studies carried out with minimal frameshift signals may lead to inaccurate estimates of the stoichiometry of synthesized viral protein products during infection (Makelainen and Makinen, 2005).

Cis-acting elements for Sobemoviruses

A seven Nt slippery sequence is the first *cis*-acting element for -1 PRF. It is the place where the ribosome will slip back by one Nt. The sequence motif for the slip-site is XXXYYYZ, where X can be any Nt, Y either A or U, and Z can be any Nt except G (Brierley *et al.*, 1992). Second element is the sequence that is forming a downstream secondary structure. Those elements can be divided into three groups. The first and the best studied of them are hairpin-type RNA pseudoknot structures formed when the Nt of hairpin loop base pair with single stranded downstream complementary Nt. The second group of structures consists of pseudoknots with an unusual structure. For example, a three stemmed pseudoknot structure has been reported for *Severe acute respiratory syndrome* (SARS) coronavirus (Baranov *et al.*, 2005; Plant *et al.*, 2005; Su *et al.*, 2005; Brierley and Dos Ramos, 2006). This structure is composed of two double-stranded stems connected by a single-stranded loop and a second loop which itself folds into a stem-loop. The stimulatory RNA of the *Visna-Maedi virus* −1 RFS signal is another unusual pseudoknot with a seven Nt interstem element between two stems (Pennell *et al.*, 2008). The third group includes stable stem-loop structures sufficient to promote efficient −1 PRF. Examples include the frameshift-promoting elements at the gag-pol junction in *Giardia lamblia virus* (GLV), *Human immunodeficiency virus type 1* (HIV- 1), *Simian immunodeficiency virus* (SIV) and related lentiviruses (Li *et al.*, 2001; Staple and Butcher, 2005; Brierley and Dos Ramos, 2006; Marcheschi *et al.*, 2007). A simple hairpin loop in frameshifting has been demonstrated in *Human astrovirus serotype-1* (HAst-1), *Human T-cell leukemia virus type II* (HTLV-II) and *Red clover necrotic mosaic virus RNA-1* (RCNMV) (Falk *et al.*, 1993; Kim and Lommel, 1994; Marczinke *et al.*, 1994; Kim and Lommel, 1998).

CfMV is the first sobemovirus for whom the secondary structure of the stem-loop regulating −1 PRF has been verified experimentally (Lucchesi *et al.*, 2000; Tamm *et al.*, 2009). The conclusion was made that sobemoviruses cluster together with lentiviruses and few other viruses (examples GLV, HAst-1, HTLV-II and RCNMV) forming an unique group of viruses that require stable stem-loops for frameshifting. Marcheschi et al. (2007) hypothesized that the nature of the RNA structure of the frameshift site is dependent upon the slippery sequence to which it is coupled.

The optimal frameshift efficiency required for the specific virus is guaranteed with the combination of the slippery sequence, spacer region length and secondary structure element (Tamm *et al.*, 1999).

An experiment has been carried out to examine the effect of different parts of the CfMV genome on -1 frameshifting efficiency. Lucchesi et al., (2000) had shown that CfMV -1 ribosomal frameshifting was directed by the U-UUA-AAC slippery heptanucleotide and by the putative stemloop structure starting seven Nt downstream. Mutation studies on the slippery sequence revealed that mutating the repetitive nature of sets of -1-frame triplets abolishes frameshifting. Mutation in the sequence U-UUA-AAC to U-UUA-AGC affecting the second set of triplets no frameshift product could be observed in *in vitro* translation assays. The heptameric sequence is often followed by an RNA secondary structure motif, either a pseudoknot or stemloop, at an average of six Nt downstream. In CfMV the secondary structure is predicted to be a stem-loop beginning at the eighth Nt downstream from the slippery sequence. The signal for CfMV stem-loop is located at positions 1648-1676. Mutation in positions 1649-1985 deleting the stem-loop structure abolished frameshifting completely. Deletions at nucleotides 1691-1985 and nucleotides 1691-2132 leaving the stem-loop region intact had no effect on frameshifting efficiency. The *in vitro* data indicate that the stem-loop region independently of downstream flanking sequences stimulates frameshifting. Supportive evidence for this came from experiments with a 70 Nt CfMV sequence containing the consensus sequences for efficient frameshifting inserted in the middle of the *uidA* gene. Translations with the construct gave similar efficiencies as the CfMV polyprotein encoding construct. The 70 Nt can thus be considered to be a minimal frameshift sequence of CfMV for efficient frameshifting *in vitro*. This 70 Nt sequence, containing both the slippery sequence and the predicted downstream secondary structure, was shown to direct PRF in wheat germ extract (WGE) with an efficiency of 12.7±1.4%. Similar *in vitro* frameshifting efficiency, 10.6±1.4%, was determined for the entire ORF2a–2b encoding region. The minimal frameshift sequence required for efficient *in vitro* frameshifting was mapped at the nucleotides 1621–1690 (Lucchesi *et al.*, 2000).

Polyprotein processing

Polyprotein processing is a major strategy used by many viruses to generate many functional protein products from a single ORF (Wellink and van Kammen, 1988). The viral protease domains present within the polyprotein play a major role in this maturation process. These proteases have unique specificities and belong to serine-, cysteine-, serine-like cysteine-, or aspartic-class of proteases (Dougherty and Semler, 1993).The NIa protease of *Tobacco etch virus* (TEV) is a serine-like cysteine protease which cleaves its polyprotein at the C-terminal side of specific glutamine residues both in *cis* and in *trans* (Carrington et al., 1988). The presence of protease domains and there cleavage sites involved in processing for many viruses have been predicted by sequence analysis. There are only in a few cases where their functional role been confirmed by characterization of recombinant protease.

At the beginning of SCPMV-like ORF2 polyprotein processing of *Sobemoviruses* was proposed to be catalyzed by the N-terminal serine protease domain. It has been predicted that the luteovirus and sobemovirus polyproteins are cleaved at (Q,E)/(G,S,A) sites (Gorbalenya *et al.*, 1988). The finding that the N-terminal cleavage site of the CfMV VPg is between glutamic acid (E) and asparagine (N) was thus unexpected (Tamm and Truve, 2000b). The putative catalytic triad H181, D216, and S284 is located in the N-terminal part of serine protease. This triade is present in SCPMV and also conserved in SeMV and other members of *Sobemovirus* group (Lokesh *et al.*, 2001).

SeMV polyprotein expression in *E. coli* was shown to undergo proteolytic processing. Mutational analysis of those residues (H181, D216, and S284) showed that the protease was

responsible for polyprotein processing. Analysis of the cleavage site mutants of SCPMV-like SeMV cDNA construct confirmed the cleavage between protease-VPg and VPg-RdRP at E325–T326 and E402–T403 sites. An additional suboptimal cleavage site at E498–S499 was identified which carried out further processing of RdRp in to 10- and 52-kDa proteins. It's shoved that the protease had both E/T and E/S specificities (Satheshkumar *et al.*, 2004).

After re-sequencingn of SeMV genome and genome organization was revised. New cDNA clone of SeMV was made for polyprotein processing studies using proteins isolated from recombinant *E. coli*. The results confirmed the cleavage at the earlier identified sites E325-T326, E402-T403 and E498-S499 to release protease, VPg, P10 and P8. A novel cleavage site was identified within the protease domain at position E132-S133, which was found to be essential for efficient polyprotein processing. Products, corresponding to cleavages identified in *E. coli*, were also detected in infected Sesbania leaves. Identified sites are exactly the same in polyprotein 2a-2b, the cleavage was observed between Protease-VPg but not between VPg-RdRp (Nair and Savithri, 2010b). However, the exact mechanism of sobemoviral polyprotein processing is still unclear. *In vitro* translation experiments in either rabbit reticulocyte lysate or WGM have been carried out for six of the sobemoviruses demonstrating the presence of ∼100 , ∼70 , ∼30 (CP) and ∼18 kDa proteins (ORF1 product) (Tamm and Truve, 2000b). The ∼100 and ∼70 kDa were assigned to the two polyproteins. But limited polyprotein processing was observed with these *in vitro* translated products. The cleavage site between protease domain and VPg of CfMV poliprotein was identified using the N-terminal AA sequencing of VPg. Analysis of CfMV infected plant material revealed several cleavage products. Antibodies raised against polyprotein P2a detected a 12 kDa protein in infected leaf material, which was either not detected or was detected faintly with the anti-VPg antibody from the same samples; 18 - 23 kDa polypeptides. Antibodies from P2a and VPg recognized a 24 kDa protein, which represents the part of P2a from the beginning of VPg to the C-terminus of P2a. The 12 kDa protein, which appears to correspond to VPg, is a minor product in infected plants and can only be clearly detected by Western blot analysis of viral RNA-derived samples (Makinen *et al.*, 2000).

In a model suggested by Prufer et al., (1999), it was proposed that the hydrophobic N-terminal region of PLRV polyprotein P1 targets the protein to cellular membranes and that the basic nucleic acid-binding domain at its C-terminus interacts with the PLRV RNA. This membrane-bound complex was proposed to serve as a proteolytic processing site for VPg maturation (Prufer *et al.*, 1999). Some ideas of this model could be applied for the polyprotein processing of CfMV based on similarities in the polyprotein sequence and arrangement of sobemoviruses and poleroviruses. The 60 N-terminal AA of P2a contain hydrophobic residues and have been proposed to form a transmembrane domain in the CfMV polyprotein (Ryabov *et al.*, 1996). The C-terminal part of P2a contains a strong basic region (AA 539 - 552) in the P2a sequence according to (Makinen *et al.*, 1995) and may determine the RNA-binding property of P2a (Tamm and Truve, 2000b). No proteolytic processing of the CfMV polyprotein was observed in a cell-free translation system (Tamm *et al.*, 1999). This could be due to proteolytic processing of the CfMV polyprotein was required association with the replication complex and/or a proper cellular environment (Makinen *et al.*, 2000).

In SBMV, the N-terminal AA of VPg is threonine, which is preceded by a glutamic acid, indicating that an E/T processing site is used for VPg maturation (van der Wilk *et al.*, 1998).

Sobemovirus pathology

Sobemovirus infections can cause a variety of disease symptoms in plants: mild or severe chlorosis and mottling, stunting, necrotic lesions, vein clearing, and sterility. Some infections can be symptomless or cause death of plants.

In most cases sobemoviruses is detected at relatively high concentrations in infected plants. Largest virus concentration is described in mesophyll and vascular tissues, but are reports that Sobemoviruses had been identified in epidermal, guard, and bundle sheath cells (VIDE database, http://biology.anu.edu.au/Groups/MES/vide/). The localization of virus in vascular tissues as xylem parenchyma cells and xylem vessels seems to be characteristic, detection of virus particles in phloem parenchyma cells and sieve elements are reported only in the occasional cases (VIDE database, http://biology.anu.edu.au/Groups/MES/vide/). CfMV has been reported to localize in phloem of vascular tissues (Tamm and Truve, 2000b).

Particles of sobemoviruses had been detected in the cytoplasm and vacuoles of infected cells. Formation of crystalline arrays in the cytoplasm was reported for these viruses. In several sobemovirus infection case cells contain cytoplasmic fibrils, some of which are enveloped in endoplasmic reticulum-derived vesicles. For RYMV, BSSV, or SNMoV infected cells characteristic tubules, often aggregated into bundles were reported. The nature of these structures is unknown. No particles had been detected in either chloroplasts or mitochondria of cells infected with any of these viruses (Tamm and Truve, 2000b).

There are reports of particles localization in the cell nuclei of plants infected with sobemoviruses (Francki *et al.*, 1985). Yassi et al. (Yassi *et al.*, 1994) noted that the N-terminal part of the RYMV-CI CP (AA residues 3 to 22) contains a sequence which is identical to the bipartite nuclear targeting motif (Dingwall and Laskey, 1991). A similar motif can be found at the N-terminal part of all sobemoviral CP (Yassi *et al.*, 1994; Makinen *et al.*, 1995). This finding may explain particles observation in the nuclei of infected cells during sobemovirus infection. Except for this observation, no molecular determinants have been attributed to the subcellular or tissue-specific localization of sobemovirus particles. There is practically no information about the subcellular localization of the nonstructural sobemovirus proteins. In susceptible hosts, several sobemoviruses can cause severe diseases with significant economic losses. The main lost in host plants is caused by CfMV in Norwegian and United Kingdom cocksfoot varieties, for RYMV in African rice, and for SCMoV in Australian subterranean clover. Studies of hosts displaying total or partial resistance to the corresponding sobemoviruses gives some information for mechanisms involved in resistance. Natural resistance to sobemoviruses was reported for CfMV in cocksfoot (Catherall, 1987); 82), for SCMoV in subterranean clover (Wroth and Jones, 1992), for CnMoV in *Cynosurus cristatus* (Catherall, 1987), for RYMV in *Orza sativa* (Thottapilly and Rossel, 1993; Ghesquiere *et al.*, 1997; Ndjiondjop *et al.*, 1999), and in *Oryza glaberrima* (Attere and Fatokun, 1983; Thottapilly and Rossel, 1993; Paul *et al.*, 1995; Ndjiondjop *et al.*, 1999), and for SBMV in beans (Zaumeyer and Harter, 1943). For SCPMV, it has been demonstrated that resistance in cowpeas is controlled by a single gene according to a classical gene-to-gene interaction model (Hobbs *et al.*, 1987). Resistance to RYMV in several *O. glaberrima* cultivars and in *O. sativa indica* cultivar Gigante is determined by a single recessive gene (Attere and Fatokun, 1983; Thottapilly and Rossel, 1993; Paul *et al.*, 1995; Ndjiondjop *et al.*, 1999). Partial RYMV resistance in many *japonica* rice cultivars is a polygenic trait (Ghesquiere *et al.*, 1997; Albar *et al.*, 1998; Pressoir, 1998). For resistant cultivars against RYMV transgenic rice plants were made expressing the putative RdRP sequence of RYMV and displaying resistance to several different RYMV strains have been reported (Pinto *et al.*, 1999). The protection achieved

was based on RNA homology dependent resistance. It is the only reported case of transgenic resistance to sobemoviruses (Tamm and Truve, 2000b).
Nothing was known about the molecular features of genes encoding resistance to sobemoviruses until reports of potyvirus VPg interaction with translation initiation factor 4E (eIF4E) (Leonard *et al.*, 2000; Schaad *et al.*, 2000), and the involvement of VPg in resistance breaking (Nicolas *et al.*, 1997; Keller *et al.*, 1998; Moury *et al.*, 2004).
There was no information if VPg as in potyviruses is involved in resistance breaking. The recessive gene *rymv-1*, responsible for the high resistance of *Oryza sativa* 'Gigante' to RYMV was identified. This resistance was overcome by the variant CI4* of RYMV, which emerged after serial inoculations of the non-resistance-breaking (nRB) isolate CI4. Comparison of the full-length sequences of CI4 and CI4*, a non-synonymous mutation was identified at position 1729 (G/T) in the VPg domain. In further experiments neither reversion nor any additional mutation was observed. This mutation was introduced into the VPg of CI4. Analysis of infection revealed infection symptoms in uninoculated leaves of *O. sativa* 'Gigante'. This shoves that VPg is a virulence factor in plants as it is for *Potyviridae* family members (Hebrard *et al.*, 2006).

Characterization of Sobemovirus encoded proteins

P1 description

Special class of plant virus proteins called movement proteins (MP) are involved in virus spread from cell-to-cell and in systemically infection. To overcame the cell wall barrier, they exploit and modify plasmodesmata (Pd), co-axial membranous channels that cross cell-walls of adjacent plant cells, linking there cytoplasm, plasma membranes (PM) and endoplasmic reticulum (ER) (Epel, 2009). Long-distance movement of virus is realized through phloem sieve tubes. There are no sequence similarities between MP from different plant virus taxonomic groups. Each group of MPs had developed different pathways and mechanisms for virus transport. According to movement mechanisms several strategies have been identified – cell-to-cell transport mediated by a single MP; MP and CP but not virus particles, triple gene block, multiple proteins or formation of tubules (Taliansky *et al.*, 2008).
Cell-to-cell transport mediated by a single MP is characteristic for Tobamoviruses, Dianthoviruses, Umbraviruses, Bromoviruses and Cucumoviruses. They encode single MP which associate with ER, binds RNA in sequence non-specific manner, targets Pd and increases the size exclusion limit (SEL) of Pd (Epel, 2009). The best studied MP of this group is from *Tobacco mosaic virus* (TMV). MP of TMV was the first viral protein which was identified as MP (Deom *et al.*, 1987; Meshi *et al.*, 1987). It was identified that TMV MP was localized to the Pd (Tomenius *et al.*, 1987), modifies the Pd SEL (Wolf *et al.*, 1989), binds RNA in sequence non-specific manner (Citovsky *et al.*, 1990), and associates with the ER and ER derived membrane proteins (Heinlein *et al.*, 1998; Reichel and Beachy, 1998; Brill *et al.*, 2000; Fujiki *et al.*, 2006).
Bromoviruses and Cucumoviruses for movement additional element CP is needed, but its involvement is not strict due to deletation experiments in C-terminal end of MP. CP was not necessary for MP interaction of Pd, there for it could be involved in some host response rather than being directly involved in the movement mechanism (Lucas, 2006).
In triple gene block virus transport MP is encoded by triple gene block (TGB) – three partially overlapping reading frames. In this group are two major classes – the hordai-like and the potex-like. As hordai-like are classified viruses members from Benyvirus, Hordeivirus, Pecluvirus and

Pomovirus genera which are rod-shaped. The CP is dispensable for cell-to-cell spread for members of this class. Potex-like TGB proteins are encoded by viruses in the genera Allexivirus, Carlaviruses, Foveaviruses and Potexvirus, which are filamentous viruses that require the CP for cell-to-cell spread (Epel, 2009).

For viruses using multiple proteins for cell-to-cell movement do not encode separate MP, but these functions have been identified to several proteins. Potyviruses belong to this group. In virus movement identified proteins are CP, helper component – protease, cylindrical inclusion protein (RNA helicase) and VPg (Taliansky *et al.*, 2008).

Formation of tubules is another complicatedly different strategy for intracellular movement. Members of this group are Comoviruses and Nepoviruses. Those viruses are transported as virons through specific tubular structures in extensively modified Pd. Tubule formation is made by one virus encoded MP (Taliansky *et al.*, 2008).

All of the characterized *Sobemoviruses* encode a small P1 protein from the 5'-terminal ORF1. There are no sequences and the P1 primary sequences homology of the different members of the genus *Sobemovirus*. They are also unrelated to any other known proteins. The theoretical or experimental molecular masses of different *Sobemovirus* P1 proteins range between 11.7 and 24.3 kDa.

Full-length cDNA clones of RYMV-CI an d SCPMV were constructed and used to study the functions of P1 in the viral life cycle (Bonneau *et al.*, 1998; Sivakumaran *et al.*, 1998). Analysis of mutants incapable of producing P1 or producing only truncated versions of that protein indicated that the RYMV-CI or SCPMV P1 is not needed for virus replication (Bonneau *et al.*, 1998; Sivakumaran *et al.*, 1998). At the same time, the absence of full-length P1 abolished cell-to-cell and systemic movement of the virus in rice and cowpea plants.

The RYMV-CI mutant that did not express P1, due to a mutation at the initiation codon, replicated efficiently in rice protoplasts, but at a level lower than the wild-type cDNA transcript. Transgenic rice plants expressing wild-type P1 were able to complement this initiation codon mutant and exhibited systemic infections. These data demonstrate that one or more of the P1 functions act in *trans* and are essential during infection of plants. These results suggest that the P1 of sobemoviruses can be a virus movement protein. Most plant viruses contain movement proteins which take part in to virus movement from cell-to-cell through the plasmodesmata. Some additional clues was observed like observation of virus-like particles of in plasmodesmata (Opalka *et al.*, 1998). Other mark could be that there are no proteins with such functions associated to any sobemovirus gene products. In study of recombinant CfMV-NO P1 protein RNA binding activity was demonstrated, that His-tagged P1 was able to interact with ssRNA transcripts in a non-sequence-specific manner (Tamm and Truve, 2000a). The biological significance of RNA binding and the domain(s) essential for binding remain are under investigation, but movement proteins, in general, are indeed nucleic acid-binding proteins. It is interesting that RYMV symptoms appeared more rapidly in transgenic plants overexpressing P1 additionaly inoculated with the full-length RYMV-CI transcript than in nontransformed plants (Bonneau *et al.*, 1998). The observation that expression of P1 in *trans* enhances the infection process *in planta* suggests that P1 can act as an enhancing factor for genome amplification. Recently, it was reported that P1 of RYMV-Nig functions as a suppressor of posttranscriptional gene silencing (PTGS) (Voinnet *et al.*, 1999). PTGS of a green fluorescent protein transgene in *Nicotiana benthamiana* plants was reversed by RYMV-Nig P1 expressed from the potato virus X vector. It was concluded that P1 protein of RYMV-Nig is the suppressor of maintenance of PTGS in *N. benthamiana*, although it is encoded in the genome of a virus that is not infectious on *Nicotiana* species (Voinnet *et al.*, 1999).

Similarly, P1 of SCPMV was not required for viral assembly and viral RNA synthesis but was needed to infect cowpea plants (Sivakumaran *et al.*, 1998).
Results from experiments of CfMV P1 showed that P1 was not required for virus replication similarly to in RYMV P1 and SCPMV P1 case (Sivakumaran *et al.*, 1998; Tamm and Truve, 2000b) although there is no sequence homology between these three proteins.
Further analysis of CfMV P1 infected plants indicated that the systemic infection was detected only in case of a spontaneous transversion of mutated ORF1 initiation codon to wildtype whereas the original CfMV P1 population did not infect the host systemically. No systemic spread happens in the absence of P1 expression, as it was shown also for RYMV and SCPMV (Sivakumaran *et al.*, 1998; Tamm and Truve, 2000b).
The ability of CfMV P1 to bind ssRNA but not dsDNA led to the suggestion that P1 may have a role in virus movement (Tamm and Truve, 2000a). In experiments of P1-EGFP fusion only in rare cell-to-cell movement was detected when tested independently from the context of other CfMV proteins in epidermal cells of oat leaves. Thus, it is not possible to confirm that CfMV P1 is indeed the movement protein facilitating local spread of the virus as proposed earlier (Meier *et al.*, 2006).
P1 of RYMV was also reported as a pathogenicity determinant (Bonneau *et al.*, 1998) and in the case of the RYMV-Nig it has been also described as an RNA silencing suppressor (Voinnet *et al.*, 1999). Surprisingly, however, there is no similarity between the AA sequences of sobemoviral P1s (Yassi *et al.*, 1994; Makinen *et al.*, 1995; Othman and Hull, 1995). Similar like RYMV-Nig P1 acts as a silencing suppressor in *Nicotiana benthamiana*, a non-host species (Voinnet *et al.*, 1999), CfMV-NO P1 was investigated for suppressor activity in *N. benthamiana* using the agrobacterium-mediated transient assay (Hamilton et al., 2002). CfMV-NO P1 was able to suppress the initiation and maintenance of silencing. The suppression of systemic silencing was weaker with CfMV-NO P1 than in the case of RYMV-Nig P1. In the case of suppression at the local level, the reduction in the amount of 25 Nt small interfering RNAs (siRNAs) was less pronounced for CfMV-NO P1 than it was when RYMV-Nig P1 was used. At the same time, was show that CfMV-NO P1 did not bind siRNAs (Sarmiento *et al.*, 2007).
Recent studies on RYMV P1 in non-host and host plants revealed that it have multiple effects on exogenous and endogenous gene silencing, specifically acting on the DCL4- dependent endogenous pathway in rice. The P1 protein participate not only in local gene silencing, but also in the short- and long-distance movement of the silencing signal, acting as a suppressor and as an enhancer of silencing. These multiple functions suggest a key role for this protein in the control of virus accumulation in the host, and consequently in the success of the infection (Lacombe *et al.*, 2010).

Protease description

Proteases have several classifications. One is based on their site of action: exoproteinases remove AA from amino- or carboxy-termini of proteins; endoproteinases break specific peptide bonds that exist between two AA located in internal parts of a protein. More popular classification is based on the AA constituting the catalytic site. In this clasifiaction are four different classes: serine and serine-like proteases, cysteine (thiol) proteases, aspartic (or acidic) proteases, and metalloproteases. Only the cysteine, serine and serine-like proteases are involved in proteolytic processing of polyproteins encoded by positive-strand RNA plant viruses. Serine and serine-like proteases have a catalytic triad of histidine, aspartic acid and serine AA while cysteine proteases possess a catalytic dyad of cysteine and histidine AA residues.

Based on the common properties of cellular and viral proteases positive-strand RNA plant viruses proteases could be divided into two main classes – the chymotrypsin-related cysteine and serine proteases and papain-like cysteine proteasess on the basis of (limited) similarity in sequence and spatial folding to the archetype cell enzymes (Mandahar, 2006).

All *Sobemoviruses* contains a protease domain encoded by ORF 2a N-terminal part. Sequence analysis of SCPMV putative protease domain with cellular and viral proteases revealed similarities to the serine protease, which is related to such cellular proteases as trypsin and chymotrypsin (Gayathri *et al.*, 2006).

The sequence of Sobemovirus protease is unique among plant virus proteases in that it resembles a cellular serine protease in possessing a serine residue instead of cysteine in the catalytic triad (Gorbalenya *et al.*, 1988; Dougherty and Semler, 1993). A putative catalytic triad with a consensus AA sequence for sobemovirus proteases was identified as H(x32–44)D/E(x66–67) TxxGxSG. As we reported in our publication, RGMoV protease AA linker between H and D was longer than for other sobemoviruses, but it corresponded with 3C-like proteases from other organisms and viruses (Snijder *et al.*, 1996). The 60 N-terminal AA of P2a contain hydrophobic residues and have been proposed to form a transmembrane domain in the CfMV polyprotein (Ryabov *et al.*, 1996).

The glycine and histidine residues downstream from the putative catalytic residues are suggested to be the site of substrate binding. A biochemical demonstration of protease activity has not been made (Tamm and Truve, 2000b) until Satheshkumar et al., (2004) expressed SeMV recombinant protease in *E.coli*.

In SeMV protease activity experiments with serine protease domain lacking the N-terminal 70 AA (ΔN70-Pro) which contain transmembranal domain was found to be inactive in *trans*. In presence of VPg at the C-terminus of ΔN70-Pro rendered the polyprotein active in *cis* and *trans*. By mutational analysis, it was demonstrated that interaction of W43 in the VPg domain with the protease was responsible for both *cis* and *trans* proteolytic activities and the associated conformational changes (Satheshkumar *et al.*, 2005a).

In experiments for the interacting partner identification of W43 VPg a stretch of aromatic residues (F269, W271, Y315, and Y319) exposed in the protease domain were mutated. W271A protease-VPg mutant showed absence of cleavage activity both *in vivo* and *in trans*, F269A mutant was partially active. Mutations of both tyrosine residues did not result in loss of protease activity. But H275, though not a part of the exposed aromatic stretch, was shown to be essential for protease activity. It revealed that W271 and H275 of the protease domain mediate aromatic stacking interactions with W43 of VPg thereby rendering the protease active (Nair *et al.*, 2008).

The 3D structure of SeMV protease was determined by Gayathri and coworkers (2006). It is the first non structural *Sobemovirus* protein whom is known structure. SeMV protease belongs to the trypsin-like family of serine proteases as it was suggested earlier. The overall fold exhibits the characteristic features of the trypsin fold. It consists of two β barrels (domains I and II) connected by a long inter-domain loop. Domain I and II belong to the all β class of proteins. The active site and the substrate-binding cleft occur in between the two domains and are fairly exposed to the solvent. There are only three helices in SeMV protease (Gayathri *et al.*, 2006) (Fig. 2).

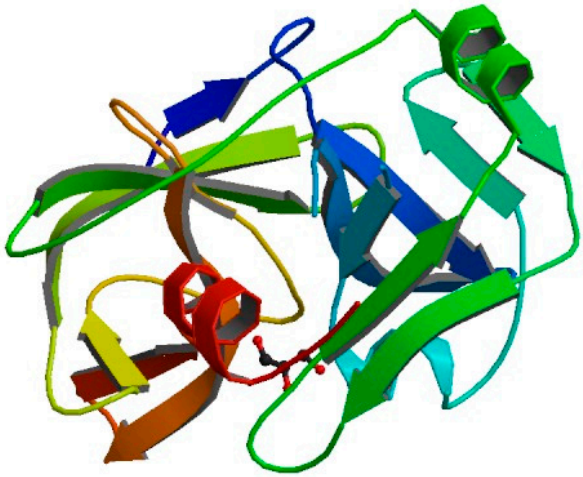

Figure 2. 3D structure of SeMV protease (Gayathri *et al.*, 2006).

VPg description

VPg of potyviruses is a multifunctional protein involved in viral genome translation and replication. Previous studies have shown that both eukaryotic translation initiation factor 4E (eIF4E) and eIF4G or their respective isoforms from the eIF4F complex, which modulates the initiation of protein translation, selectively interact with VPg and are required for potyvirus infection. Reasently two more host factors DEAD-box RNA helicase-like proteins, PpDDXL and AtRH8 from peach (*Prunus persica*) and Arabidopsis (*Arabidopsis thaliana*) were identified to interact with VPg. Using cnock out plants laking AtRH8 it was demonstrated that AtRH8 was not required for plant growth and development but necessary for potyvirus infection. As AtRH8 was identified in infected *Nicotiana benthamiana* leaf tissues colocalized with chloroplast-bound virus accumulation vesicles it could be involved in viral genome translation and replication. An AtRH8 VPg binding region was identified and comparison with PpDDXL it shoved high similarity and could contine similar secondary structure. Overexpression of the VPg-binding region from either AtRH8 or PpDDXL suppresses potyvirus accumulation in infected *N. benthamiana* leaf tissues (Huang *et al.*, 2010).
The first full length structure of an uridylylated chemically synthesized VPg of picornaviruses had been soluted by NMR. The conformation changed upon uridylylation, VPg became more stable, and could fit within the active site of the polymerase near the binding site as it is in the distantly related aphthovirus, *Foot and mouth disease virus* (FMDV) -VPgpU. This suggests that the role of VPgpU in initiating RNA synthesis is to reduce the mobility of the 5' terminal-UMP moiety. Binding of UTP by positively charged amino acids and the C-terminal glutamine residue of VPg, which in the structure of the VPgpU binds the backface of the uridine base, could play an important role in base selectivity during priming (Schein *et al.*, 2010).

VPg is usually a small protein or peptide, which serves as the protein primer for RNA synthesis in many animal and plant viruses. It is covalently linked to the 5' end of virus genRNA and sgRNA.

The VPg of *Sobemoviruses* is released from P2a after cleavage of the full-length protein by the viral serine protease. In *Sobemoviruses,* the size of VPg varies between 9 and 12 kDa (Tamm and Truve, 2000b). The N-terminus of the CfMV and SBMV VPg, and N- and C-termini of the SeMV VPg, were sequenced, showing that the CfMV cleavage occurs between a glutamate/asparagine (E/N) dipeptide (Makinen *et al.*, 2000) and the SBMV and SeMV cleavages occur between glutamate/threonine (E/T) dipeptides (van der Wilk *et al.*, 1998; Satheshkumar *et al.*, 2004). For SeMV, a third, less efficient cleavage also occurred between a glutamate/serine (E/S) dipeptide (McGavin and Macfarlane, 2009). The motif W(A,G)D followed by a D- and E-rich region is the only sequence motif conserved between VPgs of *Sobemoviruses* and related viruses (Makinen *et al.*, 2000). The new putative *Sobemoviruse* member *Rubus chlorotic mottle virus* (RuCMV) VPg instead has a WNK motif and a DE-rich sequence (McGavin and Macfarlane, 2009).

Detection of CfMV VPg in infected plants practically failed. It was hardly to detect the mature 12-kDa VPg was with the antibodies raised against VPg. This may indicate that CfMV VPg does not exist in its free form and that it is immediately linked to the viral RNA (Mäkeläinen, 2006).

There are pure knowledge's of VPg functions and properties. NMR studies on *Cowpea mosaic virus* (CPMV) VPg suggested that it does not have any ordered structure (van de Ven *et al.*, 1990). VPg of SeMV was analysed with CD Spectral Analysis, which revealed unfolded nature of VPg. This result was confirmed with fluorescence spectroscopic analysis, which revealed that VPg lacks tertiary structure**.** SeMV VPg is a natively unfolded protein lacking both secondary and tertiary structures (Satheshkumar *et al.*, 2005a)**.**

"Natively unfolded" proteins are a unique class of proteins that exhibit their function in the absence of ordered structure. These proteins are believed to adopt a rigid conformation stabilized *in vivo* upon interaction with natural substrates (Dyson and Wright, 2002; Uversky, 2002).

Recent biochemical studies on the protease–VPg domains of SeMV have shown that the interaction of VPg with the protease domain modulates the protease activity (Satheshkumar *et al.*, 2005a).

When present at the C terminus of the protease domain (ΔN70Pro-VPg), VPg modulates both *cis*- and *trans*-catalytic activities of the protease by inducing a conformational change in the individual domains. W43 of VPg was identified to be crucial for these interactions. Deletion of the VPg domain from the polyprotein resulted in the complete inhibition of proteolytic processing despite the protease domain was intact. These results identified that the natively unfolded VPg regulates the proteolytic maturation of the polyprotein in sobemoviruses (Satheshkumar *et al.*, 2005a)**.**

In experiments with CPMV, expression of the protease with the N-terminal VPg sequence enhanced its proteolytic activity (Dessens and Lomonossoff, 1992). The mechanism by which VPg activates the proteolytic activity is not known.

Unfolded VPg ineraction with protease could be needed for regulation of polyprotein processing or for biological functions of VPg. The release of different polyprotein precursors at appropriate stages of viral life cycle could be an important strategy employed by viruses for their efficient multiplication and survival within the host cell. As for VPg biological properties it could apply to the function of VPg acting as a primer for RNA synthesis (Satheshkumar *et al.*, 2005a). There is information that AA side chain hydroxyl group (from tyrosine, serine, or threonine) can mimic the 3'-OH group of a ribonucleotide. Those residues are highly specific; for example, it is Tyr-

1860 in *Tobacco vein mottling virus* (TVMV) that links VPg to the TVMV genRNA (Murphy *et al.*, 1991). As the genome organization of the potyvirues are similar to the picornaviruses and comoviruses (Momier et al., 1987). VPg of the poliovirus was identified to link to the viral RNA by a tyrosine residue (Ambros and Baltimore, 1978; Rothberg *et al.*, 1978), and for *Cowpea mosaic virus* it was a serine residue (Jaegle *et al.*, 1987).
In the genRNA isolated from native virus, both the processed and the unprocessed forms of VPg are linked to the 5' end of the genome (Merits et al., 2002). If a protein is unfolded, then the probability of the same AA acting as a primer would be highly unlikely as the protein can be in multiple conformations. VPg fusion with the protease could result in the conformation in which the hydroxyl group of a specific AA can act as the primer (Satheshkumar *et al.*, 2005a).
Acording to recent studies another function for VPg has been assigned in viral RNA translation. It is proposed that VPg might function in a way similar to the 5' 7-methyl guanosine cap structure present at the 5' end of eukaryotic mRNAs in recruiting the translation initiation factors (IFs). For *Norwalk virus* was shown that VPg interact with eIF3 (Daughenbaugh *et al.*, 2003). In *Turnip mosaic virus* (TuMV) case, VPg interacted with eIF4 and poly(A)-binding protein only in the form of 6K2-VPg-Pro or VPg-Pro polyprotein precursors *in planta.* Based on this result, a possible role of VPg-Pro in the formation of the assembly of the translation initiation complex has been proposed (Leonard et al., 2004). Thus, by using the strategy of polyprotein processing, viruses can generate several proteins and their precursors, which could have multiple functions (Satheshkumar *et al.*, 2005a).
In investigation of TuMV VPg and *Arabidopsis thaliana* translation IF of their binding properties and kinetics an affinity chromatography on m7GTP-sepharose was made. Those results showed that bound *A. thaliana* eIF(iso)4E was eluted with crude TuMV VPg. Further column studies with purified VPg and other *A. thaliana* eIF4E isoforms showed that VPg preferentially bound eIF(iso)4E. Structural data implicate Trp46 and Trp92 in eIF(iso)4E in cap recognition. When Trp46 or Trp92 were changed to Leu, eIF(iso)4E lost the ability to form a complex with both VPg and m7GTP-sepharose. This suggests that the VPg-binding site is located in or near the cap-recognition pocket on eIF(iso)4E. Affinity constants for the interactions with eIF(iso)4E of VPg and capped RNA oligomer were determined using surface plasmon resonance (SPR). The *KD* values showed that the binging affinity of VPg for eIF(iso)4E is stronger than that of capped RNA. This suggests that viral VPg can interfere with formation of a translational initiation complex on host plant cellular mRNA by sequestering eIF(iso)4E. Additional experiments with affinity chromatography showed that VPg forms a ternary complex with eIF(iso)4E and eIF(iso)4G. So, VPg may participate in viral translational initiation by functioning as an alternative cap-like structure (Miyoshi *et al.*, 2006).

P16 analog description

Helicases are well-established class of enzymes that unwind double-stranded DNA (dsDNA), double-stranded DNA/RNA hybrids, or double stranded RNA (dsRNA) structures during replication and/or transcription of cellular doubleand viral genomes by disrupting the hydrogen bonds by a reaction that is coupled with hydrolysis of an NTP (Kadare and Haenni, 1997). They contain nucleoside 5`-triphosphate (NTP)-binding motif, separate the duplex oligonucleotides into single strands in an ATP-dependent reaction, generate energy, and are encoded by most of the positive-strand RNA viruses (Mandahar, 2006).
The helicase enzyme had been definitely detected, identified, and established in only in two RNA plant viruses - *Plum pox potyvirus* (PPV) and *Tamarillo mosaic potyvirus* (TaMV) (Kadare

and Haenni, 1997). In all other cases, the helicase motif has been detected in the genome sequence. These are putative helicases since they are determined only on the basis of sequence similarities/homologies with definitive helicases but possess no supporting biochemical data. It is possible that many of these putative helicases have helicase activity that is dependent upon ATP hydrolysis (Mandahar, 2006).

The relationship between genome size and existence of putative helicases could corelate in ss positive-sense RNA viruses (Gorbalenya *et al.*, 1989; Koonin, 1991b). Around 80% of such viruses, whose genome sequence is known, possess at least one potential helicase except those having genome smaller than < 5.8 kb. The observations suggest that viruses with a small double-stranded or ssRNA genome may not need RNA unwinding for its replication (Mandahar, 2006).

All helicases and putative helicases of ss positive-sense RNA viruses are clasified in to three superfamilies.

Viruses of helicase superfamily 1 (SF1) contain the helicase-like proteins of alpha-like plant viruses. Following positive-strand RNA plant virus genera contain the proposed SF1-type helicase - the particular replicative virus protein (p) that contains the helicase domain, Alfamovirus (p1a), Bromovirus (p1a), Capillovirus (p241), Carlavirus (p26), Closterovirus (p295), Cucumovirus (p1a), Furovirus (p237), Hordeivirus (p130), Idaeovirus (p190), Potexvirus (p180), Tobamovirus (p126), Tobravirus (p134), and Tymovirus (p206). SF1 helicases contines seven conserved sequence motifs called I, Ia, and II to VI. Helicases of SFs 1 and 2 share motifs, which align with ATP-binding and ATP hydrolyzing Walker A and B motifs. Lysine residue of motif I (K844 in AMV P1) is essential for ATP binding; its mutation inhibits ATP hydrolysis and unwinding of duplex. SF1 helicase motif VI is thought to be interacting with ATP and functioning in transition of energy from ATP hydrolysis to duplex unwinding.

Helicase superfamily 2 (SF2) contains potyvirus-flavivirus-pestivirus helicase-like proteins. 2-type helicases are: Bymovirus (p270) and Potyvirus (CI protein). The PPV CI protein was the first helicase iedentified and characterized of positive strand RNA viruses. It was shoven that it it hydrolyses NTP and is Mg^{2+} depenante. CI efficiently binds poly(A) and is specific for PPV RNA duplex. Unwinding is made from 3'- to 5'-end and 3'-single-strand terminal is needed.

Helicase superfamily 3 (SF3) includes small putative helicase domains of ~100 AA residues found in DNA and RNA viruses, has only three conserved motifs, two of them are classical ATP-binding motifs. SF3 are associated with supergroup 1 polymerases. This superfamily contains picornavirus-like helicase-like proteins. This group containing members Comovirus (p58), Nepovirus (p27), Sequivirus (p336), and Waikavirus (ORF 1). The NTP-binding elements A (I) and B (II) are apparently absent in several families of the plus-strand RNA viruses. This is true of some viruses possessing supergroup I polymerases of sobemo lineage (Sobemovirus and Luteovirus subgroup II).

There is little known about other putative Sobemoviruse protein domains. The information about new class of proteins appeared in last few years. This class could be grouped as putative RNA binding proteins, probably having the function of helicase, reported to be present in other plant viruses. These domains are located in proximal part of 2a polyprotein, directly after VPg. As several *Sobemovirus* nonstructural proteins, these lack molecular mass identity and sequence similarities to each other. Mäkeläinen (2006) cloned and expressed CfMV protein P27. As revealed experiments with the protein, it repressed translation of proteins encoded downstream from the -1 PRF signals. Thus, P27 may play a role in regulating the amount of RdRp produced during the infection process. Alternatively, P27 may function in regulating the ribosomal load at the 3'end of the RNA (Mäkeläinen, 2006).

All SeMV 2a proximal protein domains were cloned and expressed in *E.coli*. The region corresponding to P10 and P8, and P8 alone was cloned, over-expressed and purified using Ni-NTA affinity chromatography. The purified P8 His-tagged protein moved abnormally as 15 kDa on the SDS-PAGE. The mass spectrometric analysis of this protein showed that it had a molecular mass of 9.766 kDa corresponding to the His-tagged P8. P8 has high positive charge (theoretical pI-11.75), which could explain its anomalous behavior on the SDS-PAGE (Nair and Savithri, 2010b).
As revealed functional analysis of P10 and P8, the domains posses ATPase activity. ATPase activity was detected for P10 in experiment with and without P8. Both domains containing P18 protein demonstarted ATPase activity but deletion of P8 domain it reduced. P8 could be classified as ATPase activator. Another property for P8 was detected: it was able to bind nucleic acids in sequence-independent way (Nair and Savithri, 2010a). Our data for P16 revealed similar nucleic acid binding properties as well as unusual behavior in the SDS-PAGE. Nair and Savithri (2010) was identified in P18 that P8 is natively unfolded protein. It is in good agreement with our data about RGMoV P16 N-terminal domain.

RNA dependent RNA polymerase description

The RdRps of positive-strand RNA viruses are classified into three supergroups (1, 2, and 3) on the basis of sequence similarities between more than 300 AA. They shared a set of conserved sequence motifs.
RdRp Supergroup 1: It contains plant viruses of picorna-like lineage (picornaviruses, comoviruses, nepoviruses), viruses of the poty lineage (potyviruses and bymoviruses), and viruses with small genomes (sobemoviruses and luteoviruses).
RdRp Supergroup 2: It contains plant viruses with small genomes (dianthoviruses, carmoviruses, tombusviruses, and necroviruses).
RdRp Supergroup 3: It contains tymo-like (tymoviruses, carlaviruses, potexviruses, and capilloviruses), rubi-like (*Beet necrotic yellow vein virus* and possibly alphaviruses), and tobamo-like viruses (tobamoviruses, trichoviruses, hordeiviruses, tobraviruses, and closteroviruses) (Mandahar, 2006).
There are some reports according to factors involved in ss positive-sense RNA replication. For AMV is reported that CP acts as integral component of replicase. It binds to the viral RNA 3′-termini and induces the formation of multiple new base pairs that organize the RNA conformation so defining template selection and positioning the RdRp at the initiation site for minus strand synthesis (Reichert *et al.*, 2007).
There are reports that host factor for RNA synthesis are nessery. Isolated replication complexes containing virus-encoded RdRp, but free from host RdRp, are able to synthesize full-length viral RNA in several plant viruses (BMV, CMV, CNV, CPMV, TCV, TBSV, TYMV). The RdRps in those cases replicated and transcribed viral RNAs specifically. It means that viral polymerase binding to the RNA template is mediated by cellular factors. Absence of thise factors cose failure to bind directly to *cis*-acting regulatory or promoter sequences on viral RNA. It is generally assumed now that polymerases/replicases of all positive-strand RNAs are composed of both virus- and host-coded proteins. The interactions between viral RdRp and cellular factors may form an RdRp complex even in the absence of RNA template in situations when cellular factor is already a part of RdRp as in BMV and TMV. In other cases, cellular factors first dock with viral RNA template, as in TYMV, followed by interaction of this RNA template with viral RdRp.

Thus, viral RdRps and cellular factors together form the transcription or replication complexes on viral RNA.
In BMV, CMV, TMV, and TYMV case RdRp replication complex contained copurified host membranes and host proteins. There was identified that the eukaryotic translation initiation factor (elF3) binds to RdRp complex of BMV and TMV.
In vitro mutations in GDD box of several animal and plant viruses greatly reduced or abolished RNA replication as in PVX and TYMV. For example, when GDD box of 166-kDa protein of PVX was mutated to GED, ADD or GAD, virus infectivity was abolished and RNA replication in protoplasts was greatly reduced (Longstaff *et al*., 1993). Similarly, TYMV RNA replication was abolished by substitution of G to R in GDD box of 66-kDa TYMV protein (Weiland and Dreher, 1993).
There are several 3D structures for animal RdRp which shoves some similarities in structure. There could be identified a basic right hand-like structure with 'palm', 'thumb', and 'finger' domains and with a cleft existing between fingers, palm, and thumb subdomains. It is the palm domain structure that is especially conserved and has four sequence motifs that are conserved in polymerases of all RNA and DNA viruses. However, the finger domains differ significantly among viral RNA polymerases. This may be due to specific adaptations to structurally diverse substrates. The basic polymerase right hand shape provides the correct geometrical arrangement of substrate molecules and metal ions at the active site for catalysis (Mandahar, 2006).
On the basis of structural similarities and the conserved motifs, it has been proposed that all polymerases make use of a common two-metal mechanism of catalysis. This involves two conserved aspartic acid residues from A and C motifs, respectively, and two divalent metal ions for the formation of the phosphodiester bonds. The carboxylates anchor a pair of divalent metal ions that perform the major role in catalysis. One divalent ion (Mg^{2+}) promotes the deproteination of the 3`- hydroxyl of the nascent strand, while the second (Mg^{2+}) facilitates the formation of the pentacovalent transition state at 7-phosphate of dNTP and the exit of the inorganic pyrophosphate group.
Sobemovirus RdRp is encoded by ORF2b and translated as polyprotein Pro-VPg-RdRp by -1 ribosomal frameshift signal and released from polyprotein via processing mechanism held by protease. Putative RdRp domain of Sobemoviruses was identified based on the presence of the GDD motif and surrounding conserved motifs characteristic of RdRps (Koonin, 1991a; Koonin and Dolja, 1993). The putative sobemovirus RdRps show extensive similarities to RdRps of a number of positive-strand ssRNA viruses, which include poleroviruses (PLRV, BWYV, BMYV, CYDV-RPV, and CABYV), an enamovirus (PEMV-1), and a barnavirus (MBV). Such similarities have been used to evaluate the taxonomic position of SCPMV in relation to other positive-strand RNA viruses (Koonin, 1991a; Koonin and Dolja, 1993). SCPMV, BWYV, PLRV, and PEMV-1 have been grouped into the Sobemo-lineage of polymerase supergroup 1 (Koonin and Dolja, 1993), indicating that the RdRPs of poleroviruses are more similar to sobemoviruses than they are to luteoviruses (subgroup I luteoviruses). There is practically no information about the replication signals needed for initiation of plus- and minus-strand synthesis in *Sobemoviruses*. Several primer extension experiments have been used to identify the precise 5' ends of the genRNAs as a start point for plus-strand synthesis. The 5'-terminal Nt of the SCPMV genRNA were identified as ACAAAA (Hacker and Sivakumaran, 1997). Similar sequence was detected for LTSV (ACAAA) and RYMV-CI (ACAA) (Yassi *et al.*, 1994; Jeffries *et al.*, 1995). For SBMV and SBMV-Ark genRNAs was identified similar motif, CACAAAA (Othman and Hull, 1995; Lee and Anderson, 1998). Similar sequence was detected for 5' ends of the SCPMV and SBMV sgRNAs locating at 3241 and 3163 (Hacker and Sivakumaran, 1997),

and, based on the sequences of the SCPMV and SBMV genomes (Wu *et al.*, 1987; Othman and Hull, 1995), also possess the sequence ACAAAA. Those similar sequences could be potentially important for virus replication. In CfMV case different sequence motif present at the 5' end of the CfMV genome (Makinen *et al.*, 1995; Ryabov *et al.*, 1996). It could be due to different hosts. All *Sobemoviruses* have a polypurine tract, including the sequence aAGgAAA (lowercase indicates less conservation of that base) just at the beginning of the genRNA (Hacker and Sivakumaran, 1997).

Those sequence motives were used as possible predictions for the 5' termini of other *Sobemovirus* sgRNAs. The sequence ACAAAA (Nt 3222 to 3227) was identified for the SBMV-Ark CP initiation codon. In RYMV-CI case the sequence ACAAA (Nt 3441 to 3445) was identified and determined location 6 Nt upstream of the ORF4 AUG codon. For the LTSV sgRNA also has the ACA AAA sequence motif (Nt 3285 to 3290).

The ACAAAa sequence was common for poleroviruses 5' termini of genRNAs and sgRNAs (Miller *et al.*, 1995). This motif was found at the 5' end of RCNMV (genus *Dianthovirus*, family *Tombusviridae*) RNA1 and sgRNA (Xiong and Lommel, 1989; Zavriev *et al.*, 1996). The 5' termini of the MBV genRNAs and sgRNAs also begin with this sequence (Revill *et al.*, 1994; Revill *et al.*, 1998). These sequences at the 5' ends of genRNAs and sgRNAs of several viruses belonging to different groups suggests that it or its complementary sequence in the minus-strand RNA may function in viral RNA synthesis. Perheps the minus-strand sequence complementary to the ACAAA domain may act as a promoter or enhancer for viral replicase binding and initiation of RNA synthesis (Miller *et al.*, 1995). But there are no experimental data confirming it. There is practically no information about the replication signals needed to initiate minus-strand synthesis at the 3' ends of *Sobemovirus* genRNAs. A potential tRNA-like structure has been attributed to the 3' end of some *Sobemovirus* genRNAs (Yassi *et al.*, 1994; Ryabov *et al.*, 1996). But for SBMV and SCPMV, it has been impossible to find an RNA sequence at the 3' end that has the potential to fold into a tRNA-resembling secondary structure (Wu *et al.*, 1987; Othman and Hull, 1995).

The only soluble and purified RdRP was obtained from corresponding cDNA of SeMV in a form of thioredoxin-tagged protein. The recombinant SeMV RdRp was able to synthesize RNA from genomic or a sgRNA template even in the absence of the protein primer, VPg. Analysis of the synthesized product revealed the double-stranded RNA and the mode of initiation was *de novo.* The results suggest that Sobemovirus RdRp can be classified as primer independent polymerase (Govind and Savithri, 2010). The examination of the 3′ UTR of sgRNA by mutation analysis revealed that a stem–loop structure located at the 3 end was import ant. Additional analysis revealed that the SeMV RdRp was capable of recognizing stem–loop structures of various lengths and forms (Govind and Savithri, 2010).

Coat protein description

The CP of *Sobemoviruses* is encoded by their 3'-proximally located ORFs. The AA sequence of the SCPMV CP reported by Hermodson (Hermodson *et al.*, 1982) showed that translation initiation occurs at the second AUG (Nt 3271 to 3273) of ORF3 and is followed by the hydrolysis of the N-terminal methionine and acetylation of the subsequent alanine. Direct sequencing of the N-terminus of the RYMV-CI and CfMV-NO CP showed that they commence at the first AUG codon of the last ORF, at Nt 3447 and 3093 (Yassi *et al.*, 1994; Makinen *et al.*, 1995). A tentative phylogenetic tree anlalysis of icosahedral plant virus CP revealed three distinct subdivisions. The CP of SCPMV grouped together with the *Tobacco necrosis virus*

(TNV; genus *Necrovirus*, family *Tombusviridae*) CP rather than with the CPs of poleroviruses. The CP of RYMV-CI is also more closely related to that of TNV (Yassi *et al.*, 1994), indicating that CP of sobemoviruses and necroviruses are phylogenetically related (Dolja and Koonin, 1991). This differs from phylogenetic analysis of sobemovirus proteases, VPgs, and RdRps, which are related to poleroviruses (Tamm and Truve, 2000b).

The CP is the single structural protein required for sobemovirus icosahedal particle building. The three-dimensional structure of SCPMV has been determined at 2.8 Å resolution (Abad-Zapatero *et al.*, 1980), SBMV has been determined at 2.9 Å resolution (Silva and Rossmann, 1987), SeMV has been determined at 3 Å resolution (Bhuvaneshwari *et al.*, 1995), RYMV has been determined at 2.8 Å resolution (Qu *et al.*, 2000), CfMV has been determined at 2.7 Å (Tars *et al.*, 2003), and in this work reported RGMoV has been determined at 2.9 Å resolution (Fig. 4).

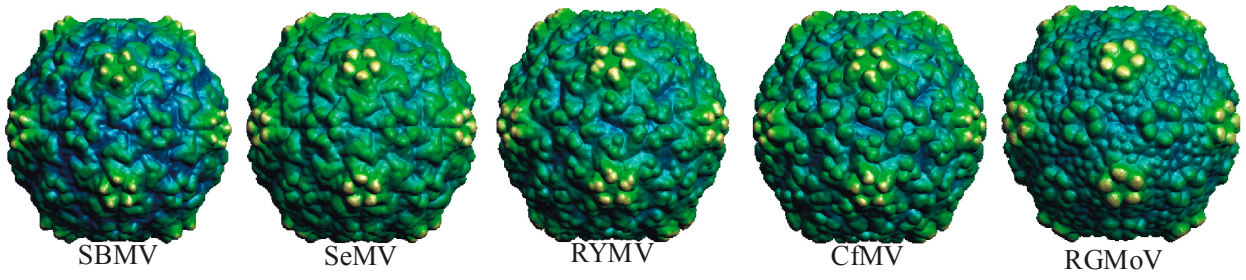

Figure 4. 3D structures of crystallized Sobemoviruses (Figures obtained from VIPER data base) (Silva and Rossmann, 1987; Bhuvaneshwari *et al.*, 1995; Qu *et al.*, 2000; Tars *et al.*, 2003), Plevka et al., 2007).

Each icosahedral unit of the particle is made of three quasiequivalent subunits, A, B, and C, whose individual conformations is slightly different. The A subunits cluster are located at the fivefold axes, whereas sets of three B and three C subunits cluster at quasi-sixfold vertices. Analysis of the X-ray structures revealed that the CP is composed of two functional domains, the R (random) domain and the S (shell or surface) domain, connected by an arm (Abad-Zapatero *et al.*, 1980; Hermodson *et al.*, 1982; Rossmann *et al.*, 1983b). The S domain is contains eight antiparallel β-sheets (termed a β-barrel) and five α-helices. The function of S domain is to create subunit-subunit interaction in virus particle (Hermodson *et al.*, 1982; Rossmann *et al.*, 1983b). The R domain is located in N-terminal part of the polypeptide chain and contains high number of positively charged a - argentine, lysine, praline, and glutamine. R domain together with some similar residues on the inner surface of the S domain is responsible for CP binding with the RNA (Hermodson *et al.*, 1982; Rossmann *et al.*, 1983b). Those basic residues are facing the RNA is able to dock a 9-bp double-helical A-RNA structure with surprising accuracy. Each basic residues are associated with a different phosphate, and the protein can interact with five bases in the minor groove. R domain of the CP of SCPMV (AA residues 1 to 54) was expressed in *Escherichia coli* as recombinant protein and in experiments *in vitro* showed that it had nonspecific RNA binding activity (Lee and Hacker, 2001).

Studies of N-terminal alanine mutants were confirmed that those residues were required for RNA binding. RNA and CfMV-NO CP (native and recombinant) RNA binding experiments were made (Tamm and Truve, 2000a). Both CP bound ssRNA in a non-sequencespecific manner *in vitro* and were selective for ssRNA over double-stranded DNA molecules. In addition, SCPMV CP possesses sequence-specific viral RNA binding properties (Hacker, 1995). The AA residues responsible for SCPMV CP RNA binding *in vitro* were mapped to the arginine-rich region (Lee and Hacker, 2001).

In experiments with swollen, trypsin-digested SBMV removing basic N-terminal end the CP assembly into T=1 particles was demonstrated (Sehgal *et al.*, 1979). Erickson and Rossmann (1982) described similar situation and showed that T=1 particle formation of SBMV did not require RNA. Savithri and Erickson (1983) studied RNA-CP interactions role in the assembly of SCPMV. Removal o f CP N-terminal part resulted again in T=1 particle formation. Similar observation was made from N-terminaly truncated SeMV CP studies (Lokesh *et al.*, 2002). Mutation analysis of arginine-rich region of recombinant SeMV CP revealed that only empty T = 3 particles were formed proposing the importance of arginine residues for RNA encapsidation (Satheshkumar *et al.*, 2005b).

Full-length cDNA clone mutants of RYMV-CI (deletion of C-terminal n and frameshift) and SCPMV (mutation of initiation codon and insertions) served as models for the function analysis of the CP (Brugidou *et al.*, 1995; Sivakumaran *et al.*, 1998). No systemic or local infection was observed by plant inoculation of both virus mutants, indicating that CP is essential for cell-to-cell and systemic virus movement. At the same time the CP was not required for RYMV-CI or SCPMV RNA synthesis in rice and cowpea protoplasts. In rice plants, RNA replication for both RYMV-CI mutants was detected in leaves 4 weeks after inoculation, indicating the importance of CP particularly in long-distance virus movement. Long distance movement of SCPMV (Fuentes and Hamilton, 1993) and RYMV (Opalka *et al.*, 1998) has been observed to be dependent on particle formation. Direct evidence that CP determines the range of systemic hosts of sobemoviruses was provided by characterizing the resistance-breaking mutant of SBMV-Ark, SBMV-S (Lee and Anderson, 1998). SBMV-S is able to move systemically in bean cultivars "Pinto" and "Great Northern", although the wild-type SBMV-Ark is restricted to the inoculated leaves of this host. Sequence analysis of the genomes of SBMV-Ark and SBMV-S revealed 7 Nt differences but only four deduced AA changes (Tamm and Truve, 2000b).

Another observation of the N-terminal part of SCPMV CP was made it could interact with membranes (Lee *et al.*, 2001), however, the biological properties of that feature is unknown. There are several reports that Sobemovirus particles have been found in the nucleus (Mohamed and Mossop, 1981; Fuentes and Hamilton, 1993). Sequence analysis *in silico* revealed similarity with a nuclear localization signal (NLS) located in the N-terminal region of sobemovirus CPs (Ngon *et al.*, 1994; Makinen *et al.*, 1995).

NLS deletation (AA 1–11, 1–22, and 22–33) analysis of CfMV CP indicated two separate NLS signals within the N-terminus - a strong NLS1 in the arginine-rich region (residues 22–33) and a weaker NLS2 within residues 1–22. Point mutants revealed that the basic AA residues in the region of the two NLSs were individually not sufficient to direct CP to the nucleus. But microinjection studies with fluorescently labeled RNA and CP purified from CfMV particles demonstrated that the wild-type CP was capable of transporting the RNA to the nucleus. Transmission was not sequence-specific because both CfMV and GFP mRNA were transported to the cell nucleus by CfMV CP. These results suggest that the plant cell nucleus may be involved in CfMV infection (Olspert *et al.*, 2010).

There are several hypotheses why the CP is transported to the nucleus. First could be reduction of CP concentration in cytoplasm for the regulation of virus replication. Second could be due to unknown CP functions which require its transport to the nucleus (Olspert *et al.*, 2010). There are reports of umbra- and potyviruses encode proteins that interact with a nucleolar protein fibrillarin, which has been shown to be essential for systemic infection and virus accumulation (Kim *et al.*, 2007a; Kim *et al.*, 2007b; Rajamaki and Valkonen, 2009).

Characterization of RGMoV

Ryegrass mottle virus (RGMoV) was first isolated from stunted Italian ryegrass *(Lolium multiflorum)* and cocksfoot *(Dactylis glomerata)* having mottling and necrotic symptoms on leaves (Fig. 5) The isometric particle, 28 nm in diameter contains ssRNA with a molecular weight of 1.5 X 10^6. It in the physical and some biological properties are similar to CfMV, which is prevalent in cocksfoot pastures in Japan. But it is serologically distinct from CfMV, *Cynosurus mottle virus* and *Phleum mottle virus,* which occur in European countries (Zhang *et al.*, 2001b).
The genome of RGMoV is a single-stranded, positive-sense, monopartite RNA with a covalently linked 5′ VPg protein, which may serve as a replication primer, and lacking poli(A) at 3' end (Zhang *et al.*, 2001b). The replication is carried out by a virus-encoded RdRp. During replication full size copies of genRNA are created together with sgRNAs corresponding to the 3′ part of the genome. The RGMoV genome contains four ORFs. ORF1 encodes protein P1 with predicted molecular mass 14.6 kDa containing 133 AA, which possibly posses similar properties as RYMV, SCPMV and CfMV P1. RGMoV polyproteins are encoded by two overlapping ORF2a and ORF2b. The ORF2a codes for a membrane anchored 3C-like serine protease, VPg and P16, whereas ORF2b contains the RdRP domain as revealed our research data. The CP is expressed from the 3′ proximal ORF3 containing 198 AA encoding a 25.6kDa.

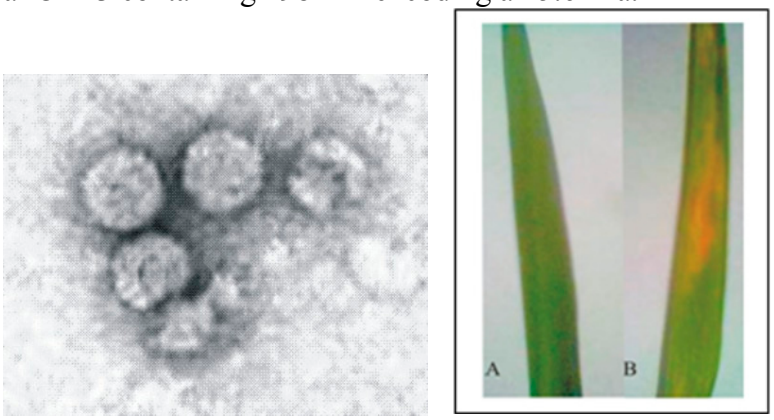

Figure 5. Native RGMoV in electronmicroscpy and comparison of non infected A and infected plant B.

Genome contains UTR regions at 5' and 3' ends as well as short UTR between ORF1 and ORF2. The protease motif is located downstream of the conserved sequence: serine protease, in AA 148 to 220 from N-terminus and P3C protease, in AA 272 to 300. The serine protease motif is well conserved between RGMoV, sobemoviruses and poliovirus. In addition, the P3C protease motif xGxS . /C . GxxxxxxxxGxxxxGxH* (the catalytic AA residue is marked with asterisk), is present just downstream. As was reported earlier (Zhang et al., 2001) instead of serine (S*) or cysteine (C*), alanine was found in RGMoV. That raised a question whether the P3C protease domain is catalytic in RGMoV or not. According to our studies, we had demonstrated that previous sequence had sequencing erros and new reported sequence of P3C had all necesery AA in catalytic and substrat binding sytes (Publication I).
Sequence analysisof RGMoV CP to the other sobemovirus members revealed 24 to 27% identity with SBMV, LTSV and RYMV, but only 15% identity with CfMV (Zhang *et al.*, 2001b).

The crystal structure of the RGMoV has been determined at 2.9 Å resolution (Fig. 6). The CP has a canonical jellyroll β-sandwich fold. In comparison to other sobemoviruses the RGMoV CP is missing several residues in two of the loop regions.

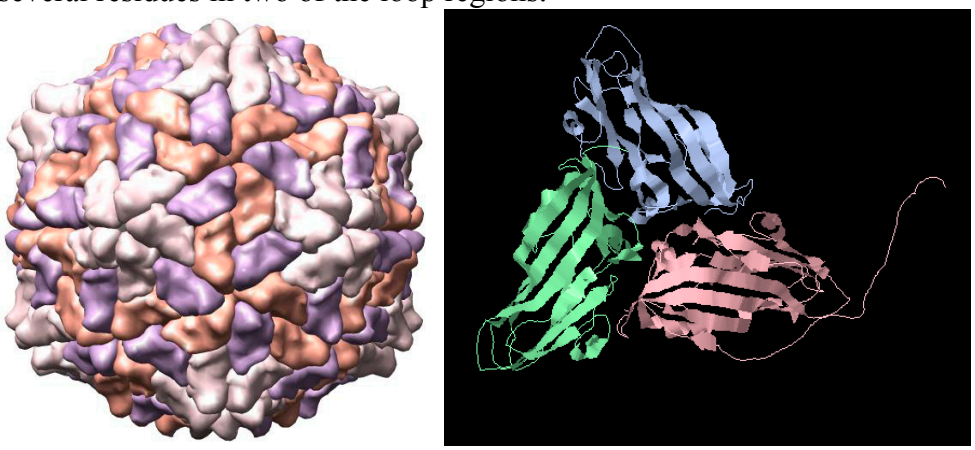

Figure 6. RGMoV 3D structure and capsomer (VIPER data base – 2IZW).

One of detected difference from other sobemoviruses is missing of C helix of FG loop. This loop has been found only in sobemoviruses and TNV. RGMoV is unique among the sobemoviruses of known sequence in lacking this helix. GH loop is other missing element of RGMoV. GH loop is interacting with β-annulus in T=3 capsid formation, but RGMoV compensates it with additional interactions between the N-terminal arms. This difference reflects as smaller diameter of particle by 8 Å smaller. Based on loop region RGMoV CP is sighty more similar to TNV than to the sobemoviruses. Those differences are giving shape close to the rhombic triacontahedron with pentamers emanating distinctively at the fivefold, shallow clefts around the twofold, and small protrusions at the threefold icosahedral symmetry axis.
The reduction of the GH loop and the absence of the C helix in the FG loop, results in the 12° bending of the interface between the asymmetric units related by the icosahedral twofold symmetry axis into the particle. The FG loops in other sobemoviruses occupy, on average, 5% of the virion volume. The decreased diameter and the missing FG loop in RGMoV have an opposing effect on the size of the space available for genome storage. Overall, the size of the RGMoV virion cavity is 7% smaller than in other sobemoviruses. The RNA packing density of 480 mg/ml is similar to other sobemoviruses and TNV. On average, 20% of the sobemovirus particle cavity is taken by the disordered N-terminal parts of the CP. It is very likely that the flexible arms do not impose any additional restraints on RNA packing. If the volume taken by the N-termini is considered available for genome storage, then the average RNA packing density in sobemoviruses would be 370 mg/ml.

Sobemovirus capside structure and formation conditions

The crystal structure of six members of the *Sobemovirus* genus has been determined: SCPMV (previously known as *Southern bean mosaic virus, cowpea strain*) (Abad-Zapatero *et al.*, 1980), SBMV has been determined at 2.9 Å resolution (Silva and Rossmann, 1987), SeMV (Bhuvaneshwari *et al.*, 1995), RYMV (Qu *et al.*, 2000) and CfMV (Tars *et al.*, 2003) and RGMoV (Plevka et al., 2007). The structure of *Tobacco necrosis virus* (TNV) CP (genus

Necrovirus; *Tombusviridae* family) is similar to sobemoviruses (Oda *et al.*, 2000). The sobemoviruses are not assigned to the *Tombusviridae* family because the gene order and the properties of the nonstructural proteins are distinctly different (Tamm and Truve, 2000b).
Sobemoviruses have a capsid arrangement with icosahedral T=3 quasi-symmetry in which the 60 icosahedral asymmetric units consist of three chemically identical copies of the CP differing in conformation - A, B and C. Five A subunits form pentamers around the icosahedral fivefold axis and three of each of the B and C subunits assemble into hexamers around the threefold axis. The difference in shape is significant - hexamers are planar and pentamers are bent. The hexamer/pentamer combination on the virus surface gives the particle its characteristic shape of a rhombic triacontahedron. Capsid size is determined by the CP N-terminal arm of the C. The N-terminal arm of the C subunit is partly ordered unlike A and B subunits, and inserted between the interacting sides of the subunits, making the contacts around the icosahedral twofold axis flat. The contacts of the A and B subunits around the quasi-twofold axis lack the inserted arms and are bent. The flat or bent type of contact between the subunits is thus caused by the presence or absence of an ordered N-terminal arm. The removal of the N-terminal arm by proteolytic cleavage (Rossmann *et al.*, 1983a) or by genetic engineering methods (Lokesh *et al.*, 2002) results in formation of T=1 particles, where all the contacts between subunits are bent. The N-terminus acts as a molecular switch, controlling the angles of the subunit contacts, which in turn control the curvature of the particle and allow successful assembly of the T=3 capsid.
The ordered N-termini of the C subunits of the sobemoviruses have been found to be arranged in two different ways. In SCPMV, SeMV, and TNV, the N-terminus folded back along the core of the C subunit and interacting with the subunits directly related by the threefold symmetry axis. In CfMV and RYMV, the N-terminal arm passes close to the icosahedral twofold axis toward the threefold icosahedral axis not directly adjacent to the core C subunit. Whichever the orientation of the N-terminus of C subunit is, it fulfils equivalent functions in particle size determination.
Hull (1977b) studied the particle stabilization of TRoV and other sobemoviruses and concluded that the particles of these viruses are stabilized by three types of bonds: pH-dependent interactions between subunits, protein-RNA interactions, and divalent-cation protein-protein bonds. The cations participating in formation of protein-protein bonds in the particle are calcium and magnesium ions. The major Ca^{2+}-binding site is on the quasithreefold axis between the A, B, and C subunits (Hermodson *et al.*, 1982). The sequence of TRoV shows that Glu194 is the ligand in methal binding site. In SBMV this residue is lysine (Mang *et al.*, 1982), suggesting a different mode of subunit association. The second Ca^{2+}-binding site is also between the quasi-threefold-related subunits, with interactions between Asp138 and Asp141 on one subunit and the main-chain carbonyls 199 and 259 on the other. The proposed Mg^{2+}-binding sites contain residues His132, Glu229, and Glu77 (Rossmann, 1984). The virus swells upon removal of cations at alkaline pH values (Hull, 1977b).
Brugidou et al. (2002) examined RYMV stability with anion-exchange chromatography and identified three markedly stable forms of RYMV particles from infected plants: first, an unstable swollen form lacking Ca^{2+} and dependent upon basic pH; second, a more stable transitional form lacking Ca^{2+} but dependent upon acidic pH; and third, a pH-independent, stable, compact form containing Ca^{2+}(Fig. 7). They demonstrated that particle stability increased over the time course of infection in rice plants: transitional and swollen forms were abundant during early infection (2 weeks postinfection), whereas compact forms increased during later stages of infection. Electron microscopy of infected tissue revealed virus particles in vacuoles of xylem parenchyma and mesophyll cells early in the time course of infection and suggested that vacuoles and other vesicles were the major storage compartments for virus particles. They proposed a model in

which virus maturation is associated with the virus accumulation in vacuoles. In this acidic compartment, virus particles may bind Ca^{2+} to produce a highly stable, compact form of the virus (Brugidou *et al.*, 2002).

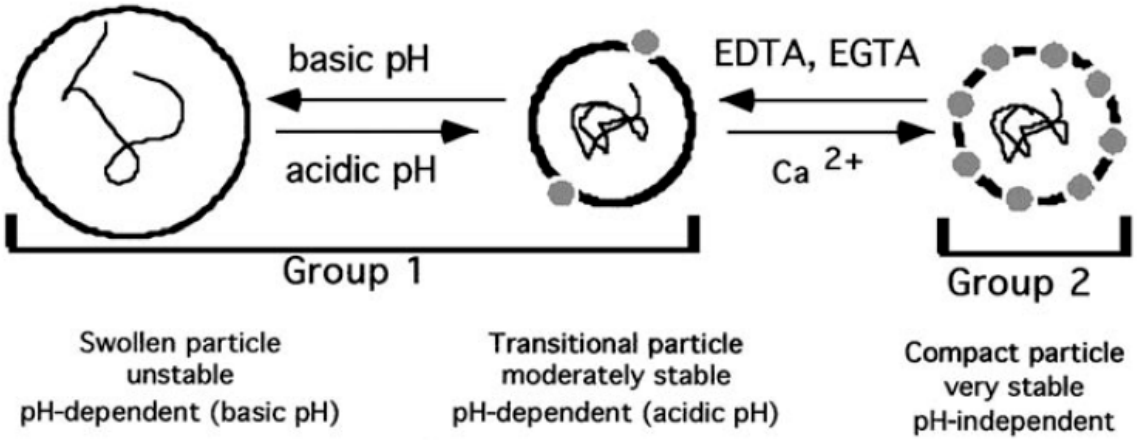

Figure 7. A schematic model for relationships between compact, transitional, and swollen RYMV particles. Group 1 (swollen and transitional) and 2 (compact) particles are separated by anion-exchange chromatography and can be interconverted by changes in pH and Ca2+ (Brugidou *et al.*, 2002).

In experiments of particle assembly with native CP it failed to assemble in either the T=1 or T=3 in the absence of RNA, suggesting that RNA-protein interactions initiates the formation of the T=3 particles. The interaction is made between viral RNA and the basic arm of the CP but no. specific CP-RNA interactions in SCPMV was observed. Dissociation of SBMV in a high-salt solution (0.4 M KCl) at a neutral pH yields a ribonucleoprotein complex (RNPC) composed of the viral RNA and about six CP subunits (Hsu *et al.*, 1977). Hacker (1995)demonstrated that the RNPC of SCPMV bind to a specific region of the viral RNA which could fold into a hairpin (Hacker, 1995). This specific CP binding site location was detected in the protease coding region in SCPMV ORF2. This region is the most highly conserved region of ORF2 among sobemoviruses. Unfortunately, these results do not directly demonstrate that CP binding to this region serves to nucleate SCPMV assembly.

As SeMV capsids are stabilized by several interactions like protein–protein, protein–RNA and calcium-mediated protein–protein interactions, and the N-terminal R domain has two important motifs, the arginine-rich motif (ARM) and the β-annulus structure for capsid formation a experiment of mutational analysis in this region was made. Mutation of all the arginine residues in the ARM almost completely abolished RNA encapsidation, although the assembly of T=3 capsids was not affected. A minimum of three arginine residues was found to be essential for RNA encapsidation. Thermal denaturation comparison of the mutant capsids devoid of RNA were and wild-type capsids were less stable. Those results show that capsid assembly is mediated by CP-dependent protein–protein inter-subunit interactions and encapsidation of genRNA so enhanced the stability of the capsids (Satheshkumar *et al.*, 2005b).

Experiments of mutation of a conserved proline within the segment that forms the β-annulus to alanine, or deletion of residues 48–53 involved in hydrogen bonding interactions with residues 54–58 of the 3-fold related subunit or deletion of all the residues (48–59) involved in the formation of β-annulus did not affect capsid assembly. These results shoved the β-annulus are not the switch in CP assembly into T=3. The role of the ordered β-annulus which was observed

in the structures of many viruses could be a consequence of assembly to optimize intersubunit interactions (Pappachan *et al.*, 2008).

For protein-protein interaction test mutation of a single Trp 170 to a charged residue (Lys or Glu) of SeMV CP was carried out. These mutations resulted in total disruption of virus assembly and resulted in the formation of stable soluble dimers. These Trp residues contribute to the hydrophobic association of the subunits at the icosahedral five-fold axes. The electrostatic repulsion between the five charged residues (Lys or Glu) resulting from the mutation was sufficient to disrupt the assembly (Pappachan *et al.*, 2009). It has been proposed that a pentamer of five dimers initiates the assembly of SeMV particles (Satheshkumar *et al.*, 2005b). The X-ray crystal structure of one of the isolated dimer mutants was determined at resolution 2.65 Å. As revealed detailed analysis of its structure a number of structural changes took place, especially in the loop and interfacial regions during the course of assembly. The isolated dimer was "more relaxed" than the dimer found in the T=3 or T=1 capsids. The isolated dimer does not bind Ca^{2+} ion and consequently four C-terminal residues are disordered. The FG loop, which interacts with RNA in the virus, has different conformations in the isolated dimer and the intact virus suggesting its flexible nature and the conformational changes that accompany assembly. The isolated dimer mutant was much less stable when compared to the assembled capsids, suggesting the importance of inter-subunit interactions and Ca^{2+} mediated interactions in the stability of the capsids (Pappachan *et al.*, 2009).

MATERIALS AND METHODS

Strains of *E.coli*

BL21 (DE3): B F⁻ *dcm ompT hsdS* ($r_k^- m_k^+$) *dcm*+ Tet^r *gal* λ (DE3) [*argU ileY leuW* Cam^r] (Stratagene)

XL 1: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F'*proAB lacI^{q}ZΔM15 Tn10* (Tet^r)] (Stratagene)

SURE: F⁻ *e14– (McrA–) Δ(mcrCB-hsdSMR-mrr)171 endA1 supE44 thi-1 gyrA96 relA1 lac recB recJ sbcC umuC::Tn5* (Kan^r) *uvrC* [F' *proAB lacI^{q}ZΔM15 Tn10* (Tet^r)] (Stratagene)

JM110: *rpsL* (Str^r) *thr leu thi*-1 *lacY galK galT ara tonA tsx dam dcm supE44 Δ*(*lac-proAB*) [F' *traD36 proAB lacIqZΔM15*] (Stratagene)

WK-6: Δ(*lac-proAB*), *galE*, *strA/F'laclq*, *ZΔM15*, *proA$^+$B$^+$* (provided by ASLA Ltd.)

Commercial plasmids

pACYC-RIL: ColE1-compatible, pACYC-based plasmid containing extra copies of the *argU*, *ileY*, and *leuW* tRNA genes, CAM^r (Stratagene)

pTZ57R: pre-digested temporary cloning plasmid, derivative of pTZ19R, contines *bla* gene Amp^r (Fermentas)

pET-Duet-1: designed for coexpression of two target genes in the same vector, Amp^r (Novagen).

pACYC-Duet-1: designed for coexpression of two target genes in the same vector, CAM^r (Novagen)

pCold-I: vector is designed to perform efficient protein expression at extremely low temperatures from cold-shock gene *csp*A, Amp^r (Takara)

pET-28a(+):carry an N-terminal His-Tag/thrombin/T7-Tag, Kan^r (Novagen)

pJET-1: pre-digested temprary cloning vector, Amp^r. As additional selection for recombinant cell selection is lethal gene *eco47IR*. It enables positive selection of the recombinants. (Fermentas)

In this work constructed plasmids

pTZ-1-FR (-2-FR, -3-FR, -4-FR, -5-FR): temporary plasmid. Contains RGMoV cDNS cloning fragments.

pTZ-P1: temporary plasmid. Contains RGMoV P1 gene coding sequence.

pTZ-Δ50SerP: temporary plasmid. Contains RGMoV protease gene coding sequence with deleted N-terminal 50 AA.

pJET-RGMoV-cDNA: plasmid contains full RGMoV cDNA copy.

pET-Du-Δ50SerP: expression plasmid contains RGMoV protease coding sequence with deleted N-terminal 50 AA.

pCold-Δ50SerP: expression plasmid containing RGMoV protease coding sequence with deleted N-terminal 50 AA for expression in low temperatures.

pCold-P1: expression plasmid. Contains RGMoV P1 coding sequence for expression in low temperatures

pET-Du-RdRpE/T1983: expression plasmid contains RGMoV RdRp coding sequence

pCold-RdRpE/T1983: expression plasmid. Contains RGMoV RdRp coding sequence for expression in low temperatures
pTZ-VPg: temporary plasmid. Contains RGMoV VPg gene coding sequence.
pACYC-Du-VPg: expression plasmid. Contains RGMoV VPg coding sequence.
pTZ-P16: temporary plasmid. Contains RGMoV P16 gene coding sequence.
pET-Du-P16: expression plasmid. Contains RGMoV P16 coding sequence.
pET-P16: expression plasmid. Contains RGMoV P16 coding sequence.

Methods used in this work

Microorganism cultivation

Table 2
Culture media used in this work.

	Type of media	Components
1.	2xTY (g/l)	Tryptone (Merck) 16 g Yeast extract (Merck) 10 g NaCl (SIGMA) 5 g
2.	LB agar (g/l)	Tryptone (Merck) 10 g Yeast extract (Merck) 5 g NaCl (SIGMA) 10 g Agar (SIGMA) 15 g
3.	SOC (g/100 ml)	Tryptone (Merck) 2 g Yeast extract (Merck) 0.5 g NaCl (SIGMA) 0.58 g KCl (SIGMA) 0.19 g $MgCl_2 \cdot 6H_2O$ (SIGMA) 0.203 g $MgSO_4$ (SIGMA) 0.12 g Glucose (SIGMA) 0.36 g

Growing of E.coli in liquid media for plasmid purification

Cells were grown in liquid 2xTY media (Table 2) in test-tube (volume 5 ml). Additional antibiotics were added, according to manufacturer protocol. Flasks with selected colonies were grown 16 h at +37°C in shaker AG CH-4103 (Infors) at 230 rpm.

Cultivation of E.coli on plates

Melt LB agar media (Table 2) add necessary volume of antibiotics. Plates incubate for 12 – 16 h at + 37ºC in thermostat.

Plasmid purification from E.coli (Miniprep)

Pour cells with a media in to 1.5 ml plastic test-tube. Centrifuge at 13.200 rpm on centrifuge S415D (Eppendorf) for 1 min. Discard supernatant and resuspend pellet in 300 µl lysis buffer P1 (Table 3). Add 300 µl of lyses buffer P2 (Table 3) mix gently for five times. Do not vortex! Allow to stay for 5 min at RT. Add 100 µl of $CHCl_3$. Mix gently for five times. Add 300 µl of P3 buffer (Table 3) mix gently. Centrifuge at 13.200 rpm on centrifuge S415D (Eppendorf) for 10 min. Transfer supernatant in to new test-tube, add 0.7 volume isopropanol (LACHEMA), and mix on vortex. Centrifuge at 13.200 rpm for 10 min. Discard supernatant, add 700 µl 75 % ethanol to pellet, vortex for 5 sec. Centrifuge at 13.200 rpm for 5 min. Discard supernatant and dry pellet at RT for 15 min. Add 50 µl of distilled H_2O or TE buffer, vortex.

Table 3
Plasmid extraction buffers

Buffer	Ingredients	Concentration
P1	TRIS-HCl (SIGMA)	50 mM pH 8.0
	EDTA (SIGMA)	10 mM
	RNase A (SIGMA)	0.1 mg/ml
P2	NaOH (SIGMA)	200 mM
	SDS (SIGMA)	1 %
P3	Potassium acetate	3 M, pH 5.2

Electrophoresis

Electrophoresis was carried out in 1xTBE buffer (1.8 g/l Tris Base (SIGMA), 5.5 g/l boric acid (LACHEMA), 0.76 g/l EDTA (SIGMA). All DNA samples were analyzed in 0.8% native agarose (USB) gel. DNA was visualized with ethidium bromide (Merck) in to UV-transiluminator 2011 MACROVUE (LKB).

Polymerase chine reaction (PCR)

Primers used in PCR reaction are summarized in to Table 4.

Table 4
Primers used in this work

Primer name	Primer sequence	Primer description
RG2-VPg-BamHI-F	5' GGATCC A AAC GGA GAG CAG GGA GCG CGC GAG A 3'	Used for VPg cloning in to pACYC-Duet-1.
RG2-VPg-HindIII-R	5' AAGCTT TCA TTC TTC ACG GGC GTC CCA ATC A 3'	Used for VPg cloning in to pACYC-Duet-1.
RG-SerP-BglII-R	5' AGATCT TCA TTC ACT GGA TTC ACA GTT TGC ATG GAA AAT GGA GA 3'	Used for Δ50SerP cloning in to pET-Duet-1.
RG-Δ50SerP-6H-F	5' CCATATG CGC GCTT GGC TAG CAA CCT CTC TGGA 3'	Used for ΔSerP cloning in to pET-Duet-1.
RG-P16-F	5'CCATATG TCC ACA GGA AAT GAT ATT CCT TTA AAC TGC CAG CA 3'	Used for P16 cloning in to pET-Duet-1.
RG-P16-BglII-R	5' AGATCT TCA AGC TGA GGG GGA CCC CTG GAC T 3'	Used for P16 cloning in to pET-Duet-1.
RG-P1-F-NdeI	5' CATATG CCT TCA GTG GTT ATC GAG GTT TGC TCAT 3'	Used for P1 cloning in to pCold-I.
RG2-P1-HindIII-R	5' AAGCTT TCA ATG ATG TCT AGT CCA AGA CTG CCC T 3'	Used for P1 cloning in to pCold-I.
RG-P1-BglII-R	5' AGATCT TCA ATG ATG TCT AGT CCA AGA CTG CCC T 3'	
RG-Pol-BglII-R	5' AGATCT TCA TTC TTC CGG GAT TTC TCC TTC A 3'	Used for cloning in to pET-Duet-1.
RG-Pol-E/T1983-NdeI-F	5'CATATG ACT GCT AGA ACG AGT ACT AAC GAG CTC T 3'	Used for cloning in to pET-Duet-1.
RG-1F	5'GTTAACAAATAGAGTTATTAAATTAACTCTATTGAACCCGT TATCCG3'	Used for RGMoV cDNA 1. fragment cloning.
RG-1R	5'TCCCATCGATGAGCACCTCTC3	Used for RGMoV cDNA 1. fragment cloning.
RG-2F	5'TATGGAGAGGTGCTCATCGATGGGA3'	Used for RGMoV cDNA 2. fragment cloning.
RG-2R	5'AGACAGCCTCCACAACCAAGGTCATGGA3'	Used for RGMoV cDNA 2. fragment cloning.
RG -3F	5'CCATGACCTTGGTTGTGGAGG3'	Used for RGMoV cDNA 3. fragment cloning.
RG-3R	5'TTCTTTACATCATGACCAAGTGCAT3'	Used for RGMoV cDNA 3.

		fragment cloning.
RG-4F	5'CTTGGTCATGAGTGTAAAGAATACTCAGCATGCGCCC3'	Used for RGMoV cDNA 4. fragment cloning.
RG-4R	5'TGACGCGTACTAAAGAGCCAGAGT3'	Used for RGMoV cDNA 4. fragment cloning.
RG-5F	5'CTCTTTAGTACGCGTCACATCCAGTGC3'	Used for RGMoV cDNA 5. fragment cloning.
RG-5R	5'AAGCTTACGAAAACCCCCGGGCCCCT3'	Used for RGMoV cDNA 5. fragment cloning.
RG-2seqF	5'ATGTGATATCGGTACAGGTCCCCGC3'	Used for RGMoV cDNA 2. fragment sequenation.
RG-2seqR	5'ACTCCAGACGATTCCGGACATGTCA3'	Used for RGMoV cDNA 2. fragment sequenation.
M13/pUC (-46) F	5'GCCAGGGTTTTCCCAGTCACGA3'	Sequenation primers for pTZ57 cloning vector.
M13/pUC (-46) R	5'GAGCGGATAACAATTTCACACAGG3'	Sequenation primers for pTZ57 cloning vector.

PCR reaction mix was made according to provided protocols of manufacturers (Fermentas or BioRad) depending on chosen polymerase. Reactions were carried out in PCR thermo-cycler Verity (Applied Biosystems) using following program steps:

1. Denaturation at + 95°C 2 min
2. Denaturation at + 95°C 30 sec
3. Primer and DNA matrices cleavage + 55°C 30 sec
4. DNA strand elongation + 72°C 1 min
5. Final elongation at +72ºC 5 min

Time and temperature in steps from 1 – 3 could be variable according to primer extension temperature, and chosen polymerase.

Restriction

Restriction was held in incubator according to necessary temperature of digestive enzyme activity. Digestive enzymes and reaction buffers were used from Fermentas according to manufacturer protocol.

Partial restriction enzyme digestion

Prepare restriction mix for first restrictase which cuts plasmid only one time. Digest plasmid for 3 h, according to used restrictase conditions. Inactivate restrictase by heating in recommended temperature. Add second restrictase 0.3 – 0.5 µl, incubate for 5 – 15 min according to used restrictase working conditions. Remove 33 µl of restriction reaction, inactivate by adding 5 µl of 6x Loading Dye. Continue restriction additional 15 min. Remove another aliquot of 33 µl, inactivate as previous. Leave the rest reaction volume additional 30 min. Inactivate it. Analyze and extract from 0.8 % native agarose gel.

Ligation

Ligation was carried out in 1.5 ml plastic test-tubes. The concentration of the vector and DNA fragment were set 1:3. T4 DNA-Ligase (Fermentas) was used for the joining of DNA molecules. Ligation was performed on ice: in first step was added vector; fragment, buffer and necessary volume of distilled H_2O (total volume 20 µl). Reaction mix was incubed for 2 min at + 62ºC, and then cooled on ice for additional 5 min. As last was added T4 DNA Ligase. Reaction was performed for 16 h.

Transformation of E.coli

Ligation was incubed for 10 min at + 65ºC before transferring to the cells. 10 µl of ligation mix was added to the competent cells (XL1, SURE or JM110). Cells were held on ice during incubation with ligation mix for 15 min than incubed for 40 sec at + 42ºC. Keep on ice for 5 min add warm 1 ml of SOC media (Table 2). Incubate cells for 1 h at + 37ºC.
Harvest cells by centrifugation, leave supernatant for 100 µl, resuspend cells and plate on LB agar plates with necessary antibiotics.

DNA fragment extraction from agarose gel

Excise gel slice containing the DNA fragment using a clean scalpel. Cut as close to the DNA as possible to minimize the gel volume. Place the gel slice into a pre-weighed 1.5 ml tube and weigh. Record the weight of the gel slice. If the DNA is extracted from a TBE agarose gel, add 1/10 of volume TBE Conversion Buffer and 4.5 volumes of Binding Buffer to a given volume of agarose. Incubate the gel mixture at + 55°C for 5 min or until the gel slice is completely dissolved. Mix the tube by inversion every few minutes to facilitate the melting process. Add 5 µl the resuspended Silica Powder Suspension to the DNA/Binding Buffer mixture. Incubate the mixture for 5 min at + 55°C to allow for binding of the DNA to the silica matrix. Mix by vortexing every few minutes to keep the silica powder in suspension. Spin the silica powder/DNA mixture for 5 s to form a pellet. Carefully remove the supernatant solution and discard. Add 500 µl of ice cold Washing Buffer, resuspend the pellet and spin for 5 sec. Discard the supernatant. Repeat this procedure three times. Air-dry the pellet for 10 - 15 min to avoid the presence residual ethanol in the purified DNA solution. Resuspend the pellet in 20 µl of sterile deionized water or TE and incubate the tube at + 55°C for 5 min. Spin the tube and remove the supernatant while avoiding the pellet. Place the recovered supernatant into a fresh tube.

RNA isolation from plant material with TRI REAGENT

Weight plant material, add 10 volume of TRI REAGENT™ (SIGMA) and homogenize material. Incubate at room temperature for 5 min. Add 0.2 ml of chloroform for every 1 ml of TRI REAGENT™ , vortex vigorously for 15 sec. Incubate at RT for 5 min. Spin for 15 min at 13.200 rpm at + 4ºC or RT. Carefully remove upper clear faze and transfer to new test-tube. Add 0.5 ml of isopropanol (LACHEMA) (for each 1 ml of TRI REAGENT™). Spin for 5 min at 13.200 rpm, RT. Discard supernatant, and wash pellet with 75 % ethanol 1 ml per 1 ml used TRI REAGENT™. Spin for 5 min at 13.200 rpm. Discard supernatant and dry pellet with air dryer for 5 min at RT. Dissolve pellet in 20 µl of distilled H_2O. For better RNA dissolvation, incubate for 10 min at + 65ºC, and pipette.

RGMoV cDNA construction

For RGMoV cDNA obtaining RT-PCR was used. cDNA was created from five RT-PCR fragments with overlapping ends for cloning in to full sequence. For full method see original publication **I**.

Reverse transcription reaction

For reverse transcription reaction (RT) all components were from Fermentas. After thawing, mix and briefly centrifuge the components of the RT. All components store on ice.
Add the following reagents into a sterile, nuclease-free tube on ice in the indicated order:
Template RNA:
Total RNA 0.1-5 µg (poly(A) mRNA 10 ng - 0.5 µg; specific RNA 0.01 pg - 0.5 µg)
Primer:
oligo $(dT)_{18}$ 0.5 µg (random hexamer 0.2 µg; *or* gene-specific 15 – 20 pmol)
DEPC-treated water fill up to 11 µl
Incubate the mix at + 70°C for 5 min and chill on ice, spin down and place the tube back on ice.
Add the following components in the indicated order:
5X Reaction Buffer 4 µl
RiboLock™ RNase Inhibitor (20 u/µl) 1 µl
10 mM dNTP Mix 2 µl
DEPC-treated water fill up to 19 µl
Incubate at + 37ºC for 5 min. If random primer is used, incubate at + 25ºC for 5 min. Add 200 units of M-MuLV Reverse Transcriptase. Incubate reaction mixture, containing oligo $(dT)_{18}$ or sequence-specific primer at + 42ºC for 60 min. If using random hexamer primer, incubate at + 25ºC for 10 min and then at + 42ºC for 60 min. The reaction terminates by heating at + 70ºC for 5 min. The reverse transcription reaction product can be directly used in PCR applications or stored at -20ºC for less than one week. For longer storage, - 70ºC is recommended.

PCR product cloning in pTZ57R/T vector

For PCR product created with *Taq* polymerase pTZ57 vector from InsT/Aclone™ PCR Product Cloning Kit (Fermentas) was used. Vector pTZ57R/T is already in cut form. Restriction is performed with *Eco*32I and treated with terminal deoxynucleotidiltransferase, creating 3’ ddT overlap at both ends, for rapid cloning of PCR products with 3’ ddA overlap. Following reaction mix was made:

0.5 µl pTZ57R/T (vector)
2 µl PCR product (insert)
1 µl 10x BSA
1 µl 10x Ligation buffer
1 µl PEG 4000
H_2O fill up to 10 µl
Total volume: 10 µl

PCR product cloning in pJET1 vector

For PCR product created with proofreading DNA polymerase which creates blunt-end PCR products pJET1 (Fermentas) vector was used. Following reaction mix was made:
2X Reaction Buffer 10 µl
PCR product (non-purify needed) 1-2 µl
pJET1/blunt Cloning Vector (50 ng/µl) 1 µl
H_2O fill up to 19 µl
T4 DNA Ligase 1 µl
Total volume: 20 µl
Vortex reaction mix briefly and centrifuge for 3-5 sec. Incubate the ligation mixture at room temperature (+ 22°C) for 5 min. Incubation time can be extended up to 30 min if the maximal number of transformants is required. Use the ligation mixture directly for bacterial transformation.

DNA sequencing

For sequenation was used a BigDye cycle sequencing kit and an ABI Prism 3100 Genetic analyzer (Applied Biosystems). Following reaction mix was made:

Big Dye reaction mix – 2 µl
(Contain Big Dye primers, dNTP's, AmpliTa q in reaction buffer)

5x reaction buffer pH 9.0 – 3 µl
Primer – 1.6 pmol
DNA of plasmid – 250 ng
H_2O fill up to 20 µl

5x reaction buffer pH 9.0:

400 mM TRIS-HCl pH 9.0
10 mM $MgCl_2$

PCR program (30 cycles):

1. Heating till + 96°C
2. + 96°C 10 sec
3. Cooling till + 50°C (according to used primer extinction temp.)
4. + 50°C 5 sec
5. Heating till + 60°C
6. + 60°C 4 min

TheClustalW 1.83 programm on the EBI server was used for the multiple sequence analysis.

DNA extraction after sequenation PCR

Add 1/10 of volume KAc 3 M pH 5.2 and 50 µl of ice cold 96 % ethanol to the PCR reaction. Incubate at – 20°C for 20 min. Spin 15 min at 13.200 rpm. Discard supernatant, and wash pellet with 200 µl of cold 75 % ethanol. Spin down for 5 min at 13.200 rpm. Carefully remove supernatant with pipette. Dry pellet for 5 min at +65°C.

Cell cultivation for protein purification

E.coli growth conditions were chosen according to recommended manufacturer protocol for used expression system. Cells were cultivated until OD reached 1.0, induced with 5 mM IPTG (Fermentas), then the cultivation was continued for additional 16 h. Cells were harvested by centrifugation and frozen in -20ºC before further treatment. For detailed growth conditions see original publication Nr. **II**.

Cell disruption with ultrasonic

Harvested cells were defrozen on ice, add and resuspended in necessary volume of buffer which contain 1xPBS, 15 mM KH_2PO_4 or 15 mM TRIS-HCl pH 8.0, 5 mM Et-SH, 1 mM PMSF (3 ml per 1 g of cells). Cells sonicate in 50 ml test-tube for 2 min with intensity 60 %, with period 0.5 on ice in ultrasonicator Soniprep 150 (MSE). Spin down sonicated cells for 30 min at 13.000 rpm. Transfer supernatant to a fresh test-tube. Store at +4°C. It is ready for further treatment and purification.

Virus production

Virus material (MAFF No. 307043) was received from the MAFF Genebank at the National Institute of Agrobiological Sciences, Japan. Oat plants cv. "Jaak" were used as host plant. Oat plants were cultivated at laboratory conditions. For detailed growth conditions see original publication Nr. **III**.

Virus material extraction from plant material

Homogenize harvested plant material after it was frothed at -70°C in 15 KH_2PO_4 buffer (pH 5.5) with masher. Remove plant over lefts by filtration thru the nylon filter, add to the supernatant 0.4 volumes $CHCl_3$ place on shaker for 30 min at RT. $CHCl_3$ remove by centrifugation at 4.000 rpm for 15 min. Carefully remove upper colorless phase to the new tube, add $(NH_4)_2SO_4$ till 30 % saturation, place on to magnetic agitator for 1 h at RT. Spin down for 20 min at 13.000 rpm. Discard supernatant, dissolve pellet in to 15 mM KH_2PO_4 pH 5.5 and dialysis against 15 mM KH_2PO_4 pH 5.5 for 16 h at +4°C. Dialyzed material concentrates with ultracentrifugation for 1 h at 72.000 rpm, + 4°C. Discard supernatant and dissolve pellet in 15 mM KH_2PO_4 pH 5.5 and two times purify thrue 20 % sucrose cushion with 15 mM KH_2PO_4 pH 5.5. For RGMoV purification for crystallography see publication Nr. **III**.

SDS-PAGE

For protein detection in samples 12% SDS-PAGE was made.

Upper SDS-PAGE (4 %):

dH_2O	3.0 ml
TRIS-HCl pH 7.0 (0.5 M)	1.25 ml
PAA:	0.6 ml
30 %	acrylamide
0.8 %	N,N'-methilen-bis-akrylamide
10 % AP	26.6 µl
10 % SDS	0.6 ml
TEMED	16.6 µl

Lower SDS-PAGE (12 %):

dH_2O	5.0 ml

TRIS-HCl pH 8.8 (1.5 M)	3.0 ml
PAA	4.0 ml
10 % AP	100 µl
10 % SDS	1.2 ml
TEMED	10 µl

Protein visualization by Coomassie Brilliant blue G-250

Plate SDS-PAGE in to plastic box; add wash solution (10 % ethanol, 10 % acetic acid). Place on shaker (BioSan) for 10 min at 80 rpm. Discard wash buffer and add coomassi blue solution. Heat for 15 sec in to microwave (Electrolux), pace on to shaker for additional 5 min. Discard coomassi blue solution and wash three times for 10 min with wash solution.

Western blot

Prepare six Watman paper and one hybridization nitrocellulose membrane 0.45 micron (Amerchame) in 5.5 x 8.2 cm dimensions. Moisten positive plate of semi-dry blotting system (The W. E. P. Company) with 1 ml of transfer buffer (1.4 % glycine (SIGMA), 3.1 g/l TRIS base, 25 % methanol). Than plate three wet Watman paper one to each other, avoid of air bubbles. On to Watmen papers place wet nitrocellulose membrane, again be sure to avoid air bubbles. Place SDS-PAGE on to membrane, smooth out. Place three remaining Watman papers on to SDS-PAGE, before placing moisten in to transfer buffer, smooth out. Moisten negative plate of blotting system. Set current intensity - for each 1 cm^2 use 1 mA (5.5 x 8.2 = 45 mA), incubate 45 min.

Western blot prestaining with antibodies

Place membrane in to 50 ml tube and add 5 ml of Blocking reagent (1 % of milk powder (Amercham); 0.3 % Bovine serum albumin (BSA) (SIGMA), 1x PBS buffer).
Place tube in to agitator for 1 h. Discard Blocking reagent. Add 10 ml of 1 x PBS and add specific antibodies, for example, rabbit-IgG against RGMoV CP (supplied by PhD. D.Skrastina) in dilatation 1:500. Incubate it for 16 h at + 4°C. Wash three times with 1x PBS for 5 min, RT. After third wash add 5 ml of 1x PBS and secondary antibodies against rabbit IgG (SIGMA) (1:1000). Incubate for 3 h at RT. Wash three times with 1x PBS for 15 min at RT.

Western blot visualization

After third wash remove membrane from tube and place in to nontransparecy container and add 50 ml of incubation mix (1 mg o-dianisidine (SIGMA), 2.5 ml 96 % ethanol (Rigas Balzams), 10 mM TRIS-HCL pH 7.0 (SIGMA), 150 mM NaCl (SIGMA), 50 µl H_2O_2 (SIA UNIKEMS), total volume 50 ml). Incubate it for 10 – 30 min in darkness (!).

Native RGMoV and protein analysis by mass spectrometer MALDI TOF

Protein (1 mg/ml) molecular mass was analyzed by autoflex MALDI TOF mass spectrometer (BRUKER DALTONICS, Germany) using 2, 5-dihydroxyacetophenone matrix (BRUKER DALTONICS, Germany). For precise description of these softwears see original publication **II.**

P16 RNA binding analysis in native agarose gel – gel shift

P16 analysis of RNA binding was carried out in 0.8 % agarose gel. In 1.5 ml test-tube combine 1 µg of nucleic acid and 25 µg of P16 in total volume of 15 µl Buffer L (10 mM TRIS-HCl pH8.0; 200 mM NaCl; 1 mM EDTA; 10 % glycerol; 1 mM PMSF; 1 mM DTT). Incubate on ice for 30 min. Add 4 µl of 5x loading buffer, mix and load on to gel. Perform electrophoresis. Visualize gel with ethidium bromide (Merck, Germany) in to UV-transiluminator 2011 MACROVUE (LKB, Sweden).

Protein Bioinformatics analysis

For protein bioinformatics analysis was used several software programs – PONDR, Predictor of Natural Disordered Regions, is available at Molecular Kinetics Inc. (Dunker *et al.*, 2002), CBS Prediction Servers TMHMM Server v. 2.0, Software FoldIndex© was used to identify protein folded and unfolded regions (Prilusky *et al.*, 2005). For precise description of thise softwears see original publication **II.**

Protein crystallization

JCSG-plus HT-96 and Structure screen I + II HT-96 protein crystallization solution kits (Molecular Dimensions) were used to find the optimum conditions for VPg and P16 crystal formation. To identify crystallization conditions for VPg, the sample was sent to high-throughput screening facility located at the Hauptman-Woodward Medical Research Institute, USA (Luft *et al.*, 2003). For precise description on this method see original publication **II**.
For RGMoV crystallization the hanging drops technique with a bottom solution containing: 300 mM Li2SO4, 50 mM Na–citrate at pH 3.0, 1% PEG 8000 was used. The drops were prepared by mixing 1.5 µl of bottom solution with an equal volume of virus solution (8 mg/ml).

I THE *RYEGRASS MOTTLE VIRUS* GENOME CODES FOR A SOBEMOVIRUS 3C-LIKE SERINE PROTEASE AND RNA-DEPENDENT RNA POLYMERASE TRANSLATED VIA -1 RIBOSOMAL FRAMESHIFTING

The original publication should be cited as follows:
Balke I., Resevica G., Zeltins A. 2007. The *Ryegrass mottle virus* genome codes for a sobemovirus 3C-like serine protease and RNA-dependent RNA polymerase translated via -1 ribosomal frameshifting. *Virus Genes* 35:395-398

Virus Genes
DOI 10.1007/s11262-007-0087-y

The *Ryegrass mottle virus* genome codes for a sobemovirus 3C-like serine protease and RNA-dependent RNA polymerase translated via –1 ribosomal frameshifting

Ina Balke · Gunta Resevica · Andris Zeltins

Received: 1 November 2006 / Accepted: 15 February 2007

Abstract In the course of sobemovirus gene cloning the complete genome of *Ryegrass mottle virus* (RGMoV) was sequenced. Sequence analysis revealed differences including missing and extraneous nucleotides in comparison to the previously published sequence (Zhang, Toriyama, Takanashi, J. Gen. Plant Pathol. **67**, 63 (2001)). A gene coding for a typical sobemovirus 3C-like serine protease was identified in ORF2a after multiple sequence alignment analysis. The newly identified 57-amino-acid stretch in ORF2a showed similarities ranging from 38.5 to 50.9% among sequenced genes of sobemovirus proteases. ORF analysis of the RGMoV polyprotein coding sequence demonstrated the arrangement of ORF2b coding for RNA-dependent RNA polymerase (RdRP) in the –1 frame in regard to ORF2a. The localization of conserved among sobemoviruses slippery sequence (UUUAAAC) at the 3′-end of ORF2a suggests the translation of RdRP via a –1 ribosomal frameshifting mechanism, allowing to include the RGMoV in the sobemovirus group with a Cocksfoot mottle virus-like (CfMV-like) genome organization.

Keywords Sobemovirus · Ryegrass mottle virus · Genome organization

Ryegrass mottle virus (RGMoV) is an isometric plant virus and represents a particle of 28 nm in diameter. It belongs to the genus *Sobemovirus* and infects *Gramineae* plants, including *Lolium multiflorum* and *Dactylis glomerata* [1]. To date, complete genome sequences of nine *Sobemovirus* species have been published: *Southern bean mosaic virus* (SBMV), *Southern cowpea mosaic virus* (SCPMV), *Lucerne transient streak virus* (LTSV), *Rice yellow mottle virus* (RYMV), *Sesbania mosaic virus* (SeMV), *Cocksfoot mottle virus* (CfMV), *Subterranean clover mottle virus* (SCMoV), *Turnip rosette virus* (TRoV), and RGMoV (for reference see Fig. 1). These viruses contain single-stranded positive-sense RNA in size varying between 4037 for TRoV and 4457 nucleotides for RYMV. The sobemovirus monopartite genomes mostly contain four predicted open reading frames (ORFs), a part of which overlap [2–4]. The 5′-ends of genomes start with an untranslated region, followed by short ORF1, that code for 109–186 residue-long sobemovirus proteins with low similarity to each other [4]. The ORF1 coded proteins have been suggested to have a role in systemic infection [5] and in suppression of post-transcriptional gene silencing [6]. The 3′-proximally located sobemovirus ORFs code for coat proteins (CP) with a comparably low sequence identity (16–31%). Regarding three- dimensional structure, icosahedral sobemovirus particles encapsulating the viral RNA are built from 180 subunits of the jelly-roll β-sandwich folded CP and have $T = 3$ quasisymmetry [7].

The analysis of sobemoviral sequences in the central part of the genomes demonstrated long ORFs coding for polyproteins with a predicted molecular mass of approximately 100 kDa. The first sequencing data of SBMV, SCPMV, LTSV, SeMV, RYMV and the analysis of genome organization suggested the formation of corresponding polyproteins from single continuous ORF coding for virus protease (Pro), viral genome-linked protein (VPg), and RNA-dependent RNA polymerase (RdRP). For all of these

The nucleotide sequence data reported in this paper have been submitted to the Genbank nucleotide sequence database and have been assigned the accession number EF091714.

I. Balke · G. Resevica · A. Zeltins (✉)
Protein Engineering, Latvian Biomedical Research and Study Centre, Ratsupites 1, Riga 1067, Latvia
e-mail: anze@biomed.lu.lv

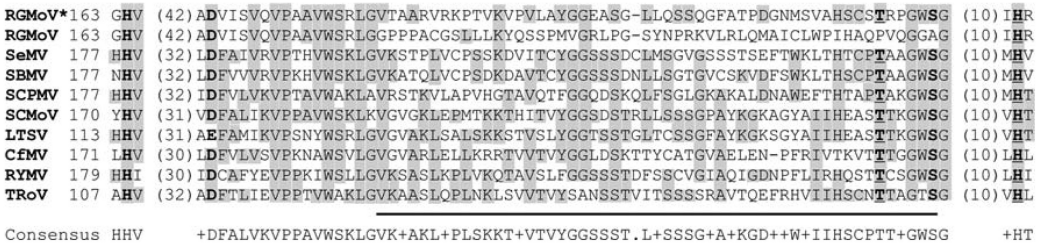

Fig. 1 RGMoV ORF2a sequence analysis, Figures in parentheses denote the distances in AA between motifs. AA residues in bold show the putative catalytic triad of 3C-like serine proteases, bold underlined residues denote possible amino acids important for substrate-binding. The line under alignment shows the region of newly identified amino acids in RGMoV ORF2a. The corresponding amino acids from the ClustalW calculated consensus sequence are highlighted. Following Genbank sequences were used for the alignment: RGMoV*- sequence of this work (EF091714), RGMoV (AB040446), SeMV (AY004291), SBMV (AF055887), SCPMV (M23021), SCMoV (AF208001), LTSV (NC_001696), CfMV (DQ680848), RYMV (AJ608215), TRoV (AY177608)

viruses it was supposed that additional proteins with an unknown function are translated from a nested ORF3 (P_{ORF3}) via the –1 ribosomal frameshifting mechanism (Pro-VPg-P_{ORF3}; for review see Ref. [2]). This type of genome organization has been denoted as a SCPMV-type. The RGMoV genome organization has been reported to belong to this type as well (Fig. 2B). For other sequenced sobemovirus genomes (CfMV, SCMoV, TRoV), a different, CfMV-type polyprotein coding strategy has been found. Instead of large polyprotein genes with nested ORF3 the genomes of these viruses contain two overlapping ORFs, and putative RdRP's are formed as a C-terminal part of polyprotein via –1 ribosomal frameshifting [2–4]. The latest sequencing of RYMV [8] and SeMV (Genbank No. AY004291) revealed errors in earlier published sequences. Removing an incorrectly detected nucleotide at 5′ part of the RdRP coding region allowed to assign the polyprotein expression strategy as a CfMV-like also for these sobemoviruses.

In the course of cloning of sobemovirus genes with the aim to construct the infectious cDNA and to characterize corresponding viral proteins, we sequenced the complete genome of *Ryegrass mottle virus* (National Institute of Agrobiological Sciences, Japan, MAFF No. 307043). Virus RNA was isolated from infected oat leaves using TRI-reagent (Sigma). For some experiments, virus particles were used as a RNA source; the purification protocol for RGMoV was the same as published earlier [7]. Purified RNA was used as a template for RT-PCR using M-MuLV H(–) reverse transcriptase and Taq polymerase (Fermentas). Following genome-specific primer pairs derived from RGMoV sequence data (nucleotide sequences and positions accordingly to Genbank. No. AB040446) were used to produce nine overlapping PCR fragments:

(1) RG-1F (nt pos. 1–43), 5′-ACAAATAGAGTTATTA AATTAACTCTATTGAACCCGTTATCCG-3′, RG-1R, (nt pos.919–899) 5′-TCCCATCGATGAGCACC TCTC-3′;

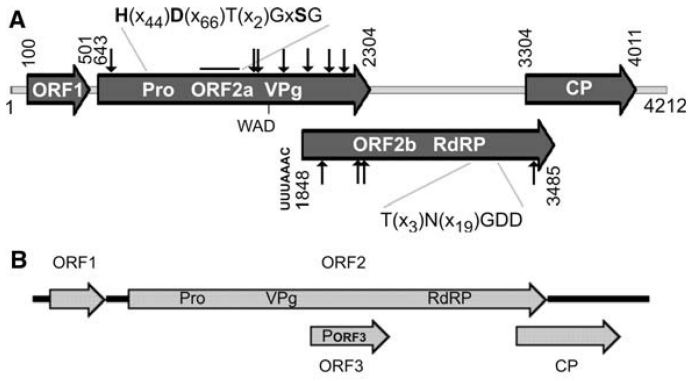

Fig. 2 The ORF map of RGMoV genome. (**A**) RGMoV genome organization according to sequencing results (EF091714). The numbers denote the nt positions in the RGMoV sequence. The line on the top of ORF2a indicate the relative localization of newly found AA sequences in the map. Important features of the genome and coded proteins are shown: consensus sequences for 3C-like serine protease, for VPg in ORF2a, and for RdRP in ORF2b, and the ribosomal frameshifting signal at the end of ORF2a (UUUAAAC). Arrows denote the localization of putative E/N, E/T and E/S protease cleavage sites. (**B**) Previously reported genome organization (Ref. 1, AB040446)

(2) RG-2F, (nt pos.895–919) 5′-TATGGAGAGGTGCTCATCGATGGGA-3′, RG-2R, (nt pos.2365–2338) 5′-AGACAGCCTCCACAACCAAGGTCATGGA-3′;
(3) RG-3F, (nt pos.2339–2359) 5′-CCATGACCTTGGTTGTGGAGG-3′, RG-3R, (nt pos.3060–3035) 5′-TTCTTTACACTCATGACCAAGTGCAT-3′;
(4) RG-4F, (nt pos.3040–3076) 5′-CTTGGTCATGAGTGTAAAGAATACTCAGCATGCGCCC-3′, RG-4R, (nt pos.3527–3504) 5′-TGACGCGTACTAAAGAGCCAGAGT-3′;
(5) RG-5F, (nt pos.3511–3537) 5′-CTCTTTAGTACGCGTCACATCCAGTGC-3′, RG-5R, (nt pos.4210–4191) 5′-ACGAAAACCCCCGGGCCCCT-3′;
(6) RG-CP-F (nt pos.3305–3332, introduced NcoI site underlined) CATCCATGGCAAGGAAGAAGGGCAAATCGGCCA, RG-CP-R (nt pos.4203–3998, introduced HindIII site underlined) TGTAAGCTTGAGCTCAGGTCTCACTGGTTGATTGT;
(7) RG-1F; RG2R;
(8) RG-2F; RG-3R;
(9) RG-3F; RG-4R.

Each genome fragment was amplified from at least two independent RNA isolations and reverse transcriptions. All PCR fragments were cloned into pTZ57R vector (Fermentas) and sequenced, using a BigDye cycle sequencing kit and an ABI Prism 3100 Genetic analyzer (Applied Biosystems). To lower the influence of RT-PCR errors on final sequencing data as much as possible, several plasmid clones from each RT-PCR were sequenced. Only the changes detected in at least 4 sequenced clones from independent reverse transcriptions were accepted as relevant differences in comparison to published sequence. The ClustalW 1.83 programm on the EBI server was used for the multiple sequence analysis. The sobemovirus sequences were obtained from Genbank searches (accession numbers indicated in the legend of Fig. 1).

Altogether five differences in the RGMoV sequence were found within polyprotein coding region. In the AA sequence of putative protease the detected G at pos. 838 results in substitution of lysine by aspartate, whereas the C at pos. 1317 in our derived sequence does not change the glycine (nucleotide positions accordingly to Genbank. No. EF091714). Also, we determined an additional *T* at pos. 1319, and missing G nucleotides at positions 1477 and 2209 in contrast to the previously published sequence [1]. ORF analysis demonstrated a completely different 52 AA stretch between nucleotide positions 1319 and 1477 and a termination codon at position 2302. Multiple AA sequence aligment including the data from all sequenced sobemoviruses and the newly obtained sequence showed a typical 3C-like serine protease AA arrangement (Fig. 1). A putative catalytic triad with a consensus AA sequence for sobemovirus proteases was identified as $H(x_{32-44})D/E(x_{66-67})TxxGxSG$. In the case of RGMoV protease, the AA linker between H and D was found to be longer than for other sobemoviruses, but it was still in a good agreement with 3C-like proteases from other organisms and viruses [9]. Our work demonstrates that the RGMoV 2a gene codes for all three catalytic triad amino acids, whereas in the previous sequence the triad serine is missing. The presence of alanine instead of the characteristic substrate-binding serine or cysteine caused uncertainty regarding the catalytic property of the protease domain in RGMoV [1]. Recently published SeMV serine protease crystallography data demonstrated the spatial arrangement of catalytic amino acids H, D, S, and in vitro experiments using the corresponding mutants confirmed the involvement of these AA in processing of SeMV polyproteins [10, 11].

Interestingly, the sobemovirus proteases including RGMoV contain also strongly conserved amino acids typical for substrate-binding (threonine and histidine), similarly to other 3C-like proteases [9]. In the previous RGMoV sequence, proline was identified instead of threonine at the position of the putative substrate-binding site (Fig. 1). The importance of H298 and T279 in substrate-binding has been recently experimentally demonstrated for the E/T or E/S cleavage in SeMV serine protease [11].

Pairwise alignment of the newly detected RGMoV protease AA sequence stretch showed comparably high similarity to corresponding regions of other sobemovirus proteases ranging from 38.5% for CfMV and 50.9% for SBMV. The same alignment using previous RGMoV data resulted only in approximately 10–17% similarity.

Another important change in genome organisation and protein synthesis strategy of RGMoV was found after genome position 2209. The detected six G stretch instead of seven at the pos. 2208, shifted the RdRP coding sequence in to the –1 frame in regard to ORF2a. Therefore, according to our data, the RGMoV genome codes for two polyproteins. RGMoV ORF2a codes for protease and putative VPg (Pro-VPg). ORF2b appears to code for RdRP, which can be translated via the –1 ribosomal frameshifting mechanism at strongly conserved [1–4] sobemovirus heptanucleotide slippery sequence (UUUAAAC; nt pos. 1842–1848; Fig. 2A). As a result, a polyprotein can be synthesized, where the RdRP is fused to the C-terminal part of the polyprotein containing protease and VPg (Pro-VPg-RdRP). An ORF3 coding for a protein with an unknown function (P_{ORF3}) in SCPMV-type sobemoviruses has not been found in the RGMoV genome.

Earlier multiple AA sequence alignments conducted by Dwyer et al. [3] surprisingly demonstrated similarity of *N*-terminal part of RdRPs of CfMV-like sobemoviruses to unknown ORF3 products of SCPMV-type sobemovirus genomes, including RGMoV, RYMV, and SeMV.

Similarly to our sequencing data, recently published corrected sequences of RYMV [8] and SeMV (Genbank No.AY004291) demonstrated excessive nucleotides in ORF2, which have led to wrongly interpreted genome organization in early publications [2]. The new findings indicate that resequencing of remaining sobemovirus cDNAs would be necessary to test the possible localization of RdRP gene in to the –1 frame in regard to protease and VPg coding sequence.

RGMoV and other sobemovirus genomes contain frameshift slippery sequence and downstream-located ATG codons (distances between 42 and 108 nt). An extreme long stretch of 774 nt has been detected in TRoV and a different start codon (CUG) has been suggested as a translation iniciation codon [4]. However, mature sobemovirus RdRPs are possibly formed as a result of proteolytic processing of corresponding polyproteins [10, 12] and not as separate proteins translated from ATG.

As shown from investigation of the formation of mature VPg's and RdRP, sobemovirus 3C-like serine proteases cleave virus polyproteins between glutamic acid and threonine [10, 13], asparagine [12], or serine residues [10]. Several putative E/T, E/N and E/S cleavage sites were identified also in RGMoV polyproteins (Fig. 2A), suggesting the formation of mature VPg and RdRP by proteolysis. The smallest possible RGMoV protein including the VPg consensus sequence (WAD) between E/N and E/S sites would be a 11.4 kDa protein, which is in a good agreement with the detected size (12 kDa) for VPg covalently bound to CfMV genomic RNA [12]. The AA sequence of RdRP contains only 4 putative E/T or E/S cleavage sites close to *N*- or *C*-termini. Assuming that all cleavage sites are accessible in mature RGMoV the completely processed protein would be 42.5 kDa in size, which is comparable to 52 kDa of processed RdRP of SeMV [10].

In conclusion, additional experiments are needed to confirm the adequacy of the cloned cDNA to the RGMoV RNA sequence. In the near future we plan to test the ability of the obtained cDNA to code for infectious viral cDNA and functionally active RGMoV proteins. Our analysis of the RGMoV genome sequence revealed the following important differences to previously published data: (i) virus RNA codes for a typical sobemovirus 3C-like serine protease with strongly conserved triad and substrate-binding amino acids; (ii) RGMoV genome analysis demonstrate the possible RdRP translation via –1 ribosome frameshifting mechanism; (iii) comparisions with other sobemovirus genomes allow to classify the genome organization of RGMoV as a CfMV-like.

A paper demonstrating the sobemovirus common CfMV-like genomic organization (M. Meier, E. Truve. Arch Virol (2006) DOI 10.1007/s00705-006-0867-z) appeared during the reviewing process of this article.

Acknowledgements We thank Prof. P. Pumpens for fruitful discussions and critical reading of the manuscript. This work was supported by Latvian Science Council grant No. 04.1149 and EU Project No.VPD1/ERAF/CFLA/05/APK/2.5.1./000031/016.

References

1. F. Zhang, S. Toriyama, M. Takanashi, J. Gen. Plant Pathol. **67**, 63 (2001)
2. T. Tamm, E. Truve, J. Virol. **74**, 6231 (2000)
3. G.I. Dwyer, R. Njeru, S. Williamson, J. Fosu-Nyarko, R. Hopkins, R.A.C. Jones, Arch. Virol. **148**, 2237 (2003)
4. A.S. Callaway, C.G. George, S.A. Lommel, Virology **330**, 186 (2004)
5. M. Meier, H. Paves, A. Olspert, T. Tamm, E. Truve, Virus Genes **32**, 321 (2006)
6. O. Voinnet, Y.M. Pinto, D.C. Baulcombe, Proc. Natl. Acad. Sci. USA **96**, 14147 (1999)
7. K. Tars, A. Zeltins, L. Liljas, Virology **310**, 287 (2003)
8. D. Fargette, A. Pinel, Z. Abubakar, O. Traore, C. Brugidou, S. Fatogoma, E. Hebrard, M. Choisy, Y. Sere, C. Fauquet, G. Konate, J. Virol. **78**, 3252 (2004)
9. E.J. Snijder, A.L.M. Wassenaar, L.C. van Dinten, W.J.M. Spaan, A.E. Gorbalenya, J. Biol. Chem. **271**, 4864 (1996)
10. P.S. Satheshkumar, G.L. Lokesh, H.S. Savithri, Virology **318**, 429 (2004)
11. P. Gayathri, P.S. Satheshkumar, K. Prasad, S Nair, H.S. Savithri, M.R.N. Murthy, Virology **346**, 440 (2006)
12. K. Mäkinen, K. Mäkeläinen, N. Arshava, T. Tamm, A. Merits, E. Truve, S. Zavriev, M. Saarma, J. Gen. Virol. **81**, 2783 (2000)
13. F. van der Wilk, M. Verbeek, A. Dullemans, J. van den Heuvel, Virus Genes **17**, 21 (1998)

II EXPRESSION AND CHARACTERIZATION OF THE *RYEGRASS MOTTLE VIRUS* NON-STRUCTURAL PROTEINS

The original publication should be cited as follows:
Ina Baļķe, Gunta Resēviča, Ieva Kalnciema, Dace Skrastiņa, and Andris Zeltiņš. Expression and characterization of the Ryegrass mottle virus non-structural proteins. Proc. Latvian Acad. Sci., Section B /manuscript/

Expression and characterization of the *Ryegrass mottle virus* non-structural proteins

Ina Baļķe, Gunta Resēviča, Ieva Kalnciema, Dace Skrastiņa, and Andris Zeltiņš

Latvian Biomedical Research and Study Centre, Ratsupites Str. 1, Riga, LV-1067, Latvia
E-mail: anze@biomed.lu.lv

Communicated by Pauls Pumpēns

Ryegrass mottle virus (RGMoV) single stranded RNA genome is organized into four open reading frames (ORF) which encodes several proteins: ORF1 encodes a protein P1, ORF2a contains a membrane-associated 3C-like serine protease, genome linked protein VPg and the P16 protein gene. ORF2b encodes replicase RdRp, and the only structural proteins, coat protein which is synthesized from ORF3. To obtain the non-structural proteins in preparative quantities and characterize them, the corresponding RGMoV gene cDNAs were cloned in pET- and pColdI-derived expression vectors and expressed in several E. coli host cells. For protease and RdRP, the best expression system containing pColdI vector and E.coli WK6 strain was found. VPg and P16 proteins were obtained from the pET or pACYC vectors and E. coli BL21 (DE3) host cells and purified using Ni- Sepharose affinity chromatography. Attempts to crystallize VPg and P16 were unsuccessful, possibly due to non-structured amino acid sequences in both protein structures. Methods based on bioinformatic analysis indicated that the entire VPg domain and the C-terminal part of P16 are unstructured proteins, which probably prevented the formation of crystals.

INTRODUCTION

Plant virus genomes encode several types of proteins involved in various activities of the virus life cycle. First of all, virus gene expressions lead to formation of structural or coat proteins protecting the virus nucleic acid from environmental damage and several types of unstructural proteins including polymerases replicating the genomes, and movement proteins necessary for virus spread from one cell the next. Additionally, plant virus genomes can code proteases, helicases, methyltransferases, suppressors of posttranscriptional gene silencing (PTGS), and other proteins (Mandahar, 2006).

Sobemoviruses contain single-stranded, positive-sense RNA in size varying between 4.0 and 4.5 kb. Their monopartite genomes contain four open reading frames (ORF). To maximize the

coding capacity, sobemoviruses exploit the polyprotein sythesis and translational frameshifting gene expression strategies (Tamm and Truve, 2000b). The coding part of the *Sobemovirus* genomes start with a short ORF1 coding for 11.7-24.3 kDa proteins P1 with low similarity to each other (Tamm and Truve, 2000b; Callaway *et al.*, 2004). The functions of the sobemoviral P1 proteins are still under discussions. First, the *Sobemovirus* P1 was reported to possess the suppressor activity against PTGS, which is a major host defence response against plant viruses. Additionally, cysteine residues in putative Zn-finger structures were found to be essential for cell-to-cell movement of the P1 protein (Sire *et al.*, 2008). The second possible function of P1 protein is the role in systemic infection as demonstrated for the *Sobemovirus* member *Cocksfoot mottle virus* (CfMV) and other viruses. P1-deficient CfMV mutant was able to replicate in oat suspension culture but not to infect oat plants systemically (Meier *et al.*, 2006). Interesting observation was made for P1 protein from *Sobemovirus* RYMV with C-terminally fused green fluorescence protein (GFP). P1-GFP fusion protein was shown to possess an autoproteolytic activity in *E.coli* and plant cells, and the P1 N-terminus is involved in the cleavage process. The biological role of the autoproteolytic activity is not clear yet (Weinheimer *et al.*, 2010).

All sobemoviruses exploit the -1 ribosomal frameshifting mechanism to produce the characteristic polyproteins (Balke *et al.*, 2007; Meier and Truve, 2007). As a result of the frameshifting, overlapping ORF2a and ORF2b direct the expression of two different polyproteins with a molecular mass of around 100 and 60 kDa, and N-terminal parts of both proteins contain 3C-like serine protease (Pro) and genome-linked virus protein (VPg). Typical genome organization of sobemoviruses can be found in our recent article (Balke *et al.*, 2007). ORF2a encoded serine proteases are well characterized *Sobemovirus* non-structural proteins. The main function of *Somebovirus* proteases is the autoproteolytic polyprotein processing into several mature proteins through E/T, E/S or E/N cleavage sites (Makinen *et al.*, 2000; Satheshkumar *et al.*, 2004). However, unusual proteolytic cleavage site A/V was also observed in some experiments (Gayathri *et al.*, 2006). The best studied sobemoviral protease from *Sesbania mosaic virus* (SeMV) can be purified from recombinant *E.coli* cultures, if coding sequence of N-terminal transmembrane region were removed from the protease gene cDNA (Satheshkumar *et al.*, 2004). Accordingly to crystal structure analysis at resolution of 2.4 Å, N-terminally truncated recombinant SeMV serine protease exhibited the characteristic trypsin fold. Several surface-

exposed aromatic amino acids (AA) were identified in the protease structure, probably interacting with VPg (Gayathri *et al.*, 2006).

The next functional domain of sobemoviral polyproteins is VPg, which contains a conserved AA motif (WAD or WGD). It is covalently attached to the genomic RNA at the 5'-end and its function is supposed to serve as protein primer for replicase (Makinen *et al.*, 2000). Very recently obtained data suggest, that VPg is not necessary for viral RNA synthesis at least *in vitro* experiments. However, it is possible that the RNA *de novo* synthesis is still VPg-dependent requiring other viral/host factors in infected plants. Additionally, SeMV VPg serves as an activator for serine protease *in cis* and *trans,* if prepared in a form of fusion protein Pro-VPg from recombinant *E.coli* (Govind and Savithri, 2010).

The role of ORF2a C-terminal part beyond VPg was under discussion for long time. It was not clear whether it has a functional role, or it is only the byproduct of proteolytic cleavage. Recent studies, using purified ORF2a encoded C-terminal fragments from SeMV, the ATPase and nucleic acid binding activity was demonstrated for corresponding recombinant proteins (Nair and Savithri, 2010a). Such properties are characteristic, for example, for plant virus-encoded helicases. As demonstrated for recombinant TGBp1 proteins from potex- and hordeiviruses, helicase RNA unwinding activity is dependent on Mg ions and ATP (Kalinina *et al.*, 2002).

Sobemovirus replicases or RdRP are expressed as C-terminal parts of polyprotein after ORF2a frameshifting into ORF2b, resulting in polyprotein with a domain arrangement of Pro-VPg-RdRP. To characterize the SeMV RdRP biochemically, the enzyme was expressed in recombinant *E.coli* as a thioredoxin-fusion protein. Purified SeMV RdRP is able to synthesize RNA from specific genomic and subgenomic RNA with and without the protein primer VPg. Stem-loop structure at the 3'-end of the genome is important for the enzyme activity (Govind and Savithri, 2010).

Ryegrass mottle virus (RGMoV) is a plant virus, which belongs to the genus *Sobemovirus* and infects several species of *Gramineae* plants (Zhang *et al.*, 2001a). Accordingly to our three-dimensional (3D) crystal structure analysis data, RGMoV particles have icosahedric T=3 quasisymmetry and contain 180 coat protein subunits with the canonical jellyroll β–sandwich structure (Plevka *et al.*, 2007). Nucleotide sequence alignments complete genomes demonstrate phylogenetic grouping of Sobemoviruses in two groups, monocotyledon- and dicotyledon plants infecting viruses, respectively. Interestingly, RGMoV infecting monocotyledon plants but

grouping with dicotyledon-infecting ones is an exception from this rule (Sereme *et al.*, 2008). Recently we resequenced the complete genome of RGMoV and found that unlike that described previously (Zhang *et al.*, 2001a) RGMoV RdRP is translated via -1 ribosome frameshifting mechanism similarly to other members of the *Sobemovirus* genus. Besides, accordingly to the sequence analysis, ORF2a of RGMoV encodes typical sobemovirus 3C-like serine protease with strongly conserved triad and substrate-binding amino acids. Several putative protease cleavage sites can be identified in RGMoV polyproteins (the localization of the sites is indicated in the map - Balke et al., 2007).

As already mentioned, protease and coat protein 3D structures are solved, but very little is known about other *Sobemovirus* protein structure. CD spectral analysis suggest, that VPgs are expected to have "natively unfolded" structure (Satheshkumar *et al.*, 2005a). No data can be found about the structure of P1, RdRP and other sobemoviral polyprotein functional domains.

To characterize these proteins structurally and functionally, first of all it is necessary to purify them in quantitative amounts. In contrast to viral coat proteins, which are actively synthesized in many cases, the expression level of virus nonstructural proteins is very low in infected plants. For example, the output of plant virus RdRP, if purified from natural sources, can be as low as 0.002 mg/kg plant material (Bates *et al.*, 1995). Therefore, corresponding protein genes have to be cloned in heterologous expression vectors and purified from recombinant organisms.

Here we describe attempts to overexpress, purify and characterize the RGMoV unstructural proteins both experimentally and using computer-assisted programs.

MATERIALS AND METHODS

Construction of RGMoV nonstructural protein expression plasmids. The source of coding sequences for nonstructural protein genes (P1, Δ50Pro, VPg, P16 and RdRP) was the plasmid containing full length cDNA copy of RGMoV (Balke *et al.*, 2007). Nucleotide and AA sequences can be found in Genbank No.EF091714. For each protein construction, specific primers containing restriction sites for rapid cloning in corresponding expression vector were designed. Primers used for PCR-mediated cloning are summarized in Table 1. For production of recombinant proteins in *E.coli* following commercially available expression vectors were used: pET-28a(+), pET-Duet-1, pACYC-Duet-1 (all Novagen, USA), pCold-I (Takara, Japan). pET-derived vectors contain six-histidine coding sequence upstream of the recombinant protein insertion site allowing to produce His-tagged proteins and purify them using affinity

chromatography columns. pET-Duet-1 and pACYC-Duet-1 are designed for coexpression of two target genes in the same vector. pCold-I vector is designed to perform efficient protein expression at extremely low temperatures from cold-shock gene *csp*A. This vector also contains the His-tag sequence.

To amplify corresponding RGMoV genes or coding sequences of polyprotein domains, a high-fidelity DNA polymerase iProof (BioRad, USA) was used in PCR reactions accordingly to the protocol of the manufacturer. PCR reactions were carried out in Verity 96 Well Thermal Cycler (Applied Biosystems, USA). Obtained PCR products were first ligated in pTZ57 vector (Fermentas, Lithuania). After selection of insert containing clones, several clones were sequenced, using a BigDye cycle-sequencing kit and ABI Prism 3100 Genetic analyzer (Applied Biosystems, USA). PCR product clones with expected cDNA sequences were excised from pTZ57 and cloned in corresponding expression vector. P16, Δ50Pro, and RdRp were cloned in pET-Duet-1 vector using NdeI and BglII restriction sites. VPg was cloned using BamHI and HindIII restriction sites in to pACYC-Duet-1 vector. P1 was cloned using NdeI and BamHI restriction sites in to pCold-I vector. Additionally, P16 from pET-Duet-1 was recloned in to pET-28a(+) vector using restriction sites NdeI and XhoI. Similarly, genes of Δ50Pro and RdRp were recloned in pCold-I vector using the same restriction sites NdeI/BglII. All restriction enzymes were from Fermentas (Lithuania).

Nonstructural protein expression and purification. After isolating of corresponding expression vector DNA *E.coli* BL21(DE3) (New England Biolabs, USA) or WK-6 (Zell and Fritz, 1987) were transformed with corresponding plasmid to obtain RGMoV protein expression clones. In all cases, at least 3 expression clones were prepared from single colonies. Expression cultures in 20 ml of 2TY-medium with corresponding antibiotics (kanamicin, chloramphenicol or amicillin) were cultivated accordingly to the expression vector protocol, provided by Novagen or Takara, respectively. pET-containing cultures were grown at +30^{o}C, 200 rpm until optical density (OD) reached 0.8 measured at 600 nm. After induction with 0.5 mM IPTG the cultivation was continued at +20 ^{o}C for additional 16 h. For pColdI-containing cultures, cultivation was started at +30 ^{o}C and, after induction with IPTG at OD_{600}=0.3 continued overnight at +15 ^{o}C. For preparative cultures, the volume was scaled up to 200 ml.

Cell biomass were harvested by low speed centrifugation, suspended in ProFound lysis buffer (Thermo Scientific, USA) and sonicated on ice using cell disintegrator Soniprep 150 (MSE, UK). Soluble protein fraction was separated by centrifugation for 30 min at 13000 rpm and +4ºC. Proteins containing N-terminal His-tag were purified using N-sepharose affinity chromatography columns and B-PER 6xHis Fusion Protein Purification Kit (Thermo Scientific, USA) accordingly to the protocol for native protein purification provided by manufacturer. In cases of VPg and P16 proteins, the final steps of purification were ammonium sulfate precipitation and dialysis against 100 volumes of 10 mM Tris, pH 7.0.

Sample analysis in denaturing polyacrylamide gels (SDS-PAGE). All samples were analyzed in 12 % SDS-PAGE. The samples for gel analyses were diluted with a buffer containing 50 mM Tris pH 7.0, 1 % SDS, 50 % glycerin; 5 % mercaptoethanol; 0.004 % Bromephenol blue and heated for 10 min at +95ºC. SDS/PAGE gels were visualized by Coomassie blue R-250 staining.

Sample analysis by mass spectrometer MALDI TOF. MALDI-suitable matrix solution was made of 2, 5-dihydroxyacetophenone (7.6 mg – 50 µmol) (2, 5-DHAP; Bruker Daltonics, Germany) suspended in 375 µl ethanol (Rīgas Balzāms, Latvia). Add 125 µl of an aqueous diammonium hydrogen citrate solution (SIGMA, USA) from stock 27/1500 suspended in H_2O. Vortex suspension for at least 1 min. Heat the suspension for at 15 min, and additionally vortex for 1 min.

Following preparation for sample analysis was made. First,2 µl of protein-containing solution (1 mg/ml) was diluted with 2 µl of aqueous 2 % trifluoro acetic acid (TFA) solution. Second, add 2 µl of 2, 5-DHAP matrix solution. Mix the solution by pipeting up and down until the liquid gets cloudy. Third, apply 1 µl of the mixture onto 600 µm anchor of a Bruker AnchorChip 600 (Bruker Daltonics, Germany) target and let it crystallize. After drying, 600 µm anchor should be well-covered with matrix/analyt mixture. The analysis was performed using MALDI TOF mass spectrometer (Bruker Daltonics, Germany).

Protein crystallization. JCSG-plus HT-96 and Structure screen I + II HT-96 protein crystallization solution kits (Molecular Dimensions, UK) were used to find the optimum conditions for VPg and P16 crystal formation. Ni-Sepharose-purified VPg and P16 (6 mg/ml) were mixed 1:1 with a corresponding solution from the kit and spotted on the cover of 96 well

Microbatch plate. The wells of the plate were filled with 40 µl of the same buffer. The crystal formation was monitored under light microscope every 3-7 days in 8 week period.

Additionally, to identify crystallization conditions for VPg, the sample was sent to high-throughput screening facility located at the Hauptman-Woodward Medical Research Institute, USA. The screening facility ensures microbatch-under-oil crystallization experiments at 1536 different conditions (Luft et al., 2003).

Computer-assisted protein sequence analysis. To identify folded and unfolded regions of proteins, the software FoldIndex was used (Prilusky *et al.*, 2005), which is based on the average amino acid residue hydrophobicity and net charge of the sequence. The predictions were performed at window size 51 and step1. Alternatively, to identify disordered regions in RGMoV proteins, the software PONDR was used (http://www.pondr.com; (Romero *et al.*, 1997; Dunker *et al.*, 2002). Prediction program of transmembrane helices in proteins (http://www.cbs.dtu.dk/ - TMHMM version 2.0) was used for RGMoV 3C-serine protease and RdRP domain analysis. All programs are based on AA sequence analysis.

RESULTS

Cloning, expression and purification of RGMoV unstructural proteins. To express separately domains of RGMoV unstructural proteins, first the sequence analysis was needed. As demonstrated for other sobemoviruses, the formation of functionally active domains is dependent on polyprotein cleavage by 3C-serine protease at sites amino acid pairs E/T, E/S or E/N (Makinen *et al.*, 2000; Satheshkumar *et al.*, 2004). As the substrate specificity for RGMoV protease is not known, it was assumed that the enzyme cleaves the polyprotein at the same sites. The relative localization of putative protease cleavage sites are shown in our previos paper (Balke *et al.*, 2007). Based on the assumption, ORF2a (551 AA) protease domain was set between amino acids 1 and 317, VPg domain from AA 318 to 382, and P16 domain from AA 383 to 551, respectively. The order of domains in ORF2a polyprotein and FoldIndex analysis is shown in Fig.1A. For sobemoviral proteases, it is known that the 5'-terminal part of the gene codes a membrane anchor domain (Satheshkumar *et al.*, 2004). As shown in Fig.1B, prediction using TMHMM program revealed the putative transmembrane anchor sequence (An) between AA 1 and 50 in the Pro sequence, which can significantly reduce the solubility of the

recombinant protein. Therefore, to stimulate the formation of soluble protein in *E.coli* cells, the coding sequence of first 50 AA was removed from the Pro expression clone.

All ORF2a fragments (Δ50Pro, VPg, P16) and coding sequences for P1 and RdRP were amplified by PCR, and cloned in expression vectors as described in Materials and Methods. First, we tested the formation of serine protease in different expression systems. As shown in Fig.2, the protease level was undetectable after expression of Δ50Pro gene from pET derived vector in BL21(DE3) cells (Fig.2A). Better expression level was observed, if the same cells contained protease gene in pCold I vector (Fig.2B). The best production level of Δ50Pro was achieved in case of the expression host *E.coli* WK6 and pColdI-derived vector. As seen from SDS/PAGE analysis, most part from Δ50Pro was found in soluble protein fraction (Fig.2C, lane 3). To purify the protein, Δ50Pro-containing fraction was loaded on Ni-Sepharose column and eluted with imidazole buffer. However, protein was not able to bind to the column material and was found in unbound protein fraction (data not shown).

Similar expression strategy was applied in case of RGMoV P1 protein and RdRP. P1 protein expression level was undetectable in all vector and host strain combinations. To construct the RdRP expression clone, the part of ORF2a-ORF2b region was chosen, which codes AA 448-947 after first putative E/T cleavage site in ORF2b protein (see Balke *et al.*, 2007 and Genbank file No.EF091714). Expression level analysis of RdRP demonstrated enhanced protein synthesis in case of pCold-derived vector and *E.coli* WK6 cells, if compared with pET-vectors and BL21(DE3) host (Fig.3). However, RdRP was found in insoluble protein fraction and the purification on Ni-Sepharose column at native conditions was impossible. Accordingly to TMHMM prediction data, expressed RdRP domain does not contain transmembrane helices in the structure.

For the expression of VPg domain, corresponding cDNA was cloned in pACYC-Duet-1 vector and *E.coli* BL21(DE3) was chosen as a host strain. As the expression level was sufficient for the preparative production of VPg (Fig.4; lane 2), other vectors and host strains were not further tested. SDS/PAGE analysis revealed that VPg was soluble and could be purified using Ni-Sepharose affinity chromatography columns. As shown in Fig.4 (lanes 7, 8), VPg bound efficiently to the column material and could be eluted from the column with the imidazole-containing elution buffer.

To obtain P16 in preparative amounts for subsequent characterization experiments, the PCR product containing P16 cDNA was cloned in pET-28a+ after His-tag coding sequence in sites NdeI/BglII. Similarly to VPg, the production level of soluble P16 in BL21(DE3) cells was acceptable and other expression systems were not tested. Similarly to VPg P16 protein could be efficiently purified using Ni-Sepharose columns (Fig.5). As surprisingly found in SDS/PAGE analysis, denaturated P16 protein (calculated MW=18928 Da) appeared on gels as a protein with MW approximately 34 kDa (Fig.5), and the abnormal behavior of P16 in SDS/PAGE was not changed even after 30 min boiling in 0.1M DTT /1% SDS buffer. To confirm the P16 origin of the purified protein, the eluted protein (Fig.5, lane 8) after precipitating with ammonium sulfate and dialysis was tested using mass spectrometry MALDI TOF. Mass spectrometry analysis revealed two major peaks (Fig.5, B). One peak of 18603 Da approximately corresponds to P16 monomer with 6 His-tag and without methionine (from AA sequence calculated value MW_{P16}=18908 Da), the second major peak of 37473 Da, which can be a signal of P16 dimmer. These data demonstrate comparably high stability of putative P16 dimmer, even at MALDI TOF experimental conditions.

Crystallization experiments. In order to obtain crystals for protein three-dimensional structure studies, purified VPg and P16 proteins were prepared for crystal formation using the hanging drop method accordingly to Materials and Methods. Two solution kits were used to test the crystal formation depending on such factors as different combinations of precipitant (PEG4000, ammonium sulfate, 2-propanol and others) at different pH in presence of different cations and anions. Altogether 192 different solutions were tested, but none of them supported the crystal formation even after 8 week-incubation. Then, VPg protein sample was sent to Hauptman-Woodward Medical Research Institute to find out the conditions for crystallization from 1536 different solutions. Also in this case the conditions for VPg crystal formation were not found. These data indicate that recombinant P16 and VPg proteins in the form of separate domains can not be crystallized, making impossible the subsequent crystallographic 3D-structure analysis.

Bioinformatic protein sequence analysis. From numerous protein crystallization experiments it is well known, that unstructured regions in protein structure can inhibit crystal formation (Pantazatos *et al.*, 2004). The reason of unsuccessful crystallization experiments with VPg and P16 can be possible disordered regions in the structure of these proteins. Already initial

FoldIndex-supported analysis of complete RGMoV ORF2a revealed several disordered regions in ORF2a polyprotein and significant part amino acids are in the disordered state (193 from 551). Accordingly to FoldIndex, complete VPg domain and at least the part of P16 domain represents clearly unfolded structure of in contrast to well structured protease domain (Fig.1A). To obtain additional data about the VPg and P16, another bioinformatics tool (PONDR) was chosen for specific analysis of both recombinant proteins. Using the program PONDR, calculations of the VPg net charge and hydropathy resulted in clear placing of VPg among the disordered proteins, however, the whole VPg domain accordingly to PONDR score was classified as rather ordered (Fig.4; B,C).

PONDR analysis results for P16 revealed the position of the protein close to the border between ordered and disordered proteins (Fig.5, C). Accordingly to results of PONDR, the C-terminal part of P16 protein appears to be clearly disordered with the PONDR score close to 1.0 (Fig.5, D).

DISCUSSION

The production of stable, soluble proteins in preparative amounts is one of the most important steps in the protein functional studies and structure determination. Here we demonstrate attempts to clone, overexpress and purify plant virus RGMoV nonstructural proteins with an aim to characterize them functionally and structurally. Several plasmid vector/*E.coli* host combinations were tested in order to maximize the output of corresponding proteins after cultivation. It was shown in cases of protease and RdRP, that the expression level of both proteins could be considerably enhanced, if pColdI-derived vector was used in combination with *E.coli* WK6 cells, whereas pET-derived vectors and BL21(DE3) turned out as inefficient. However, the system was not suitable for P1 protein expression. Apparently, different systems as expression of P1 with different N-terminal fusion partners have to be tested. In case of RdRP most part of the protein was found to be insoluble in recombinant *E.coli,* despite of 6His-tag containing leader sequence of pColdI-origin at the N-termini of RdRP and low cultivation temperature. An efficient alternative has been demonstrated in case of SeMV RdRP: soluble polymerase can be obtained in a form of thioredoxin fusion protein (Govind and Savithri, 2010).

As shown in our experiments, N-terminally truncated RGMoV 3C-like serine protease containing 6His-tag could be successfully expressed in soluble form. However, the protein was

not able to bind to Ni-Sepahrose column, indicating the absence of accessible His-tag in the protein structure. One possible reason for missing His-tag in Δ50Pro can be found by careful FoldIndex analysis: protein contains a short disordered region between AA 76 and 83 (Fig.1A) including Lys and Arg residues. This region can be cleaved by *E.coli* trypsin-like enzymes (Pantazatos *et al.*, 2004), resulting in N-terminally truncated protein without His-tag. The SDS/PAGE analysis indirectly confirmed this assumption, because the theoretical calculation of Δ50Pro MW resulted in 28.7 kDa, but the observed value was only approximately 26 kDa (Fig.2, C). On the other hand, unusual self-cleavage by Δ50Pro can not be excluded, as discussed in case of SeMV protease.

The next functional member of RGMoV structural proteins, the VPg was successfully expressed and purified using His-tag affinity chromatography. All attempts to crystallize the VPg were unsuccessful, irrespective of more than 1500 crystallization solutions tested. The reason for the inability of VPg to form crystals can be bound with unstructured regions in the protein structure. The bioinformatics analysis, using the program PONDR, demonstrated the disordered properties of VPg, if compared with other proteins. It was found, that significant part of the VPg amino acid sequence is close to the PONDR score of 0.5. Such protein segments are often predicted to be so called molecular recognition elements or features (MoRFs). MoRFs are short amino acid stretches able to undergo disorder-to-order transitions upon specific binding with functional partners (Vacic *et al.*, 2007). Plant virus VPg, including members of *Sobemovirus* group, are characterized as extended “natively unfolded proteins” lacking a distinct 3D-structure and existing at different conformations at physiological conditions (Hebrard *et al.*, 2009). Interestingly, VPg was demonstrated to modulate sobemoviral protease activity (Gayathri *et al.*, 2006), if prepared in a form of fusion protein Pro-VPg. One could expect that in the form of the fusion protein VPg could be more structured and supported the crystal formation, needed for VPg 3D-structural analysis. Interestingly, also immunization experiments confirmed disordered nature of VPg. The immune responses in mice immunized with purified VPg were highly low and obtained antibodies were not sensitive enough to detect the VPg in Western blots at the dilutions higher than 1:50 (data not shown).

Similarly to VPg, also P16 protein was not able to form crystals at different conditions. The bioinformatic analysis revealed clearly unstructured C-terminal part of the protein (Fig.5, D),

which could negatively influence the crystal formation. Additionally, P16 demonstrated several interesting properties. First, the mobility of P16 in SDS/PAGE was unusual. The reason for this property can be bound with amino acid content of P16. C-terminal part of the protein contains 14 basic AA (Arg and Lys), resulting in isoelectric point of 9.26 for P16. Positively charged proteins typically behave anomalously in SDS/PAGE as shown also in case of SeMV P8 protein (Nair and Savithri, 2010b). The second, our mass spectrometry analysis data suggest, that P16 protein can be a homodimer at physiological conditions. The third, P16 can bind nucleic acids as observed in first gel-retardation experiments (data not shown). Purified P16, if incubated with specific RGMoV RNA, as well as with unspecific RNA, caused a typical shift in nucleic acid mobility on agarose gels. Here additional studies are necessary to demonstrate the substrate specificity for potential binding partners of P16. The finding of specific binding partner is important not only from functional characterization of P16. P16 co-crystallization experiments with corresponding nucleic acid can stimulate the crystallization process of P16 protein.

ACKNOWLEDGEMENTS

Authors wish to thank Prof. P. Pumpēns and Dr. K. Tārs for numerous helpful discussions and support during the preparation of this article. This work was supported by ESF project No. 2009/0138/1DP/1.1.2.1.2/09/IPIA/VIAA/004, ESF project No. 2009/0204/1DP/1.1.1.2.0/09/APIA/VIAA/150 and Grant No. 04.1149 from Latvian Council of Science.

REFERENCES

Balke, I., Resevica, G. and Zeltins, A. (2007) The ryegrass mottle virus genome codes for a sobemovirus 3C-like serine protease and RNA-dependent RNA polymerase translated via -1 ribosomal frameshifting. *Virus Genes* **35**(2): 395-8.

Bates, H. J., Farjah, M., Osman, T. A. M. and Buck, K. W. (1995) Isolation and characterization of an RNA-dependent RNA polymerase from Nicotiana clevelandii plants infected with red clover necrotic mosaic dianthovirus. *J Gen Virol* **76**: 1483-1491.

Callaway, A. S., George, C. G. and Lommel, S. A. (2004) A Sobemovirus coat protein gene complements long-distance movement of a coat protein-null Dianthovirus. *Virology* **330**: 186-195.

Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M. and Obradovic, Z. (2002) Intrinsic disorder and protein function. *Biochemistry* **41**(21): 6573-82.

Gayathri, P., Satheshkumar, P. S., Prasad, K., Nair, S., Savithri, H. S. and Murthy, M. R. (2006) Crystal structure of the serine protease domain of Sesbania mosaic virus polyprotein and mutational analysis of residues forming the S1-binding pocket. *Virology* **346**(2): 440-51.

Govind, K. and Savithri, H. S. (2010) Primer-independent initiation of RNA synthesis by SeMV recombinant RNA-dependent RNA polymerase. *Virology* **401**(2): 280-92.

Hebrard, E., Bessin, Y., Michon, T., Longhi, S., Uversky, V. N., Delalande, F., Van Dorsselaer, A., Romero, P., Walter, J., Declerck, N. and Fargette, D. (2009) Intrinsic disorder in Viral Proteins Genome-Linked: experimental and predictive analyses. *Virol J* **6**: 23.

Kalinina, N. O., Rakitina, D. V., Solovyev, A. G., Schiemann, J. and Morozov, S. Y. (2002) RNA helicase activity of the plant virus movement proteins encoded by the first gene of the triple gene block. *Virology* **296**(2): 321-9.

Makinen, K., Makelainen, K., Arshava, N., Tamm, T., Merits, A., Truve, E., Zavriev, S. and Saarma, M. (2000) Characterization of VPg and the polyprotein processing of cocksfoot mottle virus (genus Sobemovirus). *J Gen Virol* **81**(Pt 11): 2783-9.

Mandahar, C. (2006) Positive-Sense Viral RNA. In: *Multiplication of RNA Plant Viruses*. Mandahar, C. (ed.). Springer Verl, Dordrecht, pp. 29-70.

Meier, M., Paves, H., Olspert, A., Tamm, T. and Truve, E. (2006) P1 protein of Cocksfoot mottle virus is indispensable for the systemic spread of the virus. *Virus Genes* **32**(3): 321-6.

Meier, M. and Truve, E. (2007) Sobemoviruses possess a common CfMV-like genomic organization. *Arch Virol* **152**(3): 635-40.

Nair, S. and Savithri, H. S. (2010a) Natively unfolded nucleic acid binding P8 domain of SeMV polyprotein 2a affects the novel ATPase activity of the preceding P10 domain. *FEBS Lett* **584**(3): 571-6.

Nair, S. and Savithri, H. S. (2010b) Processing of SeMV polyproteins revisited. *Virology* **396**(1): 106-17.

Pantazatos, D., Kim, J. S., Klock, H. E., Stevens, R. C., Wilson, I. A., Lesley, S. A. and Woods, V. L., Jr. (2004) Rapid refinement of crystallographic protein construct definition employing enhanced hydrogen/deuterium exchange MS. *Proc Natl Acad Sci U S A* **101**(3): 751-6.

Plevka, P., Tars, K., Zeltins, A., Balke, I., Truve, E. and Liljas, L. (2007) The three-dimensional structure of ryegrass mottle virus at 2.9 A resolution. *Virology* **369**(2): 364-74.

Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I. and Sussman, J. L. (2005) FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**(16): 3435-8.

Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E. and Dunker, A. K. (1997) Identifying Disordered Regions in Proteins from Amino Acid Sequences. Proc. I.E.E.E. International Conference on Neural Networks, p. 90-95.

Satheshkumar, P. S., Gayathri, P., Prasad, K. and Savithri, H. S. (2005) "Natively unfolded" VPg is essential for Sesbania mosaic virus serine protease activity. *J Biol Chem* **280**(34): 30291-300.

Satheshkumar, P. S., Lokesh, G. L. and Savithri, H. S. (2004) Polyprotein processing: cis and trans proteolytic activities of Sesbania mosaic virus serine protease. *Virology* **318**(1): 429-38.

Sereme, D., Lacombe, S., Konate, M., Pinel-Galzi, A., Traore, V. S., Hebrard, E., Traore, O., Brugidou, C., Fargette, D. and Konate, G. (2008) Biological and molecular characterization of a putative new sobemovirus infecting Imperata cylindrica and maize in Africa. *Arch Virol* **153**(10): 1813-20.

Sire, C., Bangratz-Reyser, M., Fargette, D. and Brugidou, C. (2008) Genetic diversity and silencing suppression effects of Rice yellow mottle virus and the P1 protein. *Virol J* **5**: 55.

Tamm, T. and Truve, E. (2000) Sobemoviruses. *J Virol* **74**(14): 6231-41.

Vacic, V., Oldfield, C. J., Mohan, A., Radivojac, P., Cortese, M. S., Uversky, V. N. and Dunker, A. K. (2007) Characterization of molecular recognition features, MoRFs, and their binding partners. *J Proteome Res* **6**(6): 2351-66.

Weinheimer, I., Boonrod, K., Moser, M., Zwiebel, M., Fullgrabe, M., Krczal, G. and Wassenegger, M. (2010) Analysis of an autoproteolytic activity of rice yellow mottle virus silencing suppressor P1. *Biol Chem* **391**(2-3): 271-81.

Zell, R. and Fritz, H. J. (1987) DNA mismatch-repair in Escherichia coli counteracting the hydrolytic deamination of 5-methyl-cytosine residues. *EMBO J* **6**: 1809-1815.

Zhang, F. Y., Toriyama, S. and Takahashi, M. (2001) Complete Nucleotide Sequence of Ryegrass Mottle Virus: A New Species of the Genus Sobemovirus. *J Gen Plant Pathol* **67**: 63-68.

http://www.cbs.dtu.dk/services/TMHMM/

Daudzziedu airenes nestukturālo proteīnu ekspresija un raksturošana

Daudzziedu airenes plankumainības vīrusa (RGMoV) vienpavediena RNS genoms ir organizēts četrās atklātajās nolasījuma fāzēs (ORF), kuras kodē vairākus proteīnus: ORF1 kodē P1 proteīnu, ORF2a satur ar membrānu saistīti 3C serīna proteāzes, ar genomu saistīto vīrusa proteīna VPg un P16 proteīna gēnu, ORF2b kodē replikāzi RdRp, un vienīgo strukturālo proteīnu, apvalka proteīnu, kas tiek sintezēts no ORF3. Lai iegūtu nestrukturālos proteīnus preparatīvos daudzumos un raksturotu tos, atbilstošo RGMoV gēnu cDNS tika klonētas pET un pColdI ekspresijas vektoros un ekspresētas vairāku *E.coli* celmu šūnās. Proteāzes un RdRP gadījumā labākā ekspresijas izrādījās sistēma, kurā tika izmantots pColdI vektors un *E.coli* WK6 celms. VPg un P16 proteīni tika iegūti no pET vai pACYC vektoriem un *E.coli* BL21(DE3) saimniekšūnām un attīrīti, izmantojot Ni-Sepahrose afinitātes hromatogrāfiju. Mēģinājumi kristalizēt VPg un P16 bija neveiksmīgi, iespējams, nestrukturētu aminoskābju sekvenču klātbūtnes dēļ abu proteīnu struktūrā. Uz bioinformātikas metodēm balstīta analīze parādīja, ka viss VPg domēns ir nestrukturēts un P16 proteīnam nestrukturēta ir C-terminālā daļa, kas, iespējams, traucēja kristālu veidošanos.

Table1. Primers used in RGMoV nonstructural protein cloning.

Primer name	Primer sequence	Introduced restriction site
RG2-VPg-BamHI-F	5’ GGATCC A AAC GGA GAG CAG GGA GCG CGC GAG A 3’	BamHI
RG2-VPg-HindIII-R	5’ AAGCTT TCA TTC TTC ACG GGC GTC CCA ATC A 3’	HindIII
RG-SerP-BglII-R	5’ AGATCT TCA TTC ACT GGA TTC ACA GTT TGC ATG GAA AAT GGA GA 3’	BglII
RG-Δ50Pro-F	5’ CCATATG CGC GCTT GGC TAG CAA CCT CTC TGGA 3’	NdeI
RG-P16-F	5’CCATATG TCC ACA GGA AAT GAT ATT CCT TTA AAC TGC CAG CA 3’	NdeI
RG-P16-BglII-R	5’ AGATCT TCA AGC TGA GGG GGA CCC CTG GAC T 3’	BglII
RG-P1-NdeI-F	5’ CATATG CCT TCA GTG GTT ATC GAG GTT TGC TCAT 3’	NdeI
RG-P1-BglII-R	5’ AGATCT TCA ATG ATG TCT AGT CCA AGA CTG CCC T 3’	BglII
RG-Pol-BglII-R	5’ AGATCT TCA TTC TTC CGG GAT TTC TCC TTC A 3’	BglII
RG-Pol-E/T1983-NdeI-F	5’CATATG ACT GCT AGA ACG AGT ACT AAC GAG CTC T 3’	NdeI

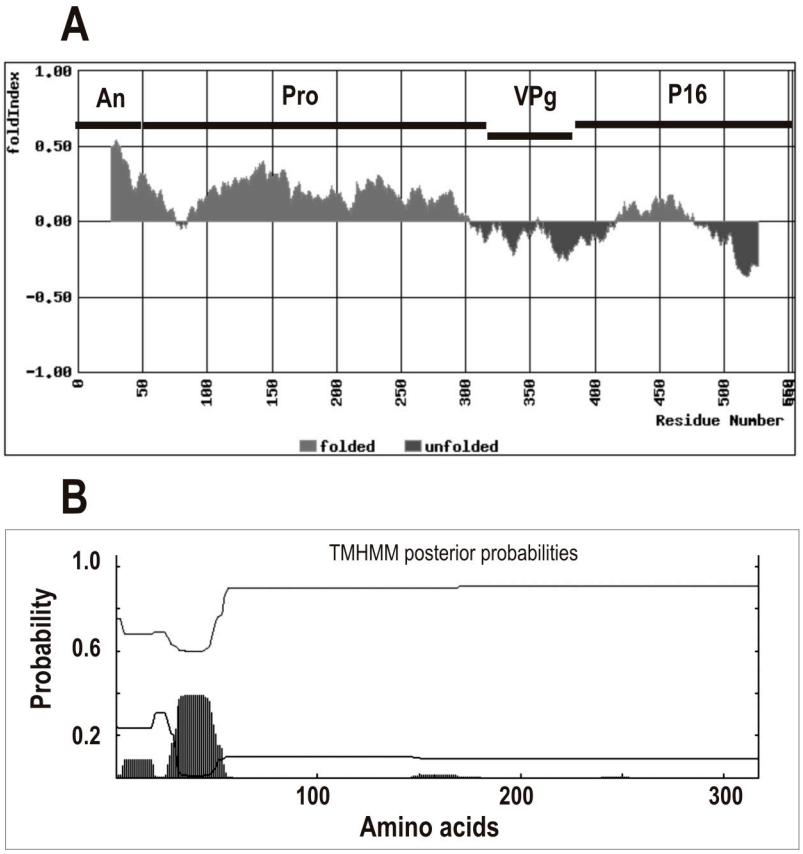

Fig.1. Prediction of RGMoV ORF2a domain structure. A - analysis of ORF2a using the programm FoldIndex (window size 51, step 1). Folded amino acid regions have positive FoldIndex values, unfolded - negative, respectively. Localization of ORF2a putative functional domains are indicated in respect to amino acid sequence: An - membrane anchor sequence of protease, Pro - 3C-like serine protease, VPg - genome-linked virus protein, P16 - ORF2a C-terminal domain. B - prediction of transmembrane helices in RGMoV protease using TMHMM programm. Putative transmembrane region is indicated as shaded curve.

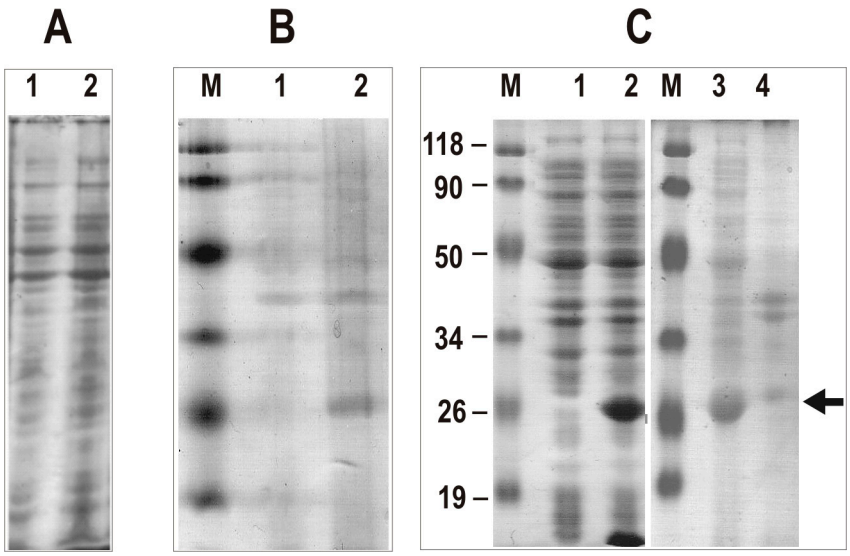

Fig. 2. Δ50Pro expression analysis. A – pET-Du-Δ50Pro expression in *E.coli* BL21(DE3); B – pCold-Δ50Pro expression in *E.coli* BL21(DE3); C – pCold-Δ50Pro expression in *E.coli* WK-6. M – protein weight marker (Fermentas, Lithuania); 1 – non induced total cell lysates; 2 – induced total cell lysates; 3 – soluble protein fraction; 4 – cell debris and insoluble proteins.

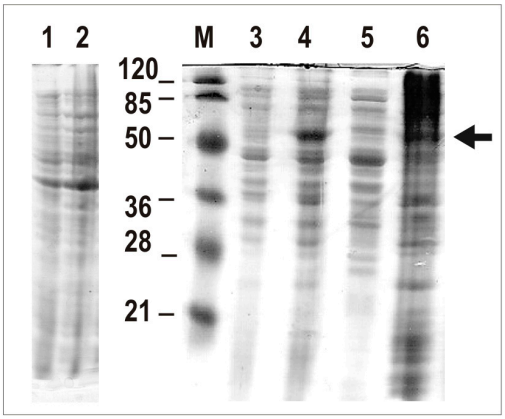

Fig. 3. Expression analysis of pET-RdRp.1, 2 - pET-Du-RdRpE/T1983 expression in BL21(DE3); 3-6 - pCold-RdRpE/T1983 expression in WK-6; 1,3 - noninduced total cell lysates; 2,4 - induced total cell lysates; 5 - soluble fraction; 6 - insoluble proteins and cell debris; M - protein marker (Fermentas, Lithuania).

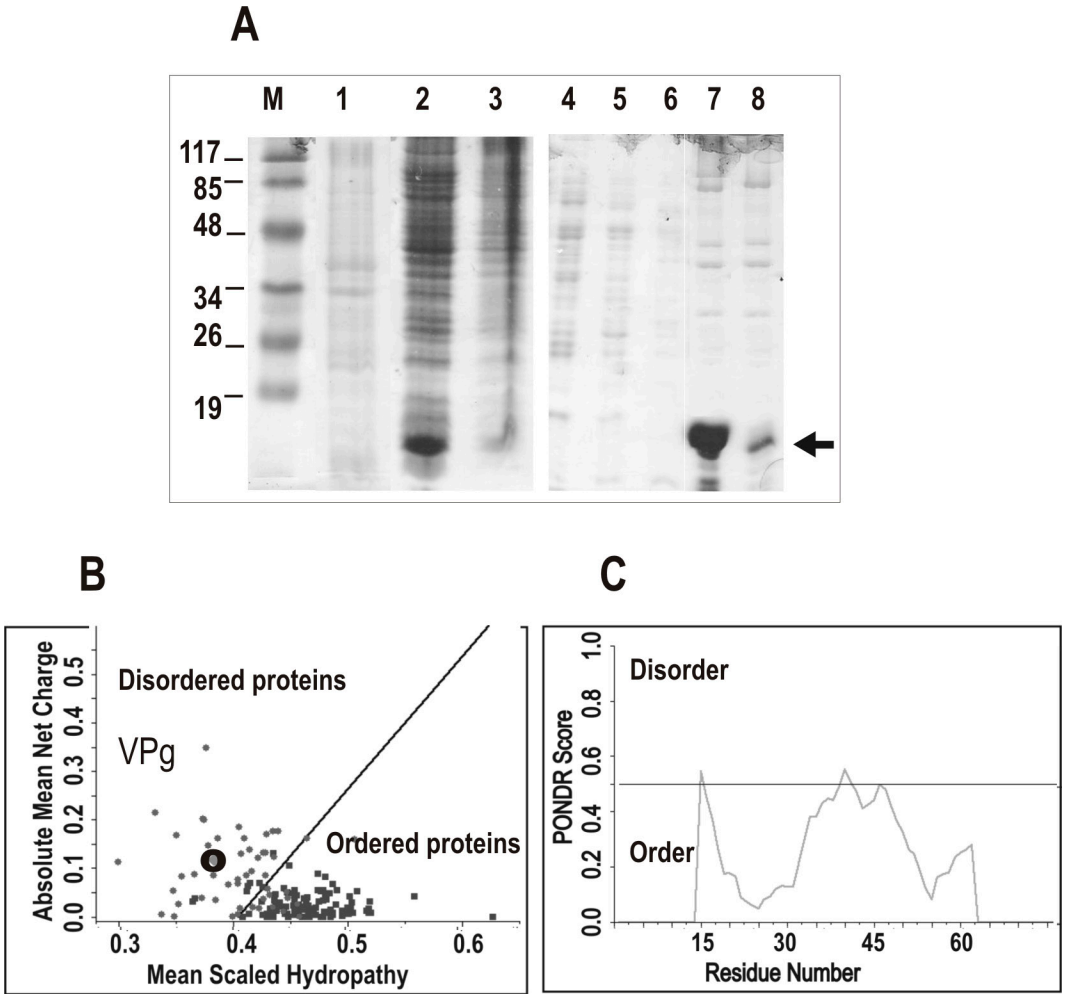

Fig.4. VPg expression, purification and sequence analysis. A - SDS/PAGE analysis: M-protein weight markers (Fermentas); 1- noninduced total cell lysates; 2 - soluble protein fraction; 3 - insoluble proteins and cell debris; 4 - proteins not bound to Ni-Sepharose column; 5, 6 - Ni-Sepharose column washing fraction; 7, 8 - proteins eluted from Ni-Sepharose column; B - prediction of folding nature of VPg in comparison with other ordered and disordered proteins; the position of VPg is labeled with "O"; C - ordered and disordered regions of VPg depending on amino acid sequence.

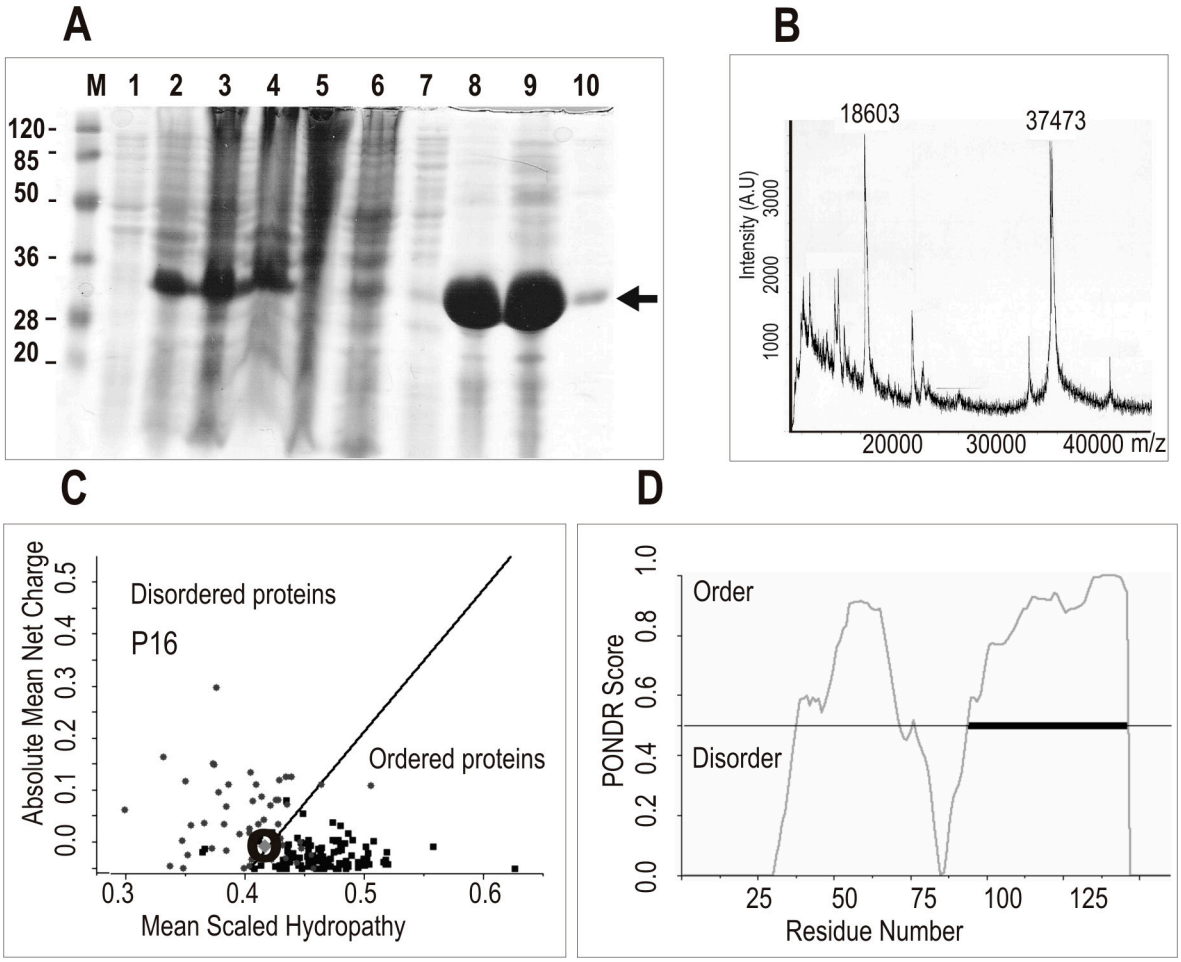

Fig. 5. Expression, purification and analysis of P16. A - SDS/PAGE analysis of pET-P16 expression and purification on Ni-Sepharose column; M - protein weight marker (Fermentas, Lithuania); 1 - noninduced cell total lysate; 2- IPTG induced cell total lysate; 3 - soluble protein fraction; 4 - insoluble proteins and cell debris ; 5 - nonbound proteins on Ni-Sepharose column; 6, 7 - proteins eluted with washing buffer; 8, 9, 10 - eluted P16 fractions. B - P16 analysis by mass spectrometer.C - prediction of folding nature of P16 in comparison with other ordered and disordered proteins; the position of P16 is labeled with "O"; D - ordered and disordered regions of P16 depending on amino acid sequence.

III THE THREE-DIMENSIONAL STRUCTURE OF RYEGRASS MOTTLE VIRUS AT 2.9 Å RESOLUTION

Contribution – virus production, purification and sequence verification.

The original publication should be cited as follows:
Plevka P., Tars K., Zeltins A., Balke I., Truve E., Liljas L. 2007. The three-dimensional structure of *Ryegrass mottle virus* at 2.9 Å resolution. *Virology* 369:364–374

Available online at www.sciencedirect.com
ScienceDirect
Virology 369 (2007) 364–374
VIROLOGY
www.elsevier.com/locate/yviro

The three-dimensional structure of ryegrass mottle virus at 2.9 Å resolution

Pavel Plevka [a], Kaspars Tars [a,*], Andris Zeltins [b], Ina Balke [b], Erkki Truve [c], Lars Liljas [a]

[a] *Department of Cell and Molecular Biology, Uppsala University, Box 596, SE-751 24 Uppsala, Sweden*
[b] *Latvian Biomedical Research and Study Centre, Ratsupites 1, LV 1067, Riga, Latvia*
[c] *Department of Gene Technology, Tallinn University of Technology, Akadeemia tee 15, 19086 Tallinn, Estonia*

Received 23 May 2007; accepted 17 July 2007
Available online 19 September 2007

Abstract

The crystal structure of the sobemovirus *Ryegrass mottle virus* (RGMoV) has been determined at 2.9 Å resolution. The coat protein has a canonical jellyroll β-sandwich fold. In comparison to other sobemoviruses the RGMoV coat protein is missing several residues in two of the loop regions. The first loop contributes to contacts between subunits around the quasi-threefold symmetry axis. The altered contact interface results in tilting of the subunits towards the quasi-threefold axis. The assembly of the T=3 capsid of sobemoviruses is controlled by the N-termini of C subunits forming a so-called β-annulus. The other loop that is smaller in the RGMoV structure contains a helix that participates in stabilization of the β-annulus in other sobemoviruses. The loss of interaction between the RGMoV loop and the β-annulus has been compensated for by additional interactions between the N-terminal arms. As a consequence of these differences, the diameter of the RGMoV particle is 8 Å smaller than that of the other sobemoviruses.

The interactions of coat proteins in sobemovirus capsids involve calcium ions. Depletion of calcium ions results in particle swelling, which is considered a first step in disassembly. We could not identify any density for metal ions in the proximity of the conserved residues normally involved in calcium binding, but the RGMoV structure does not show any signs of swelling. A likely reason is the low pH (3.0) of the crystallization buffer in which the groups interacting with the calcium ions are not charged.

Keywords: Ryegrass mottle virus; Sobemovirus; Virus structure; Virus assembly

Introduction

Ryegrass mottle virus (RGMoV), a member of the *Sobemovirus* genus, infects Italian ryegrass (*Lolium multiflorum*) and cocksfoot (*Dactylis glomerata*). It is distributed in Japan. RGMoV is a nonenveloped spherical virus with a diameter of about 290 Å.

The genome of RGMoV is a single-stranded, positive-sense, nonpolyadenylated RNA with a covalently linked 5′ VPg protein, which may serve as a replication primer (Zhang et al., 2001). The replication is carried out by a virus-encoded RNA-dependent RNA polymerase (RdRP). During replication full size copies of genomic RNA are created together with subgenomic RNAs corresponding to the 3′ part of the genome. The RGMoV genome contains four open reading frames (ORFs). ORF1 encodes protein P1, which is possibly involved in virus movement within the infected plant. RGMoV polyproteins are encoded by two overlapping ORF2a and ORF2b. The ORF2a codes for a sobemovirus 3C-like protease and VPg, whereas ORF2b gene product contains the RNA-dependent RNA polymerase (Balke et al., 2007; Meier and Truve, 2007). The coat protein is expressed from the 3′ proximal ORF3.

The crystal structure of several members of the *Sobemovirus* genus has been determined: *Southern cowpea mosaic virus* (SCPMV, previously known as *Southern bean mosaic virus*, cowpea strain) (Abad-Zapatero et al., 1980), *Sesbania mosaic virus* (SeMV) (Bhuvaneshwari et al., 1995), *Rice yellow mottle virus* (RYMV) (Qu et al., 2000) and *Cocksfoot mottle virus* (CfMV) (Tars et al., 2003). The structure is also known for *Tobacco necrosis virus* (TNV), which belongs to the genus *Necrovirus*, assigned to the *Tombusviridae* family, but has a coat protein of remarkable similarity to those of sobemoviruses

* Corresponding author. Fax: +46 18 536971.
E-mail address: kaspars@xray.bmc.uu.se (K. Tars).

0042-6822/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2007.07.028

(Oda et al., 2000). The sobemoviruses are not assigned to the *Tombusviridae* family because the gene order and the properties of the nonstructural proteins are distinctly different (Tamm and Truve, 2000).

Sobemoviruses have a capsid arrangement with icosahedral T=3 quasi-symmetry in which the 60 icosahedral asymmetric units consist of three chemically identical copies of the coat protein, which are designated A, B and C. Five A subunits form pentamers around the icosahedral fivefold axis and three of each of the B and C subunits assemble into hexamers around the threefold axis. The pentamers and hexamers differ significantly in shape—hexamers being generally planar and pentamers substantially bent. The hexamer/pentamer combination on the virus surface gives the particle its characteristic shape of a rhombic triacontahedron. The N-terminal part (called the N-terminal arm) of the C subunit of the coat protein plays a crucial role in determining the capsid size. The N-terminal arm of the C subunit is partly ordered and inserted between the interacting sides of the subunits, making the contacts around the icosahedral twofold axis flat. The N-terminal arms are completely disordered in the A and B subunits. The contacts of the A and B subunits around the quasi-twofold axis lack the inserted arms and are bent. The flat or bent type of contact between the subunits is thus caused by the presence or absence of an ordered N-terminal arm. The removal of the N-terminal arm by proteolytic cleavage (Rossmann et al., 1983) or by genetic engineering methods (Lokesh et al., 2002) results in formation of T=1 particles, where all the contacts between subunits are bent. In this manner, the N-terminus acts as a molecular switch, controlling the angles of the subunit contacts, which in turn control the curvature of the particle and allow successful assembly of the T=3 capsid.

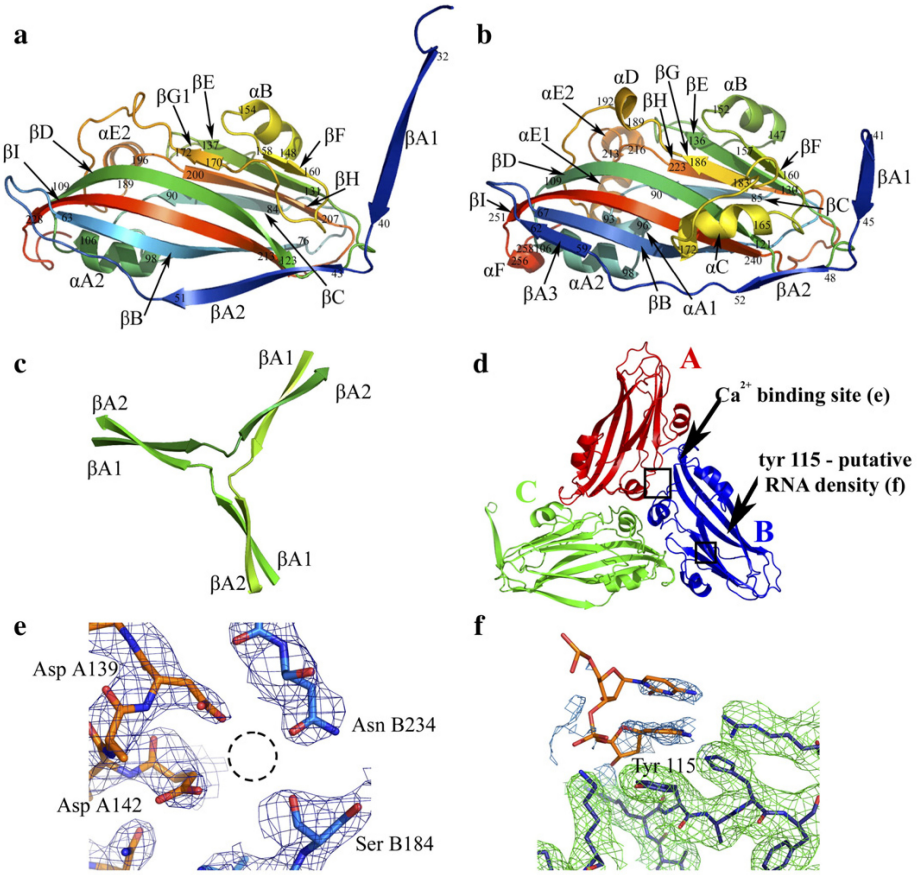

Fig. 1. Overview of the RGMoV capsid structure. Rainbow colored (N-terminus blue→C-terminus red) ribbon diagram of the RGMoV (a) and SeMV (b) C subunit as viewed from the inside of the particle. Nomenclature for α-helices and β strands and numbers of the residues at the beginning and end of the secondary structure elements are shown. (c) β-annulus formed by the three N-termini of the RGMoV C subunits (shown in different shades of green). (d) Cartoon model of the asymmetric unit of RGMoV with marked position of putative Ca^{2+} binding site between A and B subunits and a putative RNA density in the B subunit. (e) Ca^{2+} binding site between the RGMoV A (orange carbon atoms) and B (blue carbon atoms) subunits. The residues involved in calcium ion binding in the other sobemoviruses are labelled according to the RGMoV numbering. The C-terminal Gln 235, the carboxyl group of which is involved in the calcium binding, lies above the view plane and is not shown in the picture. The dashed circle indicates the putative calcium ion position as observed in other sobemoviruses. The electron density map was contoured at 3σ. No density for an ion was observed. (f) The electron density of putative RNA bases stacked to the side chain of tyrosine 115 of the RGMoV B subunit. Both the maps around the cytosine dinucleotide (blue) and the protein subunit (green) were contoured at 2σ.

The β-strands A1 and A2 in the N-terminal arm of the RGMoV interact in an antiparallel manner with the N-terminal arms of the threefold related C subunits. This arrangement of the N-terminal arms of the C subunits was initially observed in *Tomato bushy stunt virus* (TBSV) (Harrison et al., 1978) and was named β-annulus. The nomenclature was adopted for sobemoviruses as well, in spite of the β-annulus having more the shape of a three-point star than of a ring (Fig. 1c).

The ordered N-termini of the C subunits of the sobemoviruses have been found to be arranged in two different ways. In SCPMV, SeMV, and TNV, the N-terminus folded back along the core of the C subunit and interacting with the subunits directly related by the threefold symmetry axis. In CfMV and RYMV, the N-terminal arm passes close to the icosahedral twofold axis toward the threefold icosahedral axis not directly adjacent to the core C subunit. Whichever the orientation of the N-terminus of C subunit is, it fulfils equivalent functions in particle size determination.

The coat protein of RGMoV has deletions in regions that are conserved in the other sobemoviruses for which structures are known. We have determined the structure of RGMoV to provide a better understanding of sobemovirus particle stability and assembly.

Results

Quality of the model

The electron density map resulting from the 60-fold noncrystallographic averaging was continuous throughout the main chain of all the three subunits. The interpretation of the density of the side chains was predominantly straightforward. The model could be built for most of the structure except for the disordered N-terminal parts. Residues 1–27 are not visible in the C subunit, whereas amino acids 1–57 and 1–56 are disordered in subunits A and B, respectively. The mapping of individual residue positions was easy with the CfMV structure used as an initial model.

The model of the icosahedral asymmetric unit consists of 565 amino acid residues. The R factor is 0.294 for the 569,869 reflections between 35 and 2.9 Å used in the refinement. The R_{free} is very similar to the R value due to the noncrystallographic symmetry (Kleywegt and Brunger, 1996). All the measured reflections were used in the refinement. One possible reason for the high R factor may be the generally weak reflections, in particular, at high resolution. The overall quality of the structure as judged by the Ramachandran plot showed less than 2.8% residues in disallowed regions using the strict criterion of Kleywegt and Jones (1996). The RMSD of bond lengths (0.01 Å) and bond angles (1.5°) shows tight restraints during refinement, which is proper for 2.9 Å resolution. The coordinates have been deposited in the Protein Data Bank (entry 2IZW).

Coat protein fold and capsid structure

The monomers of the RGMoV coat protein have a jellyroll β-sandwich topology, common to most nonenveloped icosahedral viruses. Fig. 1a shows a cartoon representation of the RGMoV subunit with the secondary structure annotation. The secondary structure elements are named according to the sobemovirus convention—β strands are denoted A to I and helices A to F. The interconnecting loops are named according to the adjacent β strands. The two β strands A1 and A2 are visible only in the C subunit. Some of the loops of the jellyroll β-sandwich proteins contain short helices. Three of them (A1 and A2 in the CD loop and B in the EF loop) can be found in most jellyroll viral coat proteins, whereas the C helix in the FG loop has been found only in sobemoviruses and TNV (Fig. 1b). RGMoV is unique among the sobemoviruses of known sequence in lacking this helix.

Table 1 shows the rms deviations of the Cα atoms among asymmetric units of different sobemoviruses and TNV after superposition. The coat protein of RGMoV appears to be slightly more similar to the coat protein of TNV than to the sobemoviruses. The main differences among the individual sobemoviruses and TNV are located in the loop regions. The regions in which the TNV protein backbone fold is more similar to RGMoV than to the other sobemoviruses are the BC, EF and GH loops.

The overall shape of the RGMoV virion is close to the rhombic triacontahedron with pentamers emanating distinctively at the fivefold, shallow clefts around the twofold, and small protrusions at the threefold icosahedral symmetry axis. Differences between the individual subunits are with the exception of the N-terminal arm relatively small. The rms deviation between Cα atoms of the subunits A and B is 0.54 for 178 residues. The difference between the A and C is even lower (0.51/175). Slightly larger deviations (0.70/175) are observed in the comparison of the B and C subunits. The differences between the subunits are located in the loop regions. For example, the B and C subunits differ in the area around the G1 strand and in the nine C-terminal residues, where the fold of the A subunit is closer to the C subunit. The differences in the subunit conformation reflect the different environments in the quasi-equivalent positions.

Table 1
Sequence similarity and structural similarity in the coat proteins of sobemoviruses and TNV

	RGMoV	SeMV	SCPMV	RYMV	CfMV	TNV
RGMoV	–	22	21	21	12	20
SeMV	1.8/87	–	63	20	17	27
SCPMV	1.8/86	0.6/100	–	23	19	24
RYMV	1.9/82	1.4/88	1.4/88	–	30	15
CfMV	1.8/79	1.4/87	1.4/87	1.4/97	–	21
TNV	1.4/86	1.4/97	1.4/97	1.5/86	1.5/87	–

Top right: percentage identity between the respective virus coat protein sequences. Gaps were ignored in the calculation. Bottom left: rms deviations (Å) of superimposed Cα atoms of the respective 3D structures. The second number indicates the percentage of available amino acid residues used for the calculations. The icosahedral asymmetric unit consisting of subunits A, B and C was used as a rigid body in all cases. The program O (Jones et al., 1990) was used for superposition of the molecules. The cutoff for inclusion of residues for the RMSD calculation was 3.8 Å.

Discussion

An empty calcium ion binding site

Calcium ions participate in the stabilization of virion architecture in many viruses. The interactions of coat proteins within the icosahedral asymmetric unit of sobemoviruses as well as tombusviruses involve metal-mediated carboxyl clusters.

Removal of the bivalent cations at neutral or basic pH causes expansion of the particles by approximately 7% and partial disruption of some of them. Particle swelling was observed in TBSV (Robinson and Harrison, 1982), *Cowpea chlorotic mottle virus* (CCMV) (Adolph, 1975), RYMV (Hull, 1977) and SCPMV (Hsu et al., 1976). The reason for the structural change is probably the repelling force of the acidic side chains that are no longer involved in metal binding. The swollen particles can be easily dissociated by increasing the ionic strength of the solution (Adolph and Butler, 1974). The swelling of particles is considered to be the first step in an infection-connected disassembly utilizing the low calcium ion concentration in the cell.

The calcium binding sites in sobemoviruses are located 9–11 Å from the quasi-threefold symmetry axis (Fig. 1d). These clusters involve the side chain oxygen of two aspartate and one asparagine residues, one carbonyl oxygen and the C-terminal carboxyl group. The calcium ion forms five coordination bonds to oxygen atoms of the capsid protein and possibly a further hydrogen bond to a water molecule. The residues involved in calcium binding are conserved in the RGMoV coat protein sequence and the local protein fold resembles other sobemoviruses. In RGMoV, the calcium ion clusters would involve the side chain oxygen of Asp 139, Asp142 and Asn 234, the carbonyl oxygen of Ser184 and the C-terminal oxygen of the coat protein—Gln 235.

We could not identify any density for calcium ions in the proximity of residues Asp 139 and Asp 142 (Fig. 1e). The likely reason for the absence of the calcium is the low pH (3.0) of the crystallization buffer. The side chains of the aspartates 139 and 142 are not charged under these conditions, and thus, the presence of the calcium ion is not required to counteract their repelling effect that otherwise would destabilize the capsid structure. There is no evidence of particle swelling in the structure. The crystallization buffer used for CfMV had a pH of 3.3 and the pH was lower than 4 in the case of RYMV, but unambiguous density for the metal ion was observed in both cases. Nevertheless, bivalent cations are necessary for RGMoV particle stability at neutral pH, since particles disassemble in solution of high ionic strength (0.6 M NaCl) containing 0.03 M EDTA at pH 7.0 (data not shown). It was possible to obtain the diffraction quality crystals of RGMoV only at the low pH.

The structure-based alignment of the C subunits of the coat protein sequences of the sobemoviruses RGMoV, CfMV, SCPMV, SeMV, RYMV, the necrovirus TNV and tymovirus *Turnip yellow mosaic virus* (TYMV) (Canady et al., 1996) is shown in Fig. 2. Disordered residues were not used in the

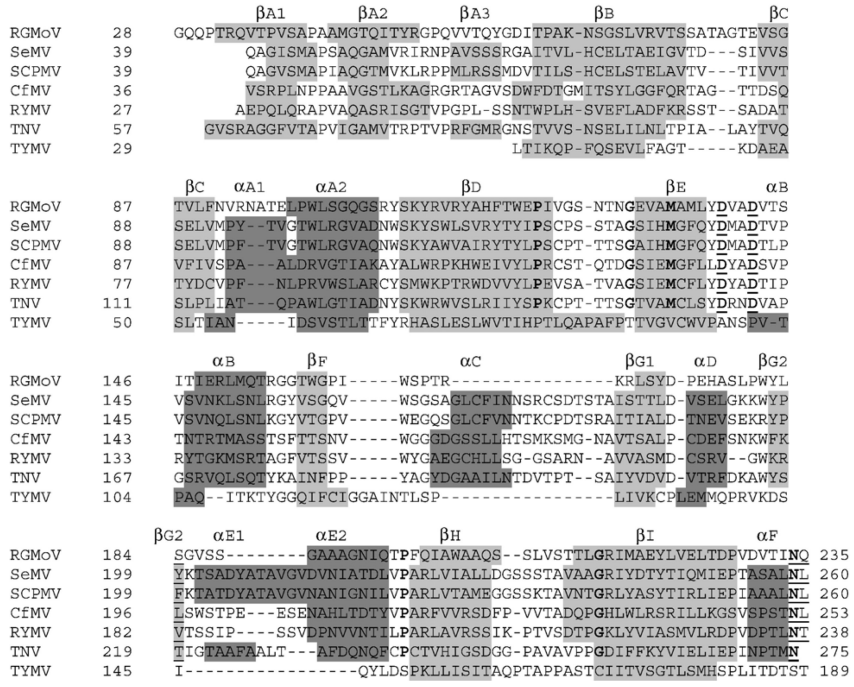

Fig. 2. Structure-based alignment of the coat protein sequences of the five sobemoviruses, TNV and TYMV. Bold letters represent conserved residues and underlined letters represent residues involved in calcium binding. The light grey blocks represent β strands and the dark grey blocks represent the α-helices. The alignment was prepared using the program Swiss-PdbViewer (Guex and Peitsch, 1997) with some manual adjustments.

alignment. Residues at eight positions are identical in the capsid protein of sobemoviruses and TNV (underlined bold letters in the alignment). Among the conserved residues, Asp 139, 142 and Asn 234 are involved in binding of the calcium ions. Of the five other invariant amino acid residues, two are prolines and two glycines, suggesting that these residues are important for the backbone conformation. The remaining conserved amino acid residue is a methionine, which is located in the E strand. The side chain of this methionine fills a hydrophobic pocket in the core of the coat protein, and thus, has a structural function in the same way as the other conserved amino acids.

RNA density in the RGMoV capsid

There are several sources of information on the genome arrangement in small T=3 viruses. In sobemoviruses it has been studied by low-angle electron scattering, cryoelectron microscopy and X-ray crystallography. Despite these methods supplying interconsistent complementary information, a precise determination of the RNA organization within particles has not yet been achieved. The low-angle neutron scattering of SCPMV supports a model where the genome is localized in three concentric shells of radii 30–55, 55–85 and 85–110 Å containing mostly RNA and about 15% protein, probably the disordered N-termini of all subunits (Kruse et al., 1982). Cryoelectron microscopy of RYMV and SCPMV virions revealed two internal layers of density that were attributed to partially ordered genomic RNA. The icosahedral symmetry of the capsid imposes icosahedral ordering on some portions of the RNA (Qu et al., 2000).

In the case of some plant T=3 RNA viruses–*Bean-pod mottle virus* (Chen et al., 1989), CCMV (Speir et al., 1995) and TYMV (Canady et al., 1996)–X-ray crystallography revealed that some parts of the genome were ordered. In these viruses, interaction of ordered amino acids with putative RNA bases was observed. The interaction could be of ionic origin involving lysine and arginine residues or based on stacking interactions with aromatic amino acid side chains. The crystal structures of sobemoviruses, as well as members of the *Tombusviridae* family, have shown only traces of the RNA.

There is only little density in the RGMoV map not corresponding to the polypeptide chains of the A, B and C subunits. Density belonging to putative RNA bases could be identified in several places stacked to side chains of aromatic amino acids exposed to the interior of virus capsid. An example illustrating electron density of two putative bases adjacent to tyrosine B115 is shown in Fig. 1f. This area reveals clear molecular boundaries of putative bases that are separate from density belonging to the capsid protein. The cytosine dinucleotide was fitted into the density to demonstrate proper distances for a stacking interaction between the bases and the tyrosine. The position of the nucleotides was not refined and they were not included into the model because most likely different bases occupy corresponding symmetry-related sites in the particle and became averaged by the random particle orientation in the crystal. Similar electron density for putative bases stacking to aromatic residues, although in different positions, has been observed earlier in CfMV (Tars et al., 2003).

Buried surface area comparison

The structured polymorphism of the T=3 sobemovirus capsid results in the same protein subunits being found in distinct structural environments within the particles generating a number of quasi-equivalent but different intersubunit interfaces. Table 2 shows the buried surface areas of respective intersubunit contacts of sobemoviruses, TNV and TYMV. The buried surface areas cannot be directly converted into binding energies, but they provide some information on the relative importance of the particular intersubunit contact in the architecture of a given virus.

All the analysed viruses have a total buried area corresponding to the contacts formed within and by a single asymmetric unit of 14,000–17,000 Å^2. The two highest values belong to CfMV and RYMV. These viruses have a different arrangement of the N-terminus of the C subunit than the remaining sobemoviruses and TNV, resulting in extended B1C1 and C1C6 interfaces. The C subunits form 40–50% of the intersubunit contacts. The C subunits form 41% of the contacts even in the TYMV capsid, where both the B and C subunits

Table 2
Buried surface area comparison

Virus	Buried surface area in 1000 Å^2 of subunits related by the icosahedral symmetry axis/named according to the standard T=3 icosahedral numbering											Fraction of total buried surface area by subunit		
	q3	q3	q3	5	q2	q6	2	q6	3	Other*	Total	A	B	C
	A1B1	B1C1	A1C1	A1A2	A2B1	B1C2	C1C6	B1C6	C2C6					
RGMoV	1.9	1.8	1.8	1.6	1.7	2.0	1.1	2.3	1.6		15.8	27%	31%	42%
SeMV	1.9	1.8	1.9	2.0	0.9	2.2	1.5	1.9	0.8		14.9	29%	29%	42%
SCPMV	2.1	2.1	2.1	2.0	1.2	2.3	1.4	1.9	0.8		15.9	30%	30%	40%
TNV	1.7	1.8	1.7	2.0	1.0	2.8	1.7	1.8	1.0		15.5	27%	29%	44%
CFMV	1.7	2.5	1.1	2.0	1.2	2.0	3.4	1.4		1.8	17.1	23%	27%	49%
RYMV	1.8	2.7	1.8	1.8	1.2	1.7	3.0	0.9		1.4	16.3	26%	27%	48%
TYMV	1.3	2.1	1.6	1.6	1.6	2.5	1.7	2.4			14.8	26%	33%	41%

Note. See Fig. 3i for the scheme of the standard T=3 icosahedral numbering.
* Other interfaces include mainly interactions of the N-termini of the C subunits.

contribute with their N-termini to the hexameric β-annulus and have identical structure within the 3 Å resolution of the model.

The relatively smaller FG loop of RGMoV results in a corresponding reduction of the area of the pentameric A1A2 and quasi-hexameric B1C2 interfaces. The contacts at A2B1, B1C6 and C1C16 are expanded making the pattern of intersubunit contacts areas of RGMoV more similar to TYMV than to other sobemoviruses. The apparent reason is the difference in the size of the FG loop of RGMoV and TYMV to other sobemoviruses and TNV. The important differences of RGMoV compared to SeMV, which was chosen as a representative of other sobemoviruses based on the highest sequence similarity to RGMoV, are summarized in Table 3 and discussed in detail below.

Quasi-threefold contacts

The sobemovirus capsids have small protrusions around the quasi-threefold axis, formed by the E1 and E2 helices in the GH loops of all the three subunits. These protrusions form part of the intersubunit contact interface. The size of the protrusions differs among sobemoviruses. The biggest protrusions together with the longest helices can be found in SeMV and SCPMV, which have both the E1 and E2 helices. RGMoV, CfMV and RYMV lack the E1 helix. The GH loop of RGMoV is missing eight residues in comparison to SeMV, which makes it the smallest among the so far structurally studied sobemoviruses. The top of the E2 helix in RGMoV is closer to the quasi-threefold symmetry axis and the whole helix is not protruding as far out from the particle as in SeMV (Figs. 3h and k). The reduced size of the GH loop in RGMoV coat protein results in the change of the relative subunit position in comparison to other sobemoviruses (Fig. 4c). The individual subunits of RGMoV are tilted 4° toward the quasi-threefold axis in comparison to SeMV (Fig. 4d). The relative tilt of the SeMV and SCPMV subunits is less than 1°. The shift in the relative subunit positions within the asymmetric unit induces changes in interactions between asymmetric units and contributes to the reduced particle size of the RGMoV as discussed in the next sections.

Deletion of the FG loop

An especially notable feature of RGMoV coat protein is the 16 amino acid deletion in comparison to SeMV in the FG-loop resulting in a complete absence of the C helix. The FG loop forms part of the A1A2, B1C2 and B1C6 quasi-equivalent interfaces in other sobemoviruses. The FG loop is involved in contacts with the F strand and the FG loop of the other subunit in the A1A2 (Fig. 3d) and B1C2 (not shown) interfaces. The region of the coat protein which is covered by the FG loop in other sobemoviruses becomes exposed to the interior of the RGMoV particle. The area of the missing C helix is partly filled by long side chains of residues from the D and I strands of the A1 subunit and Trp159 from the F strand of the A2 subunit at the A1A2 interface (Fig. 3a). The side chains do not form any interactions between the subunits and there are no extra contacts between the A1–A2 and B1–C2 RGMoV subunits to replace those of the full-size FG loop. The B1C2 interface is more open because of the flatter contacts around the threefold axis, but the residues mentioned at the A1A2 interface fulfill the same function in both of them.

The interface between the asymmetric units 1 and 2 consists of the contacts between subunits A1–A2, A2–B1, and B1–C2. The A2B1 interface is relatively enlarged in RGMoV (Table 2), most probably to compensate for the limited contacts in the other parts of the interface between subunits 1 and 2, which in other sobemoviruses involve the FG loop. The A2B1 interface

Table 3
Main structural differences between RGMoV and SeMV coat proteins

Structural element	Interface (Fig. 3i)	Structural function		Effect of the difference on the RGMoV capsid
		SeMV	RGMoV	
GH-loop (E1 and E2 helix)	A1B1, B1C1, A1C1	Parallel interaction of the E2 helices (Fig. 3k)	Missing E1 helix, limited interaction of the E2 helices (Fig. 3h)	All the RGMoV subunits are tilted 4° toward the centre of the asymmetric unit in comparison to SeMV (Fig. 4d)
FG-loop (C-helix)	A1A2	The C-helix is part of the A1A2 interface (Fig. 3d)	The volume taken by the FG loop in other sobemoviruses is partly filled by side chains of residues from neighbouring regions (Fig. 3a)	Reduced contacts around the icosahedral fivefold axis (Table 2)
	B1C6	The residues from the C-helix form hydrogen bonds to the β-annulus (Fig. 3e)	No extra interactions of the C6 subunit with the A2 strand to replace those of the missing C-helix (Fig. 3b)	FG-loop does not contribute to the stability of the β-annulus
N-terminus of the A and B subunits	A2B1	The N-termini and B strands of the A2 and B1 subunits do not form any interactions (Fig. 3j)	Residues from the N-termini and β-strands of the A2 and B1 subunits make several contacts (Fig. 3g)	The enlarged A2B1 interfaces compensates for the reduced interactions at the A1A2 and B1C2 interfaces (Table 2, Fig. 3i)
N-terminal arm of the C subunit	C2C6	Three hydrogen bonds between the A1 and A2 strands (Fig. 3f)	Eight hydrogen bonds between the A1 and A2 strands (Fig. 3c)	Extended contacts of the N-termini of the C subunits is probably required for the stability of the β-annulus

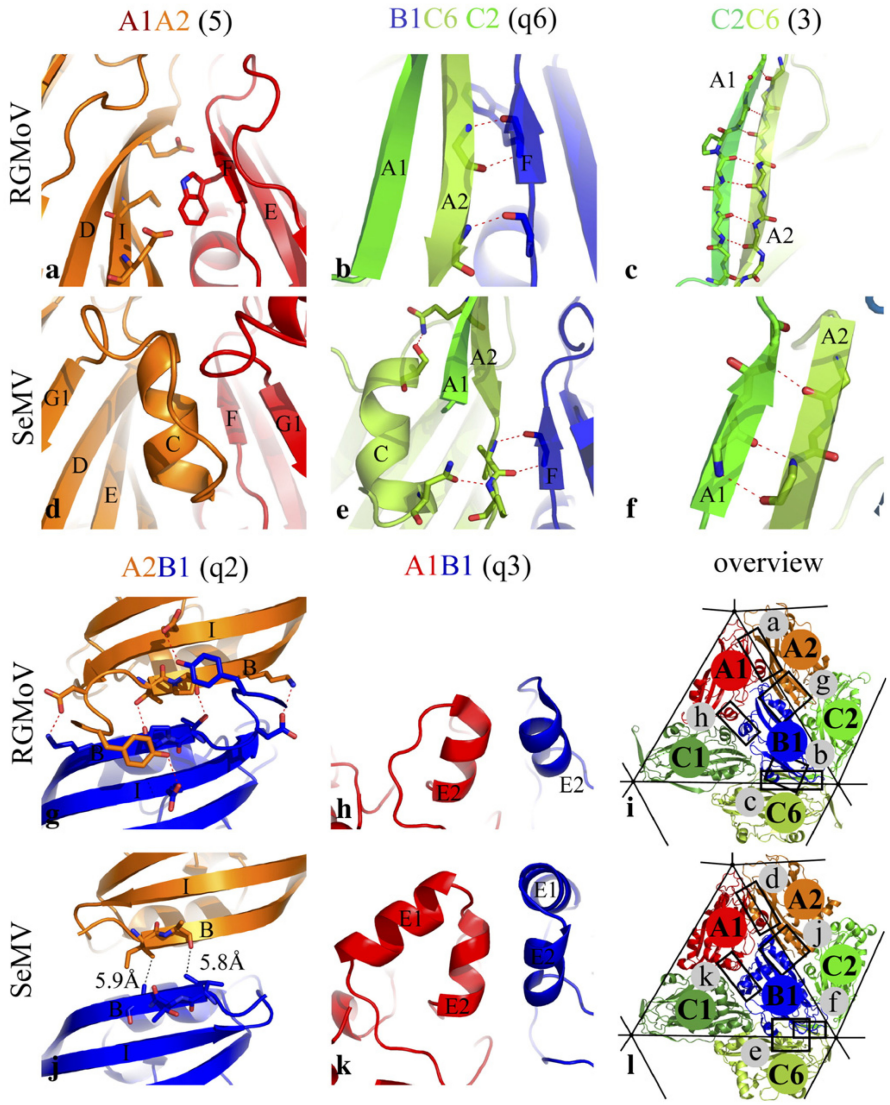

Fig. 3. Differences in the intersubunit interactions of RGMoV and SeMV. (a,d) Comparison of the interaction of the A subunits around the icosahedral fivefold axis of RGMoV (a) and SeMV (d). Side chains of RGMoV residues that are oriented toward the space occupied by the FG loop in SeMV are shown as sticks. (b,e) Interaction of the A1A2 sheet of RGMoV (b) and SeMV (e) with other parts of the capsid. In SeMV the C helix within the FG loop interacts with the A1A2 sheet and the following loop region. The A1 strand shown in dark green is from a threefold related C2 subunit. Residues forming hydrogen bonds are shown as sticks. (c,f) The interaction between the A1 and A2 strands of RGMoV (c) and SeMV (f) C2 (dark green) and C6 (green) subunits. (g,j) The interface between the A2 (orange) and B1 (blue) subunits of RGMoV (g) and SeMV (j). The side chains of the residues forming the hydrogen bonds between the RGMoV subunits are shown. In SeMV the residues from the B strands are too far from each other to form any interaction. (h,k) Comparison of the interaction between the E2 helices of the quasi-threefold related subunits A (red) and B (blue) of RGMoV (h) and SeMV (k). (i,l) Subunit annotation and schematic chart of the displayed interfaces of RGMoV (i) and SeMV (l). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

is formed by the F helices and, in the case of RGMoV, by the B strands and the ordered N-termini of the subunits (Figs. 3g and j). There are four extra residues at the N-terminus of the A subunit and five of the B subunit that are visible in the RGMoV density in comparison to SeMV (Figs. 3g, j). The asymmetric units 1 and 2 of RGMoV meet at an angle 42°, which is similar to 38° in SeMV and other sobemoviruses (Fig. 4a).

The RGMoV β-annulus formed by the extended N-termini of the C subunits is part of the B1C6 interface, making the subunit interactions different from the quasi-equivalent A1A2 and B1C2 contacts. The FG loop participates in stabilization of the β-annulus at the B1C6 interface in other sobemoviruses. In SeMV the residues from the A2 strand and the immediately following loop region of the C6 subunit interact with the C helix within the FG-loop of the same subunit (Fig. 3e). The A1A2 sheet is further stabilized by two hydrogen bonds to the F strand of the B1 subunit. The residues, which partly fill the area of the C helix at the quasi-similar A1A2 and B1C2 RGMoV contacts,

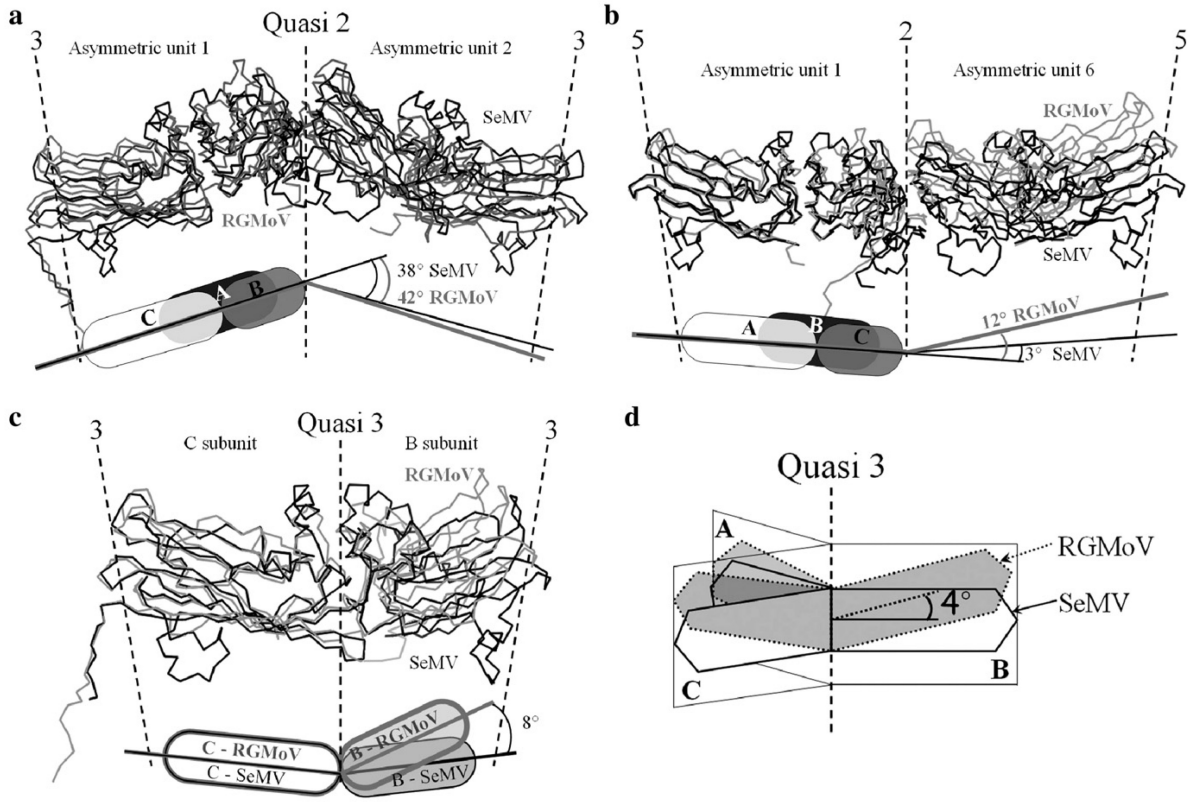

Fig. 4. Differences in the relative subunit and asymmetric unit positions between RGMoV and SeMV. (a) Comparison of the angle between asymmetric units 1 and 2 related by the icosahedral quasi-twofold symmetry axis of SeMV (black) and RGMoV (grey). The Cα atoms of the asymmetric units 1 of SeMV and RGMoV were superimposed. (b) Comparison of the angle between asymmetric units 1 and 6 related by the icosahedral twofold symmetry axis of SeMV (black) and RGMoV (grey). The Cα atoms of the asymmetric units 1 of SeMV and RGMoV were superimposed. (c) Relative tilt between the subunits B and C of SeMV (black) and RGMoV (grey). The Cα atoms of the subunits C of SeMV and RGMoV were superimposed. (d) Scheme of the relative tilt of the subunits of RGMoV (dotted line) in comparison to SeMV (solid line). All the three RGMoV subunits are tilted 4° toward the centre of the asymmetric unit in comparison to SeMV. The numeric values of the angles between the asymmetric units (subunits) were calculated as described in the Materials and methods section.

are in the B1C6 interface positioned far from the putative position of the C helix (as the Trp159 shown in the Fig. 3b) and do not interact with the N-terminus of the C subunit. Instead there is one extra hydrogen bond connecting the A2 strand of the C6 subunit to the FG-loop of the B1 subunit (Fig. 3b) and the extended interaction of the A1 and A2 strands in the RGMoV structure (Figs. 3c and f). The A1 strand faces the inner side of the particle and is extensively hydrogen bonded to the A2 strand of a threefold related subunit (Fig. 3c). Because the A1 strand does not interact with any other part of the capsid, its main role is probably in stabilization of the β-annulus. The ordered part of the N-terminal arm of the C subunit in the RGMoV structure is longer than in other sobemoviruses. There are six extra residues that could be modelled into the electron density of the C subunit in comparison to SeMV and SCPMV. The length of the β-sheet formed in the β-annulus is nine amino acids in RGMoV, while it is only five residues in all other sobemoviruses and TNV. The additional length of the sheet enlarges the binding interface (Table 2; C2C6 interface) and the number of hydrogen bonds between the A1 and A2 strands is increased from three in SeMV to eight in RGMoV (Figs. 3c and f).

A difference between the RGMoV and other sobemoviruses can be observed in the angle between the asymmetric units 1 and 6. In all the sobemoviruses, the interface is bent into the particle. In this respect, the sobemovirus virions differ from the ideal shape of the rhombic triacontahedron (angle 0°). The angle is only 3° for the SeMV, but 12° for RGMoV (Fig. 4b). The relative increase of the angle is a consequence of a combination of the shifted subunit position within the asymmetric unit and the altered interactions between the asymmetric units related by the icosahedral twofold symmetry axis.

Virion size

The icosahedral symmetry and higher T numbers enable viruses to increase the volume of the virion cavity without spending too much of the space available for the genome on the coat protein coding sequence. Table 4 shows some of the capsid

Table 4
Comparison of the particle volume and genome content of single-stranded RNA viruses

Virus	PDB ID	Particle		Fraction of particle volume taken by		Genome	
		Radius (Å)	Volume (10^6 Å^3)	FG loops	Disordered regions	Size (nt)	RNA concentration (mg/ml)
RGMoV	2IZW	125	4.9	0%	19%	4212	478
CfMV	1NG0	128	5.1	5%	19%	4082	438
RYMV	1F2N	128	5.3	4%	16%	4451	459
SCPMV	4SBV	129	5.1	6%	21%	4194	451
SeMV	1SMV	128	5.2	6%	21%	4149	438
TNV	1TNV	127	4.5	5%	30%	3762	461
TYMV	1AUY	126	5.8	0%	3%	6318	595
MS2	2B2G	119	5.4	n.a.	n.a.	3600	367
CCMV	1CWP	115	2.0	n.a.	26%	3200	870
Pariacoto	1F8V	131	3.6	n.a.	19%	4322	662
Norwalk	1IHM	157	7.4	n.a.	6%	7654	567
STMV	1A34	66	6.6	n.a.	13%	1058	878
Poliovirus	1HXS	129	5.0	n.a.	3%	7440	814

Note. Part of the coat protein of all viruses except MS2 is not visible in the crystal structure and is located inside of the particle. Volume of the disordered polypeptide chain was estimated and included in the calculations. Values were calculated as described in the Materials and methods section.
n.a.—not applicable.

and genome characteristics of some small RNA viruses. Sobemoviruses form quite a homogenous group considering the particle and genome size. However, RGMoV has about 3% smaller particle diameter than other sobemoviruses.

The important structural differences of RGMoV in comparison to other sobemoviruses contribute to the relatively smaller particle diameter of the RGMoV. The reduction of the GH loop results in relative tilting of the subunits toward the quasi-threefold axis and, together with the absence of the C helix in the FG loop, results in the 12° bending of the interface between the asymmetric units related by the icosahedral twofold symmetry axis into the particle. The relative increase of the tilt between the twofold related icosahedral asymmetric units in comparison to other sobemoviruses is the structural reason for the relative reduction of the RGMoV particle size. The extended contacts of the N-termini of the C subunits in RGMoV are probably necessary to compensate for the missing contacts of the β-annulus with the FG loop.

The FG loops in other sobemoviruses occupy, on average, 5% of the virion volume. The decreased diameter and the missing FG loop in RGMoV have an opposing effect on the size of the space available for genome storage. Overall, the size of the RGMoV virion cavity is 7% smaller than in other sobemoviruses. Nevertheless, the RNA packing density of 480 mg/ml is similar to other sobemoviruses and TNV. On average, 20% of the sobemovirus particle cavity is taken by the disordered N-terminal parts of the coat proteins. It is very likely that the flexible arms do not impose any additional restraints on RNA packing. If the volume taken by the N-termini is considered available for genome storage, then the average RNA packing density in sobemoviruses would be 370 mg/ml. The genome packing density varies significantly among different families of small single-stranded RNA viruses as demonstrated by a few examples shown in Table 4. Bromoviruses represented by CCMV, picornaviruses (poliovirus) and *Satellite tobacco mosaic virus* (STMV) have almost double RNA packing density than sobemoviruses. Caliciviruses (Norwalk virus), nodaviruses (pariacoto virus) and leviviruses have similar RNA density within particles as sobemoviruses. It is not clear whether the differences in the RNA packing density between the virus families can be attributed to a distinct type of RNA packing or if the viruses with lower RNA density simply have more space between the secondary structure elements of RNA. The genome packing represents an interesting field for further study.

Materials and methods

Virus production, purification and sequence verification

RGMoV (MAFF No. 307043) was received from the MAFF Genebank at the National Institute of Agrobiological Sciences, Japan. Oat plants cv. “Jaak” were mechanically inoculated at the 2–3 leaves growth stage and cultivated for 4 weeks at temperatures fluctuating between 21 and 23 °C under natural and artificial light 17 h/day.

For virus purification, harvested plants were homogenized in 5 volumes of 15 mM potassium phosphate buffer (pH 5.5) and filtered to remove the plant debris. The filtrate was extracted with 0.4 volumes of chloroform and centrifuged at 3000×*g* for 20 min. Solid ammonium sulphate was added stepwise to the resulting solution for a final concentration of 30%. The precipitate was collected by centrifugation, solubilized in distilled water to 1/10 of the original volume, and dialysed against 200 volumes of distilled water.

Virus particles were collected by 1 h centrifugation at 280,000×*g*. The virus pellet was resuspended in 15 mM potassium phosphate buffer (pH 5.5) and centrifuged 1 h at 280,000×*g* through a 20% sucrose cushion. After three repeats of sucrose cushion steps the virus particles were resuspended in 15 mM potassium phosphate buffer (pH 5.5) to the final concentration of approximately 15 mg/ml.

RGMoV RNA was isolated using TRI-reagent (Sigma, Saint Louis, MO, USA) according to manufacturer’s protocol. The

RGMoV cDNA copy was prepared in a M-MuLV reverse transcriptase reaction and amplified by *Pfu* DNA polymerase using two primers for RGMoV coat protein gene. RGMoV coat protein cDNA PCR product was cloned into plasmid pTZ57 and isolated plasmid DNA was sequenced in both directions using M13 direct and reverse primers. The sequence data have been submitted to the Genbank nucleotide sequence database and have been assigned the accession number EF091714.

Crystallization and data collection

Crystals of RGMoV were obtained using the hanging drop technique with a bottom solution containing: 300 mM Li_2SO_4, 50 mM Na–citrate at pH 3.0, 1% PEG 8000. The drops were prepared by mixing 1.5 μl of bottom solution with an equal volume of virus solution (8 mg/ml). Block-shaped crystals formed within 7–10 days. For data collection, crystals of size up to 0.2 mm were soaked for 30–90 s in mother liquor containing 30% PEG 4000 and immediately frozen in liquid nitrogen. Data were collected from a single crystal at 100 K on the ADSC Quantum4 CCD detector at beamline ID 14-1 at the ESRF synchrotron radiation source in Grenoble, France. An oscillation range of 0.2° was used during data collection. The crystal diffracted to 2.9 Å resolution. Data were processed and scaled using the programs Denzo and Scalepack from the HKL package (Otwinowski and Minor, 1996). The statistics from data collection and scaling is shown in Table 5.

Structure determination

The RGMoV crystals were of space group $P2_1$. A complete virus particle occupied a crystallographic asymmetric unit. Therefore, three parameters for the particle orientation and two for the position of the particle centre had to be determined. The locked self-rotation function in the program GLRF was used to identify the particle orientation (Tong and Rossmann, 1997). Reflections between 7 and 2.9 Å were used for the calculations. The radius of integration was set to 280 Å. The results suggested that the model particle needs to be rotated $\varphi = 3.3°$, $\phi = 57.6°$, $\kappa = 146.7°$ from the standard icosahedral orientation according to the polar angle convention. Translation parameters could be identified in the native Patterson function since one of icosahedral twofold symmetry axis is almost parallel to the crystallographic twofold screw axis. A clear peak appeared at fractional coordinates 0.569 1/2 0.499, defining the position of the particle centre at $x = 64.7$ Å, $z = 98.0$ Å. The model of CfMV (PDB entry 1NG0) was used for the molecular replacement. The model was placed into the approximate orientation and position in the unit cell and an initial low-resolution electron density map was calculated suggesting that the particle has a slightly smaller radius than the model. The model was moved manually toward the particle centre and subjected to rigid body refinement in CNS (Brünger et al., 1998). The total shift of the model was 3.5 Å. The new model was used to calculate phases to 10 Å resolution in CNS. The phases were refined by 25 cycles of averaging by the program AVE (Kleywegt and Read, 1997), using the 60-fold noncrystallographic symmetry. Phase extension was applied in order to obtain phases for higher resolution reflections. Addition of a small fraction of higher resolution data (one index at a time) was followed by 5 cycles of averaging. This procedure was repeated until phases were obtained for all the reflections to 2.9 Å resolution.

The model was built using the program O (Jones et al., 1990), starting from the CfMV coordinates where the respective residues were mutated to those of RGMoV. The model was refined by manual rebuilding alternating with coordinate refinement in the program CNS (simulated annealing, gradient minimization, individual b-factor refinement). Other calculations were done using CCP4 (1994). No water molecules were added to the model.

Table 5
Scaling statistics

Space group	$P2_1$
Unit cell dimensions (Å)	
a	277.64
b	298.72
c	392.49
β (°)	92.74
Resolution limits (Å)	35–2.9
High-resolution bin (Å)	2.95–2.90
Completeness (%)	40 (19)
R_{merge} *	0.12 (0.63)
I/σ	5.30 (0.82)
N reflections	569,869 (37,205)
Redundancy	1.5 (1.0)

Note. Numbers in parentheses are for the highest resolution shell.
* $R_{merge} = \Sigma h \Sigma j |Ihj - <Ih>| / \Sigma \Sigma Ihj$.

Calculation of the particle radius and volume

The particle radius was defined as a distance of the centre of mass of the asymmetric unit from the centre of the particle. Equivalent asymmetric units were chosen for all the viruses.

A single PDB file containing coordinates for the whole virus particle was generated for each virus. The volume of the particle was then calculated with the use of the program MAMA (Kleywegt and Jones, 1999). Two masks were generated: the first covering the protein shell and the second one covering the whole particle including the internal virion cavity. The difference of the two volumes defined the size of the particle cavity. The extra volume of the FG loops in comparison to RGMoV was calculated using the program Voidoo (Kleywegt and Jones, 1994). The volume of the disordered regions of the coat protein subunits was estimated using the Peptide property calculator (http://www.basic.northwestern.edu/biotools/proteincalc.html). Volumes of individual residues used in the calculation were based on the work of Harpaz et al. (1994).

Calculation of the relative tilt of the subunits in the asymmetric unit

The models of whole asymmetric units of RGMoV and SeMV were superimposed in the program O. The quasi-threefold symmetry axes of RGMoV and SeMV were compared

and found to be almost identical. Reference "atoms" Q^A, Q^B, Q^C at the centre of mass of the respective RGMoV subunits and P at the centre of mass of the RGMoV asymmetric unit were added to the coordinates of SeMV. Then the A subunit of SeMV model was superimposed on the A subunit of RGMoV. New coordinates of the reference atoms P (P^{A2}) and Q^A (Q^{A2}) were noted. The angle between the vectors $\mathbf{PQ^A}$ and $\mathbf{P^{A2}Q^{A2}}$ gives the relative tilt of the RGMoV and SeMV A subunits. The angle between the quasi-threefold axis and a plane defined by point P and vectors **PQ** and $\mathbf{P^{A2}Q^{A2}}$ was calculated and found to be smaller than 5°, which means that the tilting of the subunit is directed predominantly toward the quasi-threefold axis. The procedure was repeated for the B and C subunits. An identical comparison of SCPMV and SeMV was carried out.

Calculation of the angles at the interface between asymmetric units

Coordinates of the asymmetric units 1, 2 and 6 (according to the standard T=3 quasi-icosahedral notation) were generated (the N-terminus of the C subunit of the respective virus was trimmed to the length of the A subunit). Centres of masses of individual subunits were calculated and used to define the plane of respective asymmetric unit. Angles between the planes characterizing asymmetric units (1 and 6) and (1 and 2) were calculated.

Acknowledgments

MAFF Genebank is acknowledged for the RGMoV isolate. We thank the staff at the ESRF France. P.P. acknowledges very helpful advice on various calculations from Gerard Kleywegt. This work was supported by the Swedish Research Council, the Latvian Research Council and EU project no. VPD/ERAF/CFLA/05/APK/2.5.1./000031/016.

References

Abad-Zapatero, C., Abdel-Meguid, S.S., Johnson, J.E., Leslie, A.G.W., Rayment, I., Rossmann, M.G., Suck, D., Tsukihara, T., 1980. Structure of southern bean mosaic virus at 2.8 Å resolution. Nature 286, 33–39.

Adolph, K.W., 1975. The conformation of the RNA in cowpea chlorotic mottle virus: dye-binding studies. Eur. J. Biochem. 53, 449–455.

Adolph, K.W., Butler, P.J.G., 1974. Studies on the assembly of a spherical plant virus: I. States of aggregation of the isolated protein. J. Mol. Biol. 109, 327–341.

Balke, I., Resevica, G., Zeltins, A., 2007. The Ryegrass mottle virus genome codes for a sobemovirus 3C-like serine protease and RNA-dependent RNA polymerase translated via −1 ribosomal frameshifting. Virus Genes 35, 395–398.

Bhuvaneshwari, M., Subramanya, H.S., Gopinath, K., Savithri, H.S., Nayudu, M.V., Murthy, M.R.N., 1995. Structure of sesbania mosaic virus at 3 Å resolution. Structure 3, 1021–1030.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., Warren, G.L., 1998. Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr., D 54, 905–921.

Canady, M.A., Larson, S.B., Day, J., McPherson, A., 1996. Crystal structure of turnip yellow mosaic virus. Nat. Struct. Biol. 3, 771–781.

CCP4, 1994. The CCP4 suite: programs for protein crystallography. Acta Crystallogr., D Biol. Crystallogr. 50, 760–763.

Chen, Z., Stauffacher, C., Li, Y., Schmidt, T., Bomu, W., Kamer, G., Shanks, M., Lomonossoff, G., Johnson, J.E., 1989. Protein–RNA interactions in an icosahedral virus at 3.0 Å resolution. Science 245, 154–159.

Guex, N., Peitsch, M.C., 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18, 2714–2723.

Harpaz, Y., Gerstein, M., Chothia, C., 1994. Volume changes on protein folding. Structure 2, 641–649.

Harrison, S.C., Olson, A.J., Schutt, C.E., Winkler, F.K., Bricogne, G., 1978. Tomato bushy stunt virus at 2.9 Å resolution. Nature 276, 368–373.

Hsu, C.H., Sehgal, O.P., Pickett, E.E., 1976. Stabilizing effect of divalent metal ions on virions of southern bean mosaic virus. Virology 69, 587–595.

Hull, R., 1977. The stabilization of the particles of turnip rosette virus and of other members of the southern bean mosaic virus group. Virology 79, 58–66.

Jones, T.A., Bergdoll, M., Kjeldgaard, M., 1990. O: a macromolecule modeling environment. In: Bugg, C., Ealick, S. (Eds.), Crystallographic and Modeling Methods in Molecular Design. Springer-Verlag, New York, pp. 189–199.

Kleywegt, G.J., Brunger, A.T., 1996. Checking your imagination: applications of the free R value. Structure 4, 897–904.

Kleywegt, G.J., Jones, T.A., 1994. Halloween… masks and bones. In: Bailey, S., Hubbard, R., Waller, D. (Eds.), From First Map to Final Model. Proceedings of the CCP4 Study Weekend. SERC Daresbury Laboratory, Daresbury, pp. 59–66.

Kleywegt, G.J., Jones, T.A., 1996. Phi/psi-chology: Ramachandran revisited. Structure 4, 1395–1400.

Kleywegt, G.J., Jones, T.A., 1999. Software for handling macromolecular envelopes. Acta Crystallogr., D Biol. Crystallogr. 55, 941–944.

Kleywegt, G.J., Read, R.J., 1997. Not your average density. Structure 5, 1557–1569.

Kruse, J., Timmins, P.A., Witz, J., 1982. A neutron scattering study of the structure of compact and swollen forms of SBMV. Virology 119, 42–50.

Lokesh, G.L., Gowri, T.D., Satheshkumar, P.S., Murthy, M.R., Savithri, H.S., 2002. A molecular switch in the capsid protein controls the particle polymorphism in an icosahedral virus. Virology 292, 211–223.

Meier, M., Truve, E., 2007. Sobemoviruses possess a common CfMV-like genomic organization. Arch. Virol. 152, 635–640.

Oda, Y., Saeki, K., Takahashi, Y., Maeda, T., Naitow, H., Tsukihara, T., Fukuyama, K., 2000. Crystal structure of tobacco necrosis virus at 2.25 A resolution. J. Mol. Biol. 300, 153–169.

Otwinowski, Z., Minor, W., 1996. Processing of X-ray diffraction data collected in oscillation mode. In: Carter Jr., C.W., Sweet, R.M. (Eds.), Methods in Enzymology, vol. 276. Academic Press, New York, pp. 307–326.

Qu, C., Liljas, L., Opalka, N., Brugidou, C., Yeager, M., Beachy, R.N., Fauquet, C.M., Johnson, J.E., Lin, T., 2000. 3D domain swapping of a molecular switch for quasi-equivalent symmetry modulates the stability of an icosahedral virus. Structure 8, 1095–1103.

Robinson, I.K., Harrison, S.C., 1982. Structure of the expanded state of tomato bushy stunt virus. Nature 297, 563–568.

Rossmann, M.G., Abad-Zapatero, C., Erickson, J.W., Savithri, H.S., 1983. RNA–protein interactions in some small plant viruses. J. Biomol. Struct. Dyn. 1, 565–579.

Speir, J.A., Munshi, S., Wang, G., Baker, T.S., Johnson, J.E., 1995. Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. Structure 3, 63–78.

Tamm, T., Truve, E., 2000. Sobemoviruses. J. Virol. 74, 6231–6241.

Tars, K., Zeltins, A., Liljas, L., 2003. The three-dimensional structure of cocksfoot mottle virus at 2.7 Å resolution. Virology 310, 287–297.

Tong, L., Rossmann, M.G., 1997. Rotation function calculations with GLRF program. Methods Enzymol. 276, 594–611.

Zhang, F.Y., Toriyama, S., Takahashi, M., 2001. Complete nucleotide sequence of ryegrass mottle virus: a new species of the genus sobemovirus. J. Gen. Plant Pathol. 67, 63–68.

DISCUSSION

Sobemovirus genome is single stranded messenger sense RNA. As revealed first studies on genome organization based on SCPMV, Sobemoviruses genomes contain four ORFs. One at 5' end proximal part, large ORF in the middle part, small ORF completely nested inside of the large ORF, and 3'-end proximal ORF (Fig. 1B). This genome organization was accepted as correct and trustful until Makinen et al., (1995) reported that member of *Sobemoviruses* CfMV posses different genome organization in the middle part of the genome. Instead of large ORF2, it was demonstrated to contain two overlapping ORFs – ORF2a and ORF2b, and the small ORF3 in the middle of ORF2 was not found. So the members of Sobemovirus group were classified in two groups – viruses with SCPMV-like and CfMV-like genome organization. Even some newly discovered genus members were still characterized to have SCPMV-like genome organization, including RGMoV (Zhang *et al.*, 2001b).

CfMV-like genome organization of was regarded as unusual because CfMV was the only member with such organization for almost 10 years. Later, TRoV (Callaway *et al.*, 2002) and SCMoV (Dwyer *et al.*, 2003) were classified as new members of the genus. These memebers of *Sobemovirus* group were found to possess CfVM-like genome organization. In 2004, Fargette et al. characterized 14 different RYMV isolates; the sequence analysis revealed that the virus have the CfMV-like genome organization and not SCPMV-like, as proposed earlier.

As one of the main tasks for our work, was to obtain full-length cDNA of RGMoV, we resequenced all RGMoV genome. Sequence analysis of cDNA clones covering all the genome revealed some sequence dissimilarities. In the first moment it looked like insignificant sequence polymorphism of a virus population as it was demonstrated in case of other sobemoviruses. But deeper analysis revealed first of all the change of RGMoV genome organization from SCPMV-like to CfMV-like as it was in RYMV case. We repeated RT-PCR both from infected leaves and isolated virus particles and constructed several new RGMoV cDNA fragments containing the region under study. The second round sequencing data confirmed our previous data. The genome organization resulted change was located in seven G stretch after position 2209 where in our case was six. That resembled in formation of separate ORF coding for RdRp via – 1 ribosomal frameshift as in CfMV case. Additional analysis was made for ribosomal frameshif signal components – slippery heptonucletides and hairpin structures, identification to confirm RdRp possible translation via – 1 ribosomal frameshift. All necessary sequence elements were detected before ORF2b as it is described (Makinen *et al.*, 1995).

Some nucleotide changes were found to be neutral, but T insertion in position 1319 and deletion of G at position 1477 resulted in new 52 AA long stretch inside of the protease coding region. Multiple sequence analyses revealed tipical sobemoviral 3C-serine protease AA arrangement. Newly identified AA stretch contained putative catalytic triad of sobemovirus protease. Only the linker between H and D was longer

than in other sobemoviruses, but in still good agreement with other 3C-like proteases. This sequence contains all three catalytic triad AA, whereas in previous report of protease serine was missing. Reported alanine in catalytic triade raised question about Pro activity. Similar situation was with substrate-binding site AA, where previous was reported proline not threonin as it must be. In the new 52 AA stretch were found serine for catalytic site and threonine for substrate-biding site, so overturn speculation about RGMoV Pro activity (Publication I).

Genome sequencing errors have been observed for Sobemoviruses earlier in SBMV, SBMV-Ark (Lee and Anderson, 1998) where ORFs were miss located, and LTSV (Jeffries *et al.*, 1995), where ORF1 was divided in to two separate ORFs – ORF1a and ORF2b.

Simultaneously with our publication it was demonstrated that that all sobemoviruses possess CfMV-like genome organization and not SCPMV-like as it was proposed earlier. All *Sobemovirus* SCPMV-like genome organizations were proposed based on errors in sequencing data (Meier and Truve, 2007).

In the last years new information based on *in vitro* and bioinformatics analysis of sobemovirus encoded nonstructural proteins started to reveal there possible functions and properties. Those results are as impulse for new studies in this field. The first characterized proteins were P1 from RYMV (Bonneau *et al.*, 1998) and SCPMV (Sivakumaran *et al.*, 1998), later the first data about P1 of CfMV were obtained (Tamm and Truve, 2000a). For P1, several functions have been identified; it could serve as movement protein, similarly to other plant viruses. Its inactivation inhibited the virus movement from infected cell into the next and systemic plant infection was disabled. There was reported that P1 of CfMV-NO suppresses RNA silencing in *Nicotiana benthamiana* and *Nicotiana tabacum*, two non-host plants. CfMV-NO P1 was able to suppress the initiation and maintenance of silencing. The suppression of systemic silencing was weaker with CfMV-NO P1 than in the case of RYMV-N P1 (Sarmiento *et al.*, 2007). There was reported that P1 can bind nucleic acids in sequence nonspecific way (Tamm and Truve, 2000a). These experiments were carried out using CfMV infectious cDNA constructs or proteins obtained from bacterial expression system. RYMV P1 protein contains putative eukaryotic Zn-finger motif ($C_{64}(X2)C_{67}(X24)C_{92}(X2)C_{95}$); conserved AA C95 and C64 of the motif was shown to be involved in cell-to-cell movement and in the suppression of gene silencing (Sire *et al.*, 2008). Interestingly, similar Zn-finger motif can be found also in RGMoV P1 sequence ($C_{55}(X2)\ C_{58}(X9)H_{68}(X2)H_{71}$).

For protein structure and possible function analysis several *in silico* methods are used. Those methods were used for CfMV and SeMV proteases N-terminal domain characterization (Ryabov *et al.*, 1996; Satheshkumar *et al.*, 2004), identification of folding nature of VPg and P18 for SeMV (Satheshkumar *et al.*, 2005a; Nair and Savithri, 2010a).

RGMoV VPg and P16 nature similar as in SeMV case was analysed using *in silico*-based methods. Our results confirm VPg and P16 unfolded nature identified from analysis *in silico* and from several experiments such as crystallization and polyclonal antibodies production in VPg case (Publication II).

As for P16 additional experiment was made to test its nucleic acid binding properties. A native gel sift was carried out with RGMoV cDNA and control RNA transcript and P16 analysis of the gel revealed P16 binding properties to the nucleic acids, but as P16 banded control RNA it binding is in sequence nonspecific way.

Pro analysis *in silico* confirmed that N-terminal end contains transmembran binding domain. Expression of full length Pro failed, no detectable expression was observed (data not shown). On the other hand N-terminally truncated protease was expressed in *E.coli* and was soluble. As expression vector contined N-terminaly fused His-tag, a purification experiment was made. As revealed SDS-PAGE analysis Δ50Pro failed to purifie with His-tag column. It could be explained by possible protease autocatalyse as it was observed in SeMV protease case. It was discovered that SeMV protease cleaves at N-terminal located part betvean A/V (Gayathri *et al.*, 2006). Other explanation can be based on to FoldIndex analysis, which revealed a short disordered region between AA 76 and 83. It contines Lys and Arg residues. This region can be cleaved by *E.coli* trypsin-like enzymes (Pantazatos *et al.*, 2004), resulting in N-terminally truncated protein without His-tag.

The most characerized proteins of Sobemoviruses are their coat proteins. Alltogether, six Sobemovirus 3D structures are solved including RGMoV. Different CP capsid formation and stability tests are made. It was shown that RYMV virion svells (Fig. 7) in environment contining EDTA, that removes Ca^{2+} from virus particles and so changing their stability (Brugidou *et al.*, 2002). The N-terminal region is known for the RNA binding properties, removal of it smaller particles with a T=1 symmetry can be obtained (Rossmann *et al.*, 1983a; Lokesh *et al.*, 2002), where all the contacts between subunits are bent. Removal of β-annulus does not affect VLP formation. This segment could be involved in molecular switch that determines the T=3 capsid assembly. They were less stable than native virus (Pappachan *et al.*, 2008). Single Trp-170 change to a charged residue (Lys or Glu) of SeMV resulted in total disruption of virus assembly and resulted in the formation of soluble dimers less stable than T=3 and T=1 (Pappachan *et al.*, 2009). RGMoV structure revealed several differences in CP structure in comparison with Sobemoviruses possessing 3D structures. CP is missing several residues in two of the loop regions. The first loop contributes to contacts between subunits around the quasi-threefold symmetry axis. The altered contact interface results in tilting of the subunits towards the quasi-threefold axis. The other loop that is smaller in the RGMoV structure contains a helix that participates in stabilization of the β-annulus in other sobemoviruses. The loss of interaction between the RGMoV loop and the β-annulus has been compensated for by additional interactions between the N-terminal arms. As revealed recent data

regarding β-annulus they are not required for virus structure assembly (Pappachan *et al.*, 2008). Reported differences affect the diameter of RGMoV particles which is 8 Å smaller than that of the other sobemoviruses. The reduction of the GH loop results in relative tilting of the subunits toward the quasi-threefold axis and, together with the absence of the C helix in the FG loop, results in the 12° bending of the interface between the asymmetric units related by the icosahedral twofold symmetry axis into the particle. The FG loops in other sobemoviruses occupy, on average, 5% of the virion volume. The decreased diameter and the missing FG loop in RGMoV have an opposing effect on the size of the space available for genome storage. Overall, the size of the RGMoV virion cavity is 7% smaller than in other sobemoviruses. Nevertheless, the RNA packing density of 480 mg/ml is similar to other sobemoviruses and TNV. In obtained structure model, no Ca^{2+} and special RNA binding was observed, probably, due to acid pH of the buffer used for RGMoV crystal formation. However, the incubation the virions in EDTA and NaCl-containing buffer demonstrated RGMoV particle destabilization, suggesting that divalent cations, possibly Ca^{2+} is needed for particle stability (Publication III).

CONCLUSIONS

1. Using RGMoV RNA, isolated from infected plant material or intact virions, as template for RT-PCR five cDNA fragments were amplified and combined in to full length cDNA of RGMoV.
2. Analysis of cDNA sequence revealed several important properties of RGMoV genome:
    a. RGMoV genome organization have to be classified as CfMV-type; where RdRP is translated via -1 ribosome frameshifting mechanism;
    b. RGMoV ORF2a codes for sobemovirus 3C-like serine protease with strongly conserved catalytic triad and substrate binding amino acids.
3. All non-structural proteins of RGMoV were subcloned in *E.coli* expression vectors. Following vector and *E.coli* host strain combinations were found to to be suitable for RGMoV protein expression :
    a. for protease - pColdI vector and WK6 cells;
    b. for VPg - pACYC-Duet-1 vector and BL21(DE3) cells;
    c. for P16 - 28a(+) vector and BL21(DE3) cells;
    d. RdRp in detectable amounts was expressed from pColdI-derivate and WK6 cells, but the recombinant protein was insoluble.
4. Unsuccessful crystallization experiments and bioinformatic analysis suggested natively unfolded structure of VPg and P16.
5. P16 due to its basic C-terminal part posses RNA binding properties in sequence nonspecific way.
6. The 3D structure of native RGMoV has been determined at 2.9 Å. It contains several differences in structural elements in comparison of other sobemoviruses:
    a. RGMoV posses T=3 icosahedral symmetry similar to other crystallized sobemoviruses and the CP has a canonical jellyroll β-sandwich fold;
    b. Absence of C helix of FG loop and E1 helix in GH loop;
    c. Decrease of capsid diameter by 3 % and inner volume by 7 % due to missing structural elements.

MAIN THESIS FOR DEFENCE

1. RGMoV possess CfMV-like genome organization and RdRp is translated via -1 ribosomal frameshift.

2. RGMoV encoded 3C serine-like protease contains all amino acids for catalytic and substrate binding centers.

3. VPg and P16 accordingly to bioinformatic analysis and lack of crystal formation meet the criteria of natively disordered proteins.

4. RGMoV possess T=3 icosahedral symmetry. The coat protein has a canonical jellyroll β-sandwich fold. For RGMoV capside assembly FG and GH loops are not necessary as compensation elements in assembly are compensated by additional interactions between the N-terminal arms.

ACKNOWLEDGEMENTS

Want to thank my supervisor Dr.biol Andris Zeltiņš for opportunity to do my PhD research under his leadership, the interesting discussions, and his patient in reading and revising my Promotional paper, as well as the opportunity to work at this interesting topic.
Laboratory team Vilija Zeltiņa, Gunta Resevičа, Ieva Kalnciema and Jeļena Šaripo for the advices, support, as well as the interesting discussions.
Dr.biol. Kaspars Tārs for fruitful consultations.
Gunda Grīnberga for centrifugation.
Dr. Ligija Ignatoviča for sequenation of the analysis.
Dr. Dace Skrastiņa for polyclonal antibody preparation.
To my husband Dāvids Fridmanis and my little daughter Agate for patient and encouragement during all PhD studies and Promotion paper writing.
My parents for their help.

REFERENCES

Abad-Zapatero, C., Abdel-Meguid, S. S., Johnson, J. E., Leslie, A. G., Rayment, I., Rossmann, M. G., Suck, D. and Tsukihara, T. (1980) Structure of southern bean mosaic virus at 2.8 A resolution. *Nature* **286**(5768): 33-9.

AbouHaidar, M. G. and Paliwal, Y. C. (1988) Comparison of the nucleotide sequences of viroid-like satellite RNA of Canadian and Australasian strains of lucerne transient streak virus. *j gen virol* **69**: 2369-2373.

Albar, L., Lorieux, M., Ahmadi, N., Rimbault, I., Pinel, A., Sy, A. A., Fargette, D. and Ghesquiere, A. (1998) Genetic basis and mapping of the resistance to rice yellow mottle virus. I. QTLs identification and relationship between resistance and plant morphology. *Theor Appl Genet* **67**: 1145-1154.

Ambros, V. and Baltimore, D. (1978) Protein is linked to the 5' end of poliovirus RNA by a phosphodiester linkage to tyrosine. *J Biol Chem* **253**(15): 5263-6.

Attere, A. F. and Fatokun, C. A. (1983) Reaction of Oryza glaberrima accessions to rice yellow mottle virus. *Plant Dis* **67**: 420-421.

Balke, I., Resevica, G. and Zeltins, A. (2007) The ryegrass mottle virus genome codes for a sobemovirus 3C-like serine protease and RNA-dependent RNA polymerase translated via -1 ribosomal frameshifting. *Virus Genes* **35**(2): 395-8.

Baranov, P. V., Henderson, C. M., Anderson, C. B., Gesteland, R. F., Atkins, J. F. and Howard, M. T. (2005) Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology* **332**(2): 498-510.

Barry, J. K. and Miller, W. A. (2002) A -1 ribosomal frameshift element that requires base pairing across four kilobases suggests a mechanism of regulating ribosome and replicase traffic on a viral RNA. *Proc Natl Acad Sci U S A* **99**(17): 11133-8.

Bates, H. J., Farjah, M., Osman, T. A. M. and Buck, K. W. (1995) Isolation and characterization of an RNA-dependent RNA polymerase from Nicotiana clevelandii plants infected with red clover necrotic mosaic dianthovirus. *J Gen Virol* **76**: 1483-1491.

Bhuvaneshwari, M., Subramanya, H. S., Gopinath, K., Savithri, H. S., Nayudu, M. V. and Murthy, M. R. (1995) Structure of sesbania mosaic virus at 3 A resolution. *Structure* **3**(10): 1021-30.

Bonneau, C., Brugidou, C., Chen, L., Beachy, R. N. and Fauquet, C. (1998) Expression of the rice yellow mottle virus P1 protein in vitro and in vivo and its involvement in virus spread. *Virology* **244**(1): 79-86.

Brierley, I. (1995) Ribosomal frameshifting viral RNAs. *J Gen Virol* **76 (Pt 8)**: 1885-92.

Brierley, I. and Dos Ramos, F. J. (2006) Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. *Virus Res* **119**(1): 29-42.

Brierley, I., Jenner, A. J. and Inglis, S. C. (1992) Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *J Mol Biol* **227**(2): 463-79.

Brill, L. M., Nunn, R. S., Kahn, T. W., Yeager, M. and Beachy, R. N. (2000) Recombinant tobacco mosaic virus movement protein is an RNA-binding, alpha-helical membrane protein. *Proc Natl Acad Sci U S A* **97**(13): 7112-7.

Brugidou, C., Holt, C., Yassi, M. N., Zhang, S., Beachy, R. and Fauquet, C. (1995) Synthesis of an infectious full-length cDNA clone of rice yellow mottle virus and mutagenesis of the coat protein. *Virology* **206**(1): 108-15.

Brugidou, C., Opalka, N., Yeager, M., Beachy, R. N. and Fauquet, C. (2002) Stability of rice yellow mottle virus and cellular compartmentalization during the infection process in Oryza sativa (L.). *Virology* **297**(1): 98-108.

Buck, K. W. (1996) Comparison of the replication of positive-stranded RNA viruses of plants and animals. *Adv Virus Res* **47**: 159-251.

Callaway, A., S., Thornesbury, J. and Lommel, S., A. (2002). GenBank accession no. AY177608.

Callaway, A. S., George, C. G. and Lommel, S. A. (2004) A Sobemovirus coat protein gene complements long-distance movement of a coat protein-null Dianthovirus. *Virology* **330**: 186-195.

Catherall, P. L. (1987) Resistances of grasses to two sobemoviruses, cocksfoot mottle and cynosurus mottle. *Grass Forage Sci* **40**: 311-316.

Citovsky, V., Knorr, D., Schuster, G. and Zambryski, P. (1990) The P30 movement protein of tobacco mosaic virus is a single-strand nucleic acid binding protein. *Cell* **60**(4): 637-47.

Collins, R. F., Gellatly, D. L., Sehgal, O. P. and Abouhaidar, M. G. (1998) Self-cleaving circular RNA associated with rice yellow mottle virus is the smallest viroid-like RNA. *Virology* **241**(2): 269-75.

Daughenbaugh, K. F., Fraser, C. S., Hershey, J. W. and Hardy, M. E. (2003) The genome-linked protein VPg of the Norwalk virus binds eIF3, suggesting its role in translation initiation complex recruitment. *EMBO J* **22**(11): 2852-9.

Davies, C., Haseloff, J. and Symons, R. H. (1990) Structure, self-cleavage, and replication of two viroid-like satellite RNAs (virusoids) of subterranean clover mottle virus. *Virology* **177**(1): 216-24.

Demler, S. A. and de Zoeten, G. A. (1991) The nucleotide sequence and luteovirus-like nature of RNA 1 of an aphid non-transmissible strain of pea enation mosaic virus. *J Gen Virol* **72 (Pt 8)**: 1819-34.

Deom, C. M., Oliver, M. J. and Beachy, R. N. (1987) The 30-Kilodalton Gene Product of Tobacco Mosaic Virus Potentiates Virus Movement. *Science* **237**(4813): 389-394.

Dessens, J. T. and Lomonossoff, G. P. (1992) Sequence upstream of the 24K protease enhances cleavage of the cowpea mosaic virus B RNA-encoded polyprotein at the junction between the 24K and 87K proteins. *Virology* **189**(1): 225-32.

Dingwall, C. and Laskey, R. A. (1991) Nuclear targeting sequences--a consensus? *Trends Biochem Sci* **16**(12): 478-81.

Dinman, J. D., Richter, S., Plant, E. P., Taylor, R. C., Hammell, A. B. and Rana, T. M. (2002) The frameshift signal of HIV-1 involves a potential intramolecular triplex RNA structure. *Proc Natl Acad Sci U S A* **99**(8): 5331-6.

Dinman, J. D., Ruiz-Echevarria, M. J., Czaplinski, K. and Peltz, S. W. (1997) Peptidyl-transferase inhibitors have antiviral properties by altering programmed -1 ribosomal frameshifting efficiencies: development of model systems. *Proc Natl Acad Sci U S A* **94**(13): 6606-11.

Dinman, J. D. and Wickner, R. B. (1992) Ribosomal frameshifting efficiency and gag/gag-pol ratio are critical for yeast M1 double-stranded RNA virus propagation. *J Virol* **66**(6): 3669-76.

Dolja, V. V. and Koonin, E. V. (1991) Phylogeny of capsid proteins of small icosahedral RNA plant viruses. *J Gen Virol* **72 (Pt 7)**: 1481-6.

Dougherty, W. G. and Semler, B. L. (1993) Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. *Microbiol Rev* **57**(4): 781-822.

Dulude, D., Baril, M. and Brakier-Gingras, L. (2002) Characterization of the frameshift stimulatory signal controlling a programmed -1 ribosomal frameshift in the human immunodeficiency virus type 1. *Nucleic Acids Res* **30**(23): 5094-102.

Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M. and Obradovic, Z. (2002) Intrinsic disorder and protein function. *Biochemistry* **41**(21): 6573-82.

Dwyer, G. I., Njeru, R., Williamson, S., Fosu-Nyarko, J., Hopkins, R., Jones, R. A., Waterhouse, P. M. and Jones, M. G. (2003) The complete nucleotide sequence of Subterranean clover mottle virus. *Arch Virol* **148**(11): 2237-47.

Dyson, H. J. and Wright, P. E. (2002) Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* **12**(1): 54-60.

Epel, B. L. (2009) Plant viruses spread by diffusion on ER-associated movement-protein-rafts through plasmodesmata gated by viral induced host beta-1,3-glucanases. *Semin Cell Dev Biol* **20**(9): 1074-81.

Erickson, J. W. and Rossmann, M. G. (1982) Assembly and crystallization of a T = 1 icosahedral particle from trypsinized southern bean mosaic virus coat protein. *Virology* **116**(1): 128-36.

Falk, H., Mador, N., Udi, R., Panet, A. and Honigman, A. (1993) Two cis-acting signals control ribosomal frameshift between human T-cell leukemia virus type II gag and pro genes. *J Virol* **67**(10): 6273-7.

Farabaugh, P. J. (1996) Programmed translational frameshifting. *Microbiol Rev* **60**(1): 103-34.

Fargette, D., Pinel, A., Abubakar, Z., Traore, O., Brugidou, C., Fatogoma, S., Hebrard, E., Choisy, M., Sere, Y., Fauquet, C. and Konate, G. (2004) Inferring the evolutionary history of rice yellow mottle virus from genomic, phylogenetic, and phylogeographic studies. *J Virol* **78**(7): 3252-61.

Ferrer-Orta, C., Arias, A., Escarmis, C. and Verdaguer, N. (2006) A comparison of viral RNA-dependent RNA polymerases. *Curr Opin Struct Biol* **16**(1): 27-34.

Francki, R. I. B., Milne, R. G. and Hatta, T., Eds. (1985). Atlas of plant viruses, CRC Press, Boca Raton, Fla.

Francki, R. I. B., Randles, J. W., Hatta, T., Davies, C., Chu, P. W. G. and McLean, G. D. (1983) Subterranean clover mottle virus: another virus from Australia with encapsidated viroid-like RNA. . *Plant Pathol* **32**: 47-59.

Fuentes, A. L. and Hamilton, R. I. (1993) Failure of long-distance movement of southern bean mosaic virus in a resistant host is correlated with lack of normal virion formation. *J Gen Virol* **74 (Pt 9)**: 1903-10.

Fujiki, M., Kawakami, S., Kim, R. W. and Beachy, R. N. (2006) Domains of tobacco mosaic virus movement protein essential for its membrane association. *J Gen Virol* **87**(Pt 9): 2699-707.

Gayathri, P., Satheshkumar, P. S., Prasad, K., Nair, S., Savithri, H. S. and Murthy, M. R. (2006) Crystal structure of the serine protease domain of Sesbania mosaic virus polyprotein and mutational analysis of residues forming the S1-binding pocket. *Virology* **346**(2): 440-51.

Gesteland, R. F. and Atkins, J. F. (1996) Recoding: dynamic reprogramming of translation. *Annu Rev Biochem* **65**: 741-68.

Ghesquiere, A., Albar, L., Lorieux, M., Ahmadi, N., Fargette, D., Huang, N., McCouch, S. R. and Notteghem, J. L. (1997) A major quantitative trait locus for rice yellow mottle virus resistance maps to a cluster of blast resistance genes on chromosome 12. *Phytopathology* **87**(12): 1243-9.

Giedroc, D. P., Theimer, C. A. and Nixon, P. L. (2000) Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting. *J Mol Biol* **298**(2): 167-85.

Gorbalenya, A. E., Koonin, E. V., Blinov, V. M. and Donchenko, A. P. (1988) Sobemovirus genome appears to encode a serine protease related to cysteine proteases of picornaviruses. *FEBS Lett* **236**(2): 287-90.

Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. and Blinov, V. M. (1989) Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res* **17**(12): 4713-30.

Govind, K. and Savithri, H. S. (2010) Primer-independent initiation of RNA synthesis by SeMV recombinant RNA-dependent RNA polymerase. *Virology* **401**(2): 280-92.

Guilley, H., Richards, K. E. and Jonard, G. (1995) Nucleotide sequence of beet mild yellowing virus RNA. *Arch Virol* **140**(6): 1109-18.

Guilley, H., Wipf-Scheibel, C., Richards, K., Lecoq, H. and Jonard, G. (1994) Nucleotide sequence of cucurbit aphid-borne yellows luteovirus. *Virology* **202**(2): 1012-7.

Hacker, D. L. (1995) Identification of a coat protein binding site on southern bean mosaic virus RNA. *Virology* **207**(2): 562-5.

Hacker, D. L. and Sivakumaran, K. (1997) Mapping and expression of southern bean mosaic virus genomic and subgenomic RNAs. *Virology* **234**(2): 317-27.

Haseloff, J. and Symons, R. H. (1982) Comparative sequence and structure of viroid-like RNAs of two plant viruses. *Nucleic Acids Res* **10**(12): 3681-91.

Hebrard, E., Bessin, Y., Michon, T., Longhi, S., Uversky, V. N., Delalande, F., Van Dorsselaer, A., Romero, P., Walter, J., Declerck, N. and Fargette, D. (2009) Intrinsic disorder in Viral Proteins Genome-Linked: experimental and predictive analyses. *Virol J* **6**: 23.

Hebrard, E., Pinel-Galzi, A., Bersoult, A., Sire, C. and Fargette, D. (2006) Emergence of a resistance-breaking isolate of Rice yellow mottle virus during serial inoculations is due to a single substitution in the genome-linked viral protein VPg. *J Gen Virol* **87**(Pt 5): 1369-73.

Heinlein, M., Padgett, H. S., Gens, J. S., Pickard, B. G., Casper, S. J., Epel, B. L. and Beachy, R. N. (1998) Changing patterns of localization of the tobacco mosaic virus movement protein and replicase to the endoplasmic reticulum and microtubules during infection. *Plant Cell* **10**(7): 1107-20.

Hermodson, M. A., Abad-Zapatero, C., Abdel-Meguid, S. S., Pundak, S., Rossmann, M. G. and Tremaine, J. H. (1982) Amino acid sequence of southern bean mosaic virus coat protein and its relation to the three-dimensional structure of the virus. *Virology* **119**(1): 133-49.

Hobbs, H. A., Kuhn, C. W., Papa, K. E. and Brantley, B. B. (1987) Inheritance of nonnecrotic resistance to southern bean mosaic virus in cowpea. *Phytopathology* **77**: 1624-1629.

Honda, A. and Nishimura, S. (1996) Suppression of translation frameshift by upstream termination codon. *Biochem Biophys Res Commun* **221**(3): 602-8.

Hsu, C. H., White, J. A. and Sehgal, O. P. (1977) Assembly of southern bean mosaic virus from its two subviral intermediates. *Virology* **81**(2): 471-5.

Huang, T. S., Wei, T., Laliberte, J. F. and Wang, A. (2010) A host RNA helicase-like protein, AtRH8, interacts with the potyviral genome-linked protein, VPg, associates with the virus accumulation complex, and is essential for infection. *Plant Physiol* **152**(1): 255-66.

Hull, R. (1977a) The grouping of small spherical plant viruses with single RNS components. *J Gen Virol* **36**: 289-295.

Hull, R. (1977b) The stabilization of the particles of turnip rosette virus and of other members of the southern bean mosaic virus group. *Virology* **79**(1): 58-66.

Hull, R., Ed. (1995). Virus taxonomy. Classification and nomenclature of viruses. Sixth report of the International Committee on Taxonomy of Viruses. Sobemovirus. Vienna, Springer-Verlag.

Hung, M., Patel, P., Davis, S. and Green, S. R. (1998) Importance of ribosomal frameshifting for human immunodeficiency virus type 1 particle assembly and replication. *J Virol* **72**(6): 4819-24.

Jacks, T., Madhani, H. D., Masiarz, F. R. and Varmus, H. E. (1988) Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell* **55**(3): 447-58.

Jaegle, M., Wellink, J. and Goldbach, R. (1987) The genomelinked protein of cowpea mosaic virus is bound to the 5' terminus of virus RNA by a phosphodiester linkage to serine. *J Gen Viro* **68**: 627-632.

Jeffries, A. C., Rathjen, J. P. and Symons, R. H. (1995). GenBank accession no. U31286.

Jones, A. T. and Mayo, M. A. (1983) Interaction of lucerne transient streak virus and the viroid-like RNA-2 of Solanum nodiflorum mottle virus. . *J Gen Virol* **64**.

Jones, A. T. and Mayo, M. A. (1984) Satellite nature of the viroid-like RNA-2 of Solanum nodiflorum mottle virus and the ability of other plant viruses to support the replication of viroid-like RNA molecules. *j gen Virol* **65**: 1713-1721.

Jones, R. A. (2004) Using epidemiological information to develop effective integrated virus disease management strategies. *Virus Res* **100**(1): 5-30.

Kadare, G. and Haenni, A. L. (1997) Virus-encoded RNA helicases. *J Virol* **71**(4): 2583-90.

Kalinina, N. O., Rakitina, D. V., Solovyev, A. G., Schiemann, J. and Morozov, S. Y. (2002) RNA helicase activity of the plant virus movement proteins encoded by the first gene of the triple gene block. *Virology* **296**(2): 321-9.

Keese, P., Bruening, G. and Symons, R. H. (1983) Comparative sequence and structure of circular RNAs from two isolates of lucerne transient streak virus. *FEBS Lett* **159**: 185-190.

Keller, K. E., Johansen, I. E., Martin, R. R. and Hampton, R. O. (1998) Potyvirus genome-linked protein (VPg) determines pea seed-borne mosaic virus pathotype-specific virulence in Pisum sativum. *Mol Plant Microbe Interact* **11**(2): 124-30.

Kim, K. H. and Lommel, S. A. (1994) Identification and analysis of the site of -1 ribosomal frameshifting in red clover necrotic mosaic virus. *Virology* **200**(2): 574-82.

Kim, K. H. and Lommel, S. A. (1998) Sequence element required for efficient -1 ribosomal frameshifting in red clover necrotic mosaic dianthovirus. *Virology* **250**(1): 50-9.

Kim, S. H., Macfarlane, S., Kalinina, N. O., Rakitina, D. V., Ryabov, E. V., Gillespie, T., Haupt, S., Brown, J. W. and Taliansky, M. (2007a) Interaction of a plant virus-encoded protein with the major

nucleolar protein fibrillarin is required for systemic virus infection. *Proc Natl Acad Sci U S A* **104**(26): 11115-20.

Kim, S. H., Ryabov, E. V., Kalinina, N. O., Rakitina, D. V., Gillespie, T., MacFarlane, S., Haupt, S., Brown, J. W. and Taliansky, M. (2007b) Cajal bodies and the nucleolus are required for a plant virus systemic infection. *EMBO J* **26**(8): 2169-79.

Kim, Y. G., Maas, S. and Rich, A. (2001) Comparative mutational analysis of cis-acting RNA signals for translational frameshifting in HIV-1 and HTLV-2. *Nucleic Acids Res* **29**(5): 1125-31.

Kontos, H., Napthine, S. and Brierley, I. (2001) Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency. *Mol Cell Biol* **21**(24): 8657-70.

Koonin, E. V. (1991a) The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J Gen Virol* **72 (Pt 9)**: 2197-206.

Koonin, E. V. (1991b) Similarities in RNA helicases. *Nature* **352**(6333): 290.

Koonin, E. V. and Dolja, V. V. (1993) Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit Rev Biochem Mol Biol* **28**(5): 375-430.

Lacombe, S., Bangratz, M., Vignols, F. and Brugidou, C. (2010) The rice yellow mottle virus P1 protein exhibits dual functions to suppress and activate gene silencing. *Plant J* **61**(3): 371-82.

Lee, L. and Anderson, E. J. (1998) Nucleotide sequence of a resistance breaking mutant of southern bean mosaic virus. *Arch Virol* **143**(11): 2189-201.

Lee, S. K., Dabney-Smith, C., Hacker, D. L. and Bruce, B. D. (2001) Membrane activity of the southern cowpea mosaic virus coat protein: the role of basic amino acids, helix-forming potential, and lipid composition. *Virology* **291**(2): 299-310.

Lee, S. K. and Hacker, D. L. (2001) In vitro analysis of an RNA binding site within the N-terminal 30 amino acids of the southern cowpea mosaic virus coat protein. *Virology* **286**(2): 317-27.

Leonard, S., Plante, D., Wittmann, S., Daigneault, N., Fortin, M. G. and Laliberte, J. F. (2000) Complex formation between potyvirus VPg and translation eukaryotic initiation factor 4E correlates with virus infectivity. *J Virol* **74**(17): 7730-7.

Li, L., Wang, A. L. and Wang, C. C. (2001) Structural analysis of the -1 ribosomal frameshift elements in giardiavirus mRNA. *J Virol* **75**(22): 10612-22.

Lokesh, G. L., Gopinath, K., Satheshkumar, P. S. and Savithri, H. S. (2001) Complete nucleotide sequence of Sesbania mosaic virus: a new virus species of the genus Sobemovirus. *Arch Virol* **146**(2): 209-23.

Lokesh, G. L., Gopinath, K., Satheshkumar, P. S. and Savithri, H. S. (2006). GenBank accession no. .

Lokesh, G. L., Gowri, T. D., Satheshkumar, P. S., Murthy, M. R. and Savithri, H. S. (2002) A molecular switch in the capsid protein controls the particle polymorphism in an icosahedral virus. *Virology* **292**(2): 211-23.

Lucas, W. J. (2006) Plant viral movement proteins: agents for cell-to-cell trafficking of viral genomes. *Virology* **344**(1): 169-84.

Lucchesi, J., Makelainen, K., Merits, A., Tamm, T. and Makinen, K. (2000) Regulation of -1 ribosomal frameshifting directed by cocksfoot mottle sobemovirus genome. *Eur J Biochem* **267**(12): 3523-9.

Luft, J. R., Collins, R. J., Fehrman, N. A., Lauricella, A. M., Veatch, C. K. and DeTitta, G. T. (2003) A deliberate approach to screening for initial crystallization conditions of biological macromolecules. *J Struct Biol* **142**(1): 170-9.

Mäkeläinen, K. (2006). Lost in Translation: Translation Mechanisms in Production of Cocksfoot Mottle Virus Proteins. Helsinki, University of Helsinki.

Makelainen, K. and Makinen, K. (2005) Factors affecting translation at the programmed -1 ribosomal frameshifting site of Cocksfoot mottle virus RNA in vivo. *Nucleic Acids Res* **33**(7): 2239-47.

Mäkinen, K., Generozov, E., Arshava, N., Kaloshin, A., Morozov, S. and Zavriev, S. (2000) Detection and characterization of defective interfering RNAs associated with Cocksfoot mottle sobemovirus. *Mol. Biol. (Moscow)* **34**: 291-296.

Makinen, K., Makelainen, K., Arshava, N., Tamm, T., Merits, A., Truve, E., Zavriev, S. and Saarma, M. (2000) Characterization of VPg and the polyprotein processing of cocksfoot mottle virus (genus Sobemovirus). *J Gen Virol* **81**(Pt 11): 2783-9.

Makinen, K., Tamm, T., Naess, V., Truve, E., Puurand, U., Munthe, T. and Saarma, M. (1995) Characterization of cocksfoot mottle sobemovirus genomic RNA and sequence comparison with related viruses. *J Gen Virol* **76 (Pt 11)**: 2817-25.

Mandahar, C. L., Ed. (2006). Multiplication of RNA Plant Viruses. Dordrecht, Springer.

Mang, K. Q., Ghosh, A. and Kaesberg, P. (1982) A comparative study of the cowpea and bean strains of southern bean mosaic virus. *Virology* **116**(1): 264-74.

Marcheschi, R. J., Staple, D. W. and Butcher, S. E. (2007) Programmed ribosomal frameshifting in SIV is induced by a highly structured RNA stem-loop. *J Mol Biol* **373**(3): 652-63.

Marczinke, B., Bloys, A. J., Brown, T. D., Willcocks, M. M., Carter, M. J. and Brierley, I. (1994) The human astrovirus RNA-dependent RNA polymerase coding region is expressed by ribosomal frameshifting. *J Virol* **68**(9): 5588-95.

Matthews, R. E. F. (1982) Classification and nomenclature of viruses. Fourth report of the International Committee on Taxonomy of Viruses. . *Intervirology* **17**: 1-199.

Matthews, R. E. F., Ed. (1991). Plant Virology. Sobemovirus group. San Diego, Academic Press.

McGavin, W. J. and Macfarlane, S. A. (2009) Rubus chlorotic mottle virus, a new sobemovirus infecting raspberry and bramble. *Virus Res* **139**(1): 10-3.

Meier, M., Paves, H., Olspert, A., Tamm, T. and Truve, E. (2006) P1 protein of Cocksfoot mottle virus is indispensable for the systemic spread of the virus. *Virus Genes* **32**(3): 321-6.

Meier, M. and Truve, E. (2007) Sobemoviruses possess a common CfMV-like genomic organization. *Arch Virol* **152**(3): 635-40.

Meshi, T., Watanabe, Y., Saito, T., Sugimoto, A., Maeda, T. and Okada, Y. (1987) Function of the 30 kd protein of tobacco mosaic virus: involvement in cell-to-cell movement and dispensability for replication. *EMBO J* **6**(9): 2557-2563.

Miller, W. A., Dinesh-Kumar, S. P. and Paul, C. P. (1995) Luteovirus gene expression. *Crit Rev Plant Sci* **14**: 179-211.

Miyoshi, H., Suehiro, N., Tomoo, K., Muto, S., Takahashi, T., Tsukamoto, T., Ohmori, T. and Natsuaki, T. (2006) Binding analyses for the interaction between plant virus genome-linked protein (VPg) and plant translational initiation factors. *Biochimie* **88**(3-4): 329-40.

Mohamed, N. and Mossop, D. (1981) Cynosurus and cocksfoot mottle viruses: a comparison. . *j gen Virol* **55**: 63-74.

Morasco, B. J., Sharma, N., Parilla, J. and Flanegan, J. B. (2003) Poliovirus cre(2C)-dependent synthesis of VPgpUpU is required for positive- but not negative-strand RNA synthesis. *J Virol* **77**(9): 5136-44.

Moury, B., Morel, C., Johansen, E., Guilbaud, L., Souche, S., Ayme, V., Caranta, C., Palloix, A. and Jacquemond, M. (2004) Mutations in potato virus Y genome-linked protein determine virulence toward recessive resistances in Capsicum annuum and Lycopersicon hirsutum. *Mol Plant Microbe Interact* **17**(3): 322-9.

Murphy, J. F., Rychlik, W., Rhoads, R. E., Hunt, A. G. and Shaw, J. G. (1991) A tyrosine residue in the small nuclear inclusion protein of tobacco vein mottling virus links the VPg to the viral RNA. *J Virol* **65**(1): 511-3.

Murray, K. E. and Barton, D. J. (2003) Poliovirus CRE-dependent VPg uridylylation is required for positive-strand RNA synthesis but not for negative-strand RNA synthesis. *J Virol* **77**(8): 4739-50.

Nair, S., Gayathri, P., Murthy, M. R. and Savithri, H. S. (2008) Stacking interactions of W271 and H275 of SeMV serine protease with W43 of natively unfolded VPg confer catalytic activity to protease. *Virology* **382**(1): 83-90.

Nair, S. and Savithri, H. S. (2010a) Natively unfolded nucleic acid binding P8 domain of SeMV polyprotein 2a affects the novel ATPase activity of the preceding P10 domain. *FEBS Lett* **584**(3): 571-6.

Nair, S. and Savithri, H. S. (2010b) Processing of SeMV polyproteins revisited. *Virology* **396**(1): 106-17.

Ndjiondjop, M. N., Albar, L., Fargette, D., Fauquet, C. and Ghesquière, A. (1999) The genetic basis of high resistance to rice yellow mottle virus (RYMV) in cultivars of two cultivated rice species. *Plant Dis* **83**: 931-935.

Ngon, M., Yassi, A., Ritzenthaler, C., Brugidou, C., Fauquet, C. and Beachy, R. N. (1994) Nucleotide sequence and genome characterization of rice yellow mottle virus RNA. *j gen Virol* **75**: 249-257.

Nicolas, O., Dunnington, S. W., Gotow, L. F., Pirone, T. P. and Hellmann, G. M. (1997) Variations in the VPg protein allow a potyvirus to overcome va gene resistance in tobacco. *Virology* **237**(2): 452-9.

Oda, Y., Saeki, K., Takahashi, Y., Maeda, T., Naitow, H., Tsukihara, T. and Fukuyama, K. (2000) Crystal structure of tobacco necrosis virus at 2.25 A resolution. *J Mol Biol* **300**(1): 153-69.

Olspert, A., Paves, H., Toomela, R., Tamm, T. and Truve, E. (2010) Cocksfoot mottle sobemovirus coat protein contains two nuclear localization signals. *Virus Genes* **40**(3): 423-31.

Opalka, N., Brugidou, C., Bonneau, C., Nicole, M., Beachy, R. N., Yeager, M. and Fauquet, C. (1998) Movement of rice yellow mottle virus between xylem cells through pit membranes. *Proc Natl Acad Sci U S A* **95**(6): 3323-8.

Othman, Y. and Hull, R. (1995) Nucleotide sequence of the bean strain of southern bean mosaic virus. *Virology* **206**(1): 287-97.

Pantazatos, D., Kim, J. S., Klock, H. E., Stevens, R. C., Wilson, I. A., Lesley, S. A. and Woods, V. L., Jr. (2004) Rapid refinement of crystallographic protein construct definition employing enhanced hydrogen/deuterium exchange MS. *Proc Natl Acad Sci U S A* **101**(3): 751-6.

Pappachan, A., Chinnathambi, S., Satheshkumar, P. S., Savithri, H. S. and Murthy, M. R. (2009) A single point mutation disrupts the capsid assembly in Sesbania Mosaic Virus resulting in a stable isolated dimer. *Virology* **392**(2): 215-21.

Pappachan, A., Subashchandrabose, C., Satheshkumar, P. S., Savithri, H. S. and Murthy, M. R. (2008) Structure of recombinant capsids formed by the beta-annulus deletion mutant -- rCP (Delta48-59) of Sesbania mosaic virus. *Virology* **375**(1): 190-6.

Paul, C. N., Ng, N. Q. and Ladeinde, T. (1995) Diallel analysis of resistance to rice yellow mottle virus in African rice Oryza glaberrima Steud. *J Genet Breed* **49**: 217-222.

Paul, C. P., Barry, J. K., Dinesh-Kumar, S. P., Brault, V. and Miller, W. A. (2001) A sequence required for -1 ribosomal frameshifting located four kilobases downstream of the frameshift site. *J Mol Biol* **310**(5): 987-99.

Pennell, S., Manktelow, E., Flatt, A., Kelly, G., Smerdon, S. J. and Brierley, I. (2008) The stimulatory RNA of the Visna-Maedi retrovirus ribosomal frameshifting signal is an unusual pseudoknot with an interstem element. *RNA* **14**(7): 1366-77.

Pinto, Y. M. and Baulcombe, D. C. (1995). GenBank accession no. U23142.

Pinto, Y. M., Kok, R. A. and Baulcombe, D. C. (1999) Resistance to rice yellow mottle virus (RYMV) in cultivated African rice varieties containing RYMV transgenes. *Nat Biotechnol* **17**: 702-707.

Plant, E. P., Jacobs, K. L., Harger, J. W., Meskauskas, A., Jacobs, J. L., Baxter, J. L., Petrov, A. N. and Dinman, J. D. (2003) The 9-A solution: how mRNA pseudoknots promote efficient programmed -1 ribosomal frameshifting. *RNA* **9**(2): 168-74.

Plant, E. P., Perez-Alvarado, G. C., Jacobs, J. L., Mukhopadhyay, B., Hennig, M. and Dinman, J. D. (2005) A three-stemmed mRNA pseudoknot in the SARS coronavirus frameshift signal. *PLoS Biol* **3**(6): e172.

Plevka, P., Tars, K., Zeltins, A., Balke, I., Truve, E. and Liljas, L. (2007) The three-dimensional structure of ryegrass mottle virus at 2.9 A resolution. *Virology* **369**(2): 364-74.

Pressoir, G., L. Albar, N. Ahmadi, I. Rimbault, M. Lorieux, D. Fargette, and A. Ghesquiere (1998) Genetic basis and mapping of the resistance to rice yellow mottle virus. II. Evidence of a complementary epistasis between two QTLs. *Theor Appl Genet* **97**: 1155-1161.

Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I. and Sussman, J. L. (2005) FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**(16): 3435-8.

Prufer, D., Kawchuk, L., Monecke, M., Nowok, S., Fischer, R. and Rohde, W. (1999) Immunological analysis of potato leafroll luteovirus (PLRV) P1 expression identifies a 25 kDa RNA-binding protein derived via P1 processing. *Nucleic Acids Res* **27**(2): 421-5.

Qu, C., Liljas, L., Opalka, N., Brugidou, C., Yeager, M., Beachy, R. N., Fauquet, C. M., Johnson, J. E. and Lin, T. (2000) 3D domain swapping modulates the stability of members of an icosahedral virus group. *Structure* **8**(10): 1095-103.

Rajamaki, M. L. and Valkonen, J. P. (2009) Control of nuclear and nucleolar localization of nuclear inclusion protein a of picorna-like Potato virus A in Nicotiana species. *Plant Cell* **21**(8): 2485-502.

Reichel, C. and Beachy, R. N. (1998) Tobacco mosaic virus infection induces severe morphological changes of the endoplasmic reticulum. *Proc Natl Acad Sci U S A* **95**(19): 11169-74.

Reichert, V. L., Choi, M., Petrillo, J. E. and Gehrke, L. (2007) Alfalfa mosaic virus coat protein bridges RNA and RNA-dependent RNA polymerase in vitro. *Virology* **364**(1): 214-26.

Revill, P. A., Davidson, A. D. and Wright, P. J. (1994) The nucleotide sequence and genome organization of mushroom bacilliform virus: a single-stranded RNA virus of Agaricus bisporus (Lange) Imbach. *Virology* **202**(2): 904-11.

Revill, P. A., Davidson, A. D. and Wright, P. J. (1998) Mushroom bacilliform virus RNA: the initiation of translation at the 5' end of the genome and identification of the VPg. *Virology* **249**(2): 231-7.

Rognli, O. A., Aastveit, K. and Munthe, T. (1995) Genetic variation in cocksfoot (Dactylis glomerata L.) populations for mottle virus resistance. *Euphytica* **83**: 109-116.

Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E. and Dunker, A. K. (1997) Identifying Disordered Regions in Proteins from Amino Acid Sequences. Proc. I.E.E.E. International Conference on Neural Networks, p. 90-95.

Rossmann, M. G. (1984) Constraints on the assembly of spherical virus particles. *Virology* **134**(1): 1-11.

Rossmann, M. G., Abad-Zapatero, C., Erickson, J. W. and Savithri, H. S. (1983a) RNA-protein interactions in some small plant viruses. *J Biomol Struct Dyn* **1**(2): 565-79.

Rossmann, M. G., Abad-Zapatero, C., Hermodson, M. A. and Erickson, J. W. (1983b) Subunit interactions in southern bean mosaic virus. *J Mol Biol* **166**(1): 37-73.

Rothberg, P. G., Harris, T. J., Nomoto, A. and Wimmer, E. (1978) O4-(5'-uridylyl)tyrosine is the bond between the genome-linked protein and the RNA of poliovirus. *Proc Natl Acad Sci U S A* **75**(10): 4868-72.

Ryabov, E. V., Krutov, A. A., Novikov, V. K., Zheleznikova, O. V., Morozov, S. Y. and Zavriev, S. K. (1996) Nucleotide sequence of RNA from the sobemovirus found in infected cocksfoot shows a Luteovirus-like arrangement of the putative replicase and protease genes. . *Phytopathology* **86**: 391-397.

Sarmiento, C., Gomez, E., Meier, M., Kavanagh, T. A. and Truve, E. (2007) Cocksfoot mottle virus P1 suppresses RNA silencing in Nicotiana benthamiana and Nicotiana tabacum. *Virus Res* **123**(1): 95-9.

Sarra, S. (2005). Novel insights in the transmission of Rice yellow mottle virus in irrigated rice., Wageningen University.

Satheshkumar, P. S., Gayathri, P., Prasad, K. and Savithri, H. S. (2005a) "Natively unfolded" VPg is essential for Sesbania mosaic virus serine protease activity. *J Biol Chem* **280**(34): 30291-300.

Satheshkumar, P. S., Lokesh, G. L., Murthy, M. R. and Savithri, H. S. (2005b) The role of arginine-rich motif and beta-annulus in the assembly and stability of Sesbania mosaic virus capsids. *J Mol Biol* **353**(2): 447-58.

Satheshkumar, P. S., Lokesh, G. L. and Savithri, H. S. (2004) Polyprotein processing: cis and trans proteolytic activities of Sesbania mosaic virus serine protease. *Virology* **318**(1): 429-38.

Savithri, H. S. and Erickson, J. W. (1983) The self-assembly of the cowpea strain of southern bean mosaic virus: formation of T = 1 and T = 3 nucleoprotein particles. *Virology* **126**(1): 328-35.

Schaad, M. C., Anderberg, R. J. and Carrington, J. C. (2000) Strain-specific interaction of the tobacco etch virus NIa protein with the translation initiation factor eIF4E in the yeast two-hybrid system. *Virology* **273**(2): 300-6.

Schein, C. H., Oezguen, N., van der Heden van Noort, G. J., Filippov, D. V., Paul, A., Kumar, E. and Braun, W. (2010) NMR solution structure of poliovirus uridylyated peptide linked to the genome (VPgpU). *Peptides*.

Segundo, E., Gil-Salas, F. M., Janssen, D., Martin, G., Cuadrado, I. M. and Remah, A. (2004) First report of Southern bean mosaic virus infecting French bean in Morocco. . *Plant Dis* **88**: 1162-1162.

Sehgal, O. P., Hsu C. H., White, J. A. and Van, M. (1979) Enzymic sensitivity of conformationally altered virions of southern bean mosaic virus. *Phytopathology* **95**: 167-177.

Sehgal, O. P., Sinha, R. C., Gellatly, D. L., Ivanov, I. and AbouHaidar, M. G. (1993) Replication and encapsidation of the viroid-like satellite RNA of lucerne transient streak virus are supported in divergent hosts by cocksfoot mottle virus and turnip rosette virus. *J Gen Virol* **74 (Pt 4)**: 785-8.

Sereme, D., Lacombe, S., Konate, M., Pinel-Galzi, A., Traore, V. S., Hebrard, E., Traore, O., Brugidou, C., Fargette, D. and Konate, G. (2008) Biological and molecular characterization of a putative new sobemovirus infecting Imperata cylindrica and maize in Africa. *Arch Virol* **153**(10): 1813-20.

Silva, A. M. and Rossmann, M. G. (1987) Refined structure of southern bean mosaic virus at 2.9 A resolution. *J Mol Biol* **197**(1): 69-87.

Sire, C., Bangratz-Reyser, M., Fargette, D. and Brugidou, C. (2008) Genetic diversity and silencing suppression effects of Rice yellow mottle virus and the P1 protein. *Virol J* **5**: 55.

Sivakumaran, K., Fowler, B. C. and Hacker, D. L. (1998) Identification of viral genes required for cell-to-cell movement of southern bean mosaic virus. *Virology* **252**(2): 376-86.

Snijder, E. J., Wassenaar, A. L., van Dinten, L. C., Spaan, W. J. and Gorbalenya, A. E. (1996) The arterivirus nsp4 protease is the prototype of a novel group of chymotrypsin-like enzymes, the 3C-like serine proteases. *J Biol Chem* **271**(9): 4864-71.

Sõmera, M. (2010). Sobemoviruses: genomic organization, potential for recombination and necessity of P1 in systemic infection., Tallinn University of Technology.

Staple, D. W. and Butcher, S. E. (2005) Solution structure and thermodynamic investigation of the HIV-1 frameshift inducing element. *J Mol Biol* **349**(5): 1011-23.

Su, M. C., Chang, C. T., Chu, C. H., Tsai, C. H. and Chang, K. Y. (2005) An atypical RNA pseudoknot stimulator and an upstream attenuation signal for -1 ribosomal frameshifting of SARS coronavirus. *Nucleic Acids Res* **33**(13): 4265-75.

Taliansky, M., Torrance, L. and Kalinina, N. O. (2008) Role of plant virus movement proteins. *Methods Mol Biol* **451**: 33-54.

Tamm, T., Makinen, K. and Truve, E. (1999) Identification of genes encoding for the cocksfoot mottle virus proteins. *Arch Virol* **144**(8): 1557-67.

Tamm, T., Suurvali, J., Lucchesi, J., Olspert, A. and Truve, E. (2009) Stem-loop structure of Cocksfoot mottle virus RNA is indispensable for programmed -1 ribosomal frameshifting. *Virus Res* **146**(1-2): 73-80.

Tamm, T. and Truve, E. (2000a) RNA-binding activities of cocksfoot mottle sobemovirus proteins. *Virus Res* **66**(2): 197-207.

Tamm, T. and Truve, E. (2000b) Sobemoviruses. *J Virol* **74**(14): 6231-41.

Tars, K., Zeltins, A. and Liljas, L. (2003) The three-dimensional structure of cocksfoot mottle virus at 2.7 A resolution. *Virology* **310**(2): 287-97.

Thottapilly, G. and Rossel, H. W. (1993) Evaluation of resistance to rice yellow mottle virus in Oryza species. *Indian J Virol* **9**: 65-73.

Tien-Po, Davies, C., Hatta, T. and Francki, R. I. B. (1981) Viroid-like RNA encapsidated in lucerne transient streak virus. *FEBS Lett* **132**: 353-356.

Tomenius, K., Clapham, D. and Meshi, T. (1987) Localization by immunogold cytochemistry of the virus-coded 30K protein in plasmodesmata of leaves infected with tobacco mosaic virus. *Virology* **160**(2): 363-71.

Tu, C., Tzeng, T. H. and Bruenn, J. A. (1992) Ribosomal movement impeded at a pseudoknot required for frameshifting. *Proc Natl Acad Sci U S A* **89**(18): 8636-40.

Uversky, V. N. (2002) What does it mean to be natively unfolded? *Eur J Biochem* **269**(1): 2-12.

Vacic, V., Oldfield, C. J., Mohan, A., Radivojac, P., Cortese, M. S., Uversky, V. N. and Dunker, A. K. (2007) Characterization of molecular recognition features, MoRFs, and their binding partners. *J Proteome Res* **6**(6): 2351-66.

van de Ven, F. J., Lycksell, P. O., van Kammen, A. and Hilbers, C. W. (1990) Computer-aided assignment of the 1H-NMR spectrum of the viral-protein-genome-linked polypeptide from cowpea mosaic virus. *Eur J Biochem* **190**(3): 583-91.

van der Wilk, F., Huisman, M. J., Cornelissen, B. J., Huttinga, H. and Goldbach, R. (1989) Nucleotide sequence and organization of potato leafroll virus genomic RNA. *FEBS Lett* **245**(1-2): 51-6.

van der Wilk, F., Verbeek, M., Dullemans, A. and van den Heuvel, J. (1998) The genome-linked protein (VPg) of southern bean mosaic virus is encoded by the ORF2. *Virus Genes* **17**(1): 21-4.

van Dijk, A. A., Makeyev, E. V. and Bamford, D. H. (2004) Initiation of viral RNA-dependent RNA polymerization. *J Gen Virol* **85**(Pt 5): 1077-93.

Veidt, I., Lot, H., Leiser, M., Scheidecker, D., Guilley, H., Richards, K. and Jonard, G. (1988) Nucleotide sequence of beet western yellows virus RNA. *Nucleic Acids Res* **16**(21): 9917-32.

Verhoeven, J. T. J., W., R. J., Lesemann, D. E., Segundo, E., Velasco, L., Ruiz, L., Janssen, D. and M., C. I. (2003) Southern bean mosaic virus the causal agent of a new disease of Phaseolus vulgaris beans in Spain. *Eur J Plant Pathol* **109**: 935-941.

Vincent, J. R., Lister, R. M. and Larkins, B. A. (1991) Nucleotide sequence analysis and genomic organization of the NY-RPV isolate of barley yellow dwarf virus. *j gen Virol* **72**: 2347-2355.

Voinnet, O., Pinto, Y. M. and Baulcombe, D. C. (1999) Suppression of gene silencing: a general strategy used by diverse DNA and RNA viruses of plants. *Proc Natl Acad Sci U S A* **96**(24): 14147-52.

Walters, H. J. (1969) Beetle transmission of plant viruses. *Adv Virus Res* **15**: 339-63.

Weinheimer, I., Boonrod, K., Moser, M., Zwiebel, M., Fullgrabe, M., Krczal, G. and Wassenegger, M. (2010) Analysis of an autoproteolytic activity of rice yellow mottle virus silencing suppressor P1. *Biol Chem* **391**(2-3): 271-81.

Wolf, S., Deom, C. M., Beachy, R. N. and Lucas, W. J. (1989) Movement protein of tobacco mosaic virus modifies plasmodesmatal size exclusion limit. *Science* **246**(4928): 377-9.

Wroth, J. M. and Jones, R. A. C. (1992) Subterranean clover mottle sobemovirus: its host range, resistance in subterranean clover and transmission through seed and by grazing animals. *Ann Appl Biol* **121**: 329-343.

Wu, S. X., Rinehart, C. A. and Kaesberg, P. (1987) Sequence and organization of southern bean mosaic virus genomic RNA. *Virology* **161**(1): 73-80.

Xiong, Z. and Lommel, S. A. (1989) The complete nucleotide sequence and genome organization of red clover necrotic mosaic virus RNA-1. *Virology* **171**(2): 543-54.

Yassi, M. N., Ritzenthaler, C., Brugidou, C., Fauquet, C. and Beachy, R. N. (1994) Nucleotide sequence and genome characterization of rice yellow mottle virus RNA. *J Gen Virol* **75 (Pt 2)**: 249-57.

Zaumeyer, W. J. and Harter, L. L. (1943) Inheritance of symptom expression of bean mosaic virus 4. *J Agric Res* **67**: 295-300.

Zavriev, S. K., Hickey, C. M. and Lommel, S. A. (1996) Mapping of the red clover necrotic mosaic virus subgenomic RNA. *Virology* **216**(2): 407-10.

Zell, R. and Fritz, H. J. (1987) DNA mismatch-repair in Escherichia coli counteracting the hydrolytic deamination of 5-methyl-cytosine residues. *EMBO J* **6**: 1809-1815.

Zhang, F. Y. and Toriyama, S. (2001). GenBank accession no AB040447.

Zhang, F. Y., Toriyama, S. and Takahashi, M. (2001a) Complete Nucleotide Sequence of Ryegrass Mottle Virus: A New Species of the Genus Sobemovirus. *J Gen Plant Pathol* **67**: 63-68.

Zhang, F. Y., Toriyama, S. and Takahashi, M. (2001b) Complete nucleotide sequence of Ryegrass mottle virus: a new species of the genus Sobemovirus. *J Gen Plant Pathol* **67**: 63-68.

http://biology.anu.edu.au/Groups/MES/vide/

http://www.ncbi.nlm.nih.gov/ICTVdb/Ictv/fs_sobem.htm

http://viperdb.scripps.edu