THESIS ON NATURAL AND EXACT SCIENCES B119

Properties of VPg and Coat Protein of Sobemoviruses

ALLAN OLSPERT

TUT
PRESS

TALLINN UNIVERSITY OF TECHNOLOGY
Faculty of Science
Department of Gene Technology

Dissertation was accepted for the defence of the degree of Doctor of Philosophy in Natural and Exact Sciences on October 4, 2011

Supervisor: **Professor Erkki Truve**
Department of Gene Technology, Tallinn University of Technology, Tallinn, Estonia

Opponents: **Dr. Olivier Le Gall**
Plant-Virus Interaction group, National Institute for Agricultural Research (INRA), Villenave d'Ornon, Bordeaux-Aquitane, France

Dr. Aare Abroi
Estonian Biocentre, Tartu, Estonia

Defence of the thesis: December 2, 2011

Declaration
Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology, has not been submitted for any academic degree.

/Allan Olspert/

This dissertation was supported by the European Social Fund

Euroopa Liit
Euroopa Sotsiaalfond

Eesti tuleviku heaks

ISSN 1406-4723
ISBN 978-9949-23-198-0 (publication)
ISBN 978-9949-23-199-7 (PDF)

LOODUS- JA TÄPPISTEADUSED B119

Sobemoviiruste VPg ja kattevalgu omadused

ALLAN OLSPERT

TTÜ KIRJASTUS

CONTENTS

INTRODUCTION 7
ORIGINAL PUBLICATIONS 8
ABBREVIATIONS 9
1. REVIEW OF THE LITERATURE 11
    1.1 Sobemoviruses 11
    1.2 The characteristics of VPg proteins 22
    1.3 The CP and the movement of plant viruses 25
2. AIM OF THE STUDY 28
3. MATERIALS AND METHODS 29
4. RESULTS AND DISCUSSION 30
    4.1 Sobemoviral VPg processing 30
    4.2 Characterization of VPg-RNA linking of sobemoviruses 32
    4.3 Sobemovirus VPgs are phosphorylated 35
    4.4 Subcellular localization of CfMV CP 36
    4.5 The CP of CfMV is dispensable for virus movement 38
CONCLUSIONS 42
ACKNOWLEDGEMENTS 55
PUBLICATION I 57
PUBLICATION II 71
PUBLICATION III 81
MANUSCRIPT 91
ABSTRACT 115
KOKKUVÕTE 117
CURRICULUM VITAE 119
ELULULOOKIRJELDUS 121

INTRODUCTION

Viruses are the smallest biological entities with protein coding capacity and with the ability to determine their own replication. Nevertheless viruses are obligatory intracellular parasites that do not encode their own translation apparatus. Viruses take control of the host cell and its processes through various molecular interactions to facilitate replication and the reproduction of progeny viruses. Since they require a cellular organism for reproduction viruses co-evolve with their hosts(s) forming distinct pathosystems. Different groups of viruses have a quite diverse molecular composition, size, genome size and shape. A virus exists and survives outside the host organism in a well defined molecular complex named virion, which contains the genome of the virus and all necessary proteins for the invasion of its host. A protective layer is formed around the genome and viral proteins by the structural component(s) of the virion. Capsid or coat protein(s) (CP) is the major component of this layer.

Sobemovirus is a genus of small spherical plant viruses with a positive-sense single-stranded RNA genome. Like few other genera, Sobemoviruses have a viral protein genome-linked (VPg) covalently attached to the 5' end of the viral RNAs. Thus, sobemoviral virions consist of CP, viral RNA and VPg.

The VPg proteins of other virus genera have been demonstrated to be involved in various stages of viral infection cycle. The VPg occupies a position filled by the cap structure in case of cellular messenger RNAs and is covalently linked to the viral RNA. Therefore, for at least some viruses the VPg has been shown to be involved in translation of viral RNA. However, other viruses utilize different mechanisms for translation and the VPg is used as a primer for viral RNA synthesis. Although some amount data about the functions of sobemoviral VPgs are available, majority of important biochemical aspects of these proteins have not been characterized.

Usually viral proteins are multifunctional and the CP is no exception. Besides being the building blocks of the virion, CPs of different viruses have been demonstrated to be involved in most steps of viral infection. These include vector transmission, host cell entry together with the disassembly, viral RNA translation, suppression of host defenses, viral genome replication, assembly of progeny viruses and movement. The virion structures of several sobemoviruses have been determined and virion formation is quite well understood. However, the other possible CP functions are unknown.

The goal of the present study was to further characterize VPg processing from the polyprotein and to identify the amino acid residues used for VPg-RNA linking of several sobemoviruses. Another part of this study focuses on unraveling the roles of *Cocksfoot mottle virus* CP during virus replication and movement.

ORIGINAL PUBLICATIONS

I Olspert, A., Paves, H., Toomela, R., Tamm, T., Truve, E. (2010). Cocksfoot mottle sobemovirus coat protein contains two nuclear localization signals. Virus Genes, 40, 423 - 431.

II Olspert, A., Peil, L., Hébrard, E., Fargette, D., Truve, E. (2011). Protein-RNA linkage and post-translational modifications of two sobemovirus VPgs. Journal of General Virology, 92, 445 - 452.

III Olspert, A., Arike, L., Peil, L., Truve, E. (2011). Viral RNA linked to VPg over a threonine residue. FEBS Letters, 585, 2979 - 2985.

MANUSCRIPT

Olspert, A., Kamsol, K., Sarmiento, C., Truve, E. The CP of *Cocksfoot mottle virus* is dispensable for movement.

ABBREVIATIONS

ARM	arginine-rich motif
BMV	*Brome mosaic virus*
BSSV	*Blueberry shoestring virus*
CfMV	*Cocksfoot mottle virus*
CMV	*Cucumber mosaic virus*
CnMoV	*Cynosurus mottle virus*
CP	viral coat protein
CRE	*cis*-active RNA element
CyRSV	*Cymbidium ringspot virus*
dpi	days post inoculation
EGFP	enhanced green fluorescent protein
eIF	eukaryotic translation initiation factor
GCFV	*Ginger chlorotic fleck virus*
IRES	internal ribosomal entry site
IYMV	*Imperata yellow mottle virus*
LTSV	*Lucerne transient streak virus*
MP	movement protein
mRNA	messenger RNA
NLS	nuclear localization signal
ORF	open reading frames
P1	protein encoded by ORF1
P10	10 kDa protein processed from the C-terminal part of SeMV P2a
P2a	protein encoded by ORF2a
P2ab	protein encoded by ORF2a and ORF2b utilizing -1 PRF
P8	8 kDa protein processed from the C-terminal part of SeMV P2a
PCR	polymerase chain reaction
PLYV	*Papaya lethal yellowing virus*
PRF	programmed ribosomal frameshift
Pro	protease
PVA	*Potato virus A*
R domain	random domain
RGMoV	*Ryegrass mottle virus*
RNA	ribonucleic acid
RNAi	RNA interference
RNP	ribonucleoprotein
RoMoV	*Rottboellia yellow mottle virus*
RT-PCR	reverse transcriptase PCR
RYMV	*Rice yellow mottle virus*

S domain	shell domain
SBMV	*Southern bean mosaic virus*
SCMoV	*Subterranean clover mottle virus*
SCPMV	*Southern cowpea mosaic virus*
SeMV	*Sesbania mosaic virus*
sgRNA	subgenomic RNA
siRNA	small interfering RNAs
SMAMV	*Snake melon asteroid mosaic virus*
SNMoV	*Solanum nodiflorum mottle virus*
SoMV	*Sowbane mosaic virus*
ssRNA	single-stranded RNA
TBSV	*Tomato bushy stunt virus*
TRoV	*Turnip rosette virus*
UTR	untranslated region
VPg	viral protein genome-linked
VTMoV	*Velvet tobacco mottle virus*
wt	wild-type

1. REVIEW OF THE LITERATURE

1.1 Sobemoviruses

The genus *Sobemovirus* comprises of plant viruses that have an icosahedral virion with an approximate diameter of 25-30 nm (Tamm and Truve, 2000b; Truve and Fargette, in press). The virions are assembled from 180 molecules of a single CP of about 26-31 kDa according to T = 3 symmetry. The genome of sobemoviruses is one single-stranded positive-sense RNA molecule of approximately 4-4.5 kb in length. The genome has a viral protein genome-linked (VPg) covalently attached to the 5' end of the RNA and the 3' end does not have a poly(A) tail.

According to the 9th report of the International Committee on Taxonomy of Viruses sobemoviruses are recognized as genus but have not been assigned to any family (Truve and Fargette, in press). Currently there are 13 viruses assigned to the genus, these are: *Blueberry shoestring virus* (BSSV), *Cocksfoot mottle virus* (CfMV), *Lucerne transient streak virus* (LTSV), *Rice yellow mottle virus* (RYMV), *Ryegrass mottle virus* (RGMoV), *Sesbania mosaic virus* (SeMV), *Solanum nodiflorum mottle virus* (SNMoV), *Southern bean mosaic virus* (SBMV), *Southern cowpea mosaic virus* (SCPMV), *Sowbane mosaic virus* (SoMV), *Subterranean clover mottle virus* (SCMoV), *Turnip rosette virus* (TRoV), *Velvet tobacco mottle virus* (VTMoV). In addition, there are 5 tentative species in the genus, which are *Cynosurus mottle virus* (CnMoV), *Ginger chlorotic fleck virus* (GCFV), *Imperata yellow mottle virus* (IYMV), *Papaya lethal yellowing virus* (PLYV) and *Snake melon asteroid mosaic virus* (SMAMV). *Rottboellia yellow mottle virus* (RoMoV) has previously been proposed as a tentative member (Hull and Fargette, 2005), and is indeed most likely a separate species (Sõmera, personal communication).

Although sobemoviruses are spread all over the world and infect hosts from mono- and dicotyledon plant species, the individual host range of each virus is considered relatively narrow (Sõmera, 2010; Tamm, 2000).

Many sobemoviruses are transmitted by seeds, most have an insect vector, but the main source of spread is due to mechanical transmission (Traoré *et al.*, 2008). Different species of aphids, beetles, grasshoppers, mirids and leafhoppers have been shown to be vectors of sobemoviruses (Sõmera, 2010; Tamm, 2000).

1.1.1 Genome organization and expression

The genomes of sobemoviruses contain four open reading frames (ORF) (Figure 1) (Meier and Truve, 2007). The 5' proximal ORF1 encodes the protein P1. As the 5' end of the genome does not contain a cap structure but VPg, the

mechanism of P1 translation initiation is currently not entirely clear. The VPg of RYMV has been shown to interact with eukaryotic translation initiation factor (eIF) eIF(iso)4G (Hébrard *et al.*, 2006; Hébrard *et al.*, 2010). It has been proposed that this interaction serves for ribosome recruitment (Hébrard *et al.*, 2010). However, it has also been shown that the 5' UTR of CfMV operates as a translational enhancer and might contain an internal ribosomal entry site (IRES) (Mäkeläinen, 2006).

The central part of the genome contains two overlapping ORFs, ORF2a and ORF2b, which encode the polyprotein. Historically sobemoviruses were divided into two groups based on differences in the organization of the central part of the genome. However, re-sequencing has proved that all sobemoviruses have similar organization which includes ORF2a and ORF2b (Meier and Truve, 2007). The AUG codon of the ORF1 of sobemoviruses is in poor context for optimal translation initiation, while the AUG ORF2 is in good context. The translation initiation of ORF2 of SCPMV has been analyzed in detail (Sivakumaran and Hacker, 1998). The introduction of upstream initiation codons or mutating the ORF1 initiation codon context to more favorable one reduced protein synthesis from ORF2a, while the deletion of the AUG of ORF1 increased translation from ORF2a. These results rule out an IRES in the vicinity of ORF2a AUG and suggest that translation of ORF2a is initiated by scanning ribosomes that have missed the AUG of ORF1 due to its poor context, i.e. leaky scanning. ORF2b is expressed as a fusion product of P2ab through – 1 programmed ribosomal frameshift (PRF) mechanism (Mäkinen *et al.*, 1995a). -1 PRF is in general regulated by two RNA regions, a heptamer slippery sequence and a downstream secondary structure element, either a pseudoknot or a stem-loop. For sobemoviruses these elements were first discovered in the genome of CfMV, which has a UUUAAAC slippery sequence shortly followed by a putative stem-loop structure (Mäkinen *et al.*, 1995a). CfMV has been shown to utilize -1 PRF for the synthesis of the polyprotein and the PRF efficiency has been demonstrated to be around 10% (Tamm *et al.,* 2009; Lucchesi *et al.*, 2000; Mäkeläinen and Mäkinen, 2005; Mäkinen *et al.*, 1995a). The P2a polyprotein contains the domains of protease (Pro), VPg, P10 and P8, whereas the fusion P2ab contains Pro, VPg and the RNA-dependent RNA polymerase (RdRp) (Nair and Savithri, 2010b).

The 3’ proximal ORF3 expresses the viral CP from its corresponding subgenomic RNA (sgRNA) (Ghosh *et al.*, 1981; Rutgers *et al.*, 1980). The 5' ends of SBMV and SCPMV sgRNAs have been identified and they start with the primary sequence ACAAAA, which is identical to the 5' end of their respective genomes (Hacker and Sivakumaran, 1997). Based on the identity to 5' genomic sequence the sgRNA start sites can be predicted for several other members of the genus, but not for all sobemoviruses (Tamm and Truve, 2000b). Similarly to ORF1, the introduction of upstream AUG to the 5' UTR of SCPMV sgRNA

reduced CP translation, which indicated that the translation occurs by scanning rather than via IRES (Hacker and Sivakumaran, 1997).

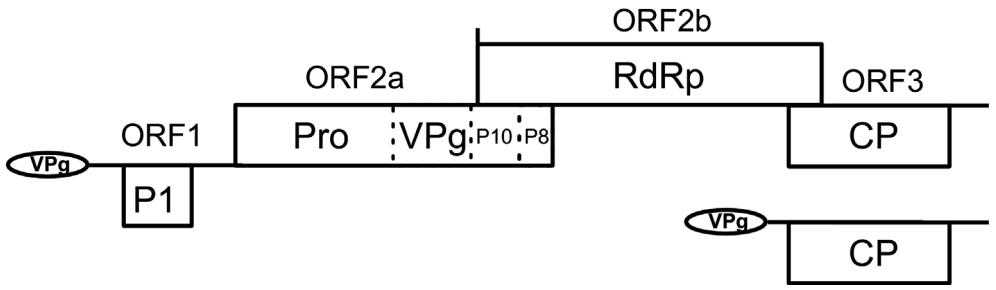


Figure 1. Genome organization of sobemoviruses
An ellipse notes the covalently attached VPg at the 5' ends of the genomic and subgenomic RNAs. Boxes correspond to open reading frames (ORF) with the according protein names indicated within. P1 – RNA silencing suppressor; viral polyprotein domains Pro – protease, VPg – viral genome linked protein, P10, P8 and RdRp – RNA-dependent RNA-polymerase; CP - viral coat protein. The -1 programmed frameshift signal is indicated with a tick in the beginning of the ORF2b coding for RdRp.

1.1.2 Sobemoviral P1 and polyprotein

P1

The ORF1 region which encodes the P1 proteins is the most variable part of sobemovirus genomes (Sõmera, 2010; Tamm and Truve, 2000b). The P1s lack any sequence similarity with other viral or non-viral proteins at both nucleotide as well as amino acid levels. The molecular weights of P1 proteins range from 11 to 25 kDa. CfMV P1 has been reported to bind ssRNA in a sequence-nonspecific manner (Tamm and Truve, 2000a). RYMV and CfMV P1s have been reported to be suppressors of RNA interference (RNAi) (Lacombe *et al.*, 2010; Sarmiento *et al.*, 2007; Siré *et al.*, 2008; Voinnet *et al.*, 1999). CfMV P1 suppressed both local and systemic RNAi, however the suppression mechanism remained unknown as it did not bind double stranded small interfering RNAs (siRNA) (Sarmiento *et al.*, 2007). It is interesting that for RYMV P1 both suppression and activation of RNAi has been demonstrated (Lacombe *et al.*, 2010). Due to the availability of large number of different isolates, the role of RYMV P1 in RNAi has been studied extensively. Previously, RYMV P1 cell-to-cell movement ability was reported to be in correlation with the efficiency of silencing suppression (Siré *et al.*, 2008). Later it was demonstrated that RYMV P1 can also enhance the short- and long-distance spread of silencing (Lacombe *et al.*, 2010). Transgenic rice expressing RYMV P1 demonstrated specific deregulation of DCL4-dependent siRNA pathway in the host (Lacombe *et al.*, 2010). According to this RYMV P1 probably suppresses RNAi through inhibition of DCL4 RNAi pathway. Recently it was shown that SeMV P1 interacts with other viral proteins such as VPg, P10 and CP (Chowdhury and Savithri, 2011; Roy Chowdhury and Savithri, 2011).

It has also been speculated that the P1 proteins are the movement proteins (MP) of sobemoviruses. The P1s of RYMV, SCPMV and CfMV have indeed been demonstrated to be dispensable for virus replication but indispensable for virus movement (Bonneau *et al.*, 1998; Meier *et al.*, 2006; Sivakumaran *et al.*, 1998). Interestingly, the movement disability was complemented *in trans*, as RYMV mutant lacking P1 produced systemic infection in transgenic rice plants that expressed RYMV P1 (Bonneau *et al.*, 1998). In addition, the CfMV P1 fusion to enhanced green fluorescent protein (EGFP) has been shown to be capable of limited movement between epidermal cells, when expressed transiently independent of other viral proteins (Meier *et al.*, 2006). However, currently it is not clear whether the P1 contributes to the virus trafficking as a suppressor of RNAi or/and as an MP.

Polyprotein

Historically the first sobemoviral polyprotein domain was identified by sequence comparisons when the SCPMV polyprotein sequence was found to

contain a putative serine protease (Gorbalenya *et al.*, 1988). The protease cleavage specifity was proposed based on the known specificities of viral proteases and SCPMV sequence to be between E/T or E/S (Gorbalenya *et al.*, 1988). By *in silico* analysis the domains for VPg and RdRp were also proposed in their respective parts of P2a and P2ab (Mäkinen *et al.*, 1995b).

SeMV is currently the only sobemovirus for which *in vitro* polyprotein processing is fully characterized (Nair and Savithri, 2010b). The SeMV P2a domain order has been identified to be Pro-VPg-P10-P8, whereas the frameshift fusion product P2ab contains the domains of Pro-VPg-RdRp. The N-terminus of SeMV Pro has been predicted to contain transmembrane helixes (Satheshkumar *et al.*, 2004a). SeMV protease has been reported to be activated through hydrophobic interactions with the flanking VPg domain (Nair *et al.*, 2008; Satheshkumar *et al.*, 2005a). SeMV protease can cleave the polyprotein either between E/T or E/S residues. Based on the analysis of processing sites, the proposed SeMV Pro specificity is N,Q-E/T,S-X, where X is an aliphatic residue (Nair and Savithri, 2010b). The VPg N-termini of SBMV, CfMV and RYMV have been determined to be cleaved between E/T, E/N and E/S residues, respectively (Hébrard *et al.*, 2008; Mäkinen *et al.*, 2000; van der Wilk *et al.*, 1998). In addition to cleavage between each domain, the SeMV polyprotein P2a is cleaved also between the core protease domain and the hydrophobic N-terminal membrane anchoring domain of the protease (Nair and Savithri, 2010b). This cleavage site upstream of the Pro core domain can also be predicted for other sobemoviruses (Mäkinen *et al.*, 2000). Furthermore, it seems that the cleavage of the N-terminal hydrophobic part of protease is required for the maturation and efficient full processing of the polyprotein, since mutation of that cleavage site abolishes cleavage between P10 and P8 domains (Nair and Savithri, 2010b). The *in vitro* cleavage between Pro and the N-terminal anchoring domain of P2ab was rather slow, which may indicate the relevance of this particular processing site mainly for the regulation of P10 and P8 processing. In addition, cleavage of P2a between N-terminal – Pro and P10 - P8 domains took place only *in cis*, which may indicate the requirement of these processing events at early stages of infection when there is not enough polyprotein for efficient *trans* cleavage. Processing of SeMV P2ab suggests that the protein is first cleaved between Pro and VPg domains releasing VPg-RdRp, which is not cleaved further *in vitro*. However, VPg present at the N-terminus of RdRp has been shown to have inhibitory effect on the *in vitro* polymerase activity of RdRp (Nair and Savithri, 2010a). The only information about the polyprotein processing of other sobemoviruses comes from the analysis of the VPgs (see below).

Protease. Gorbalenya *et al.* (1988) have proposed that the SCPMV protease is probably a serine protease with the catalytic triad H181, D216 and S284. These amino acid residues are also conserved in all other sobemoviruses. The

crystal structure of the protease domain of SeMV has been determined at a resolution of 2.4 Å (Gayathri *et al.*, 2006). The 3D structure of SeMV was found to be closer to cellular serine proteases than to other viral proteases. The SeMV protease domain belongs to the family of trypsin-like serine proteases. The protein consists of two β-barrels connected by a long loop. The active site and the substrate binding cleft occur between the two domains and are fairly exposed. There are only three helices in the structure. The catalytic triad residues H181, D216 and S284 indeed form the active site. It has been demonstrated that the mutation of any of the active site residues to alanine rendered the protease inactive (Satheshkumar *et al.*, 2004a). Residues T279, A280, H298, F301 and N308 form the proposed substrate binding S1 pocket (Gayathri *et al.*, 2006). Residues T279, H298 and N308 are conserved among the primary sequences of sobemovirus proteases. The mutation of the residues H298, T279 and N308 and the analysis of *cis* cleavage revealed that T279A and H298A mutants were inactive, while the N308A mutant was partially active, suggesting that the interactions of T279A and H298A with the substrate glutamate are crucial for the E/T,S cleavage. The structure revealed a stretch of aromatic amino acids exposed on the surface of the protease, which was presumably a protein–protein interaction interface. Indeed, it was later demonstrated that W271 and H275 of the Pro domain mediate aromatic stacking interactions with W43 of VPg which activate protease (Nair *et al.*, 2008). This regulation is proposed to be required for modulating the function of the protease. Interestingly, H275 is conserved across the proteases of sobemoviruses.

VPg. The genomes of SBMV and SCPMV were the first sobemoviruses for which the 5' end of the genome was determined to be linked to a protein (Ghosh *et al.*, 1981; Mang *et al.*, 1982). Since then the same has been demonstrated for CfMV, SeMV and RYMV (Hébrard *et al.*, 2008; Lokesh *et al.*, 2001; Mäkinen *et al.*, 2000). All sequenced sobemoviruses contain the VPg domain, however there is little similarity between the sequences. The only conserved motif among the sequences is the WAD/WGD/WNK motif followed by an E/D rich region (Mäkinen *et al.*, 2000; McGavin and MacFarlane, 2009). The VPgs of SeMV and RYMV have been demonstrated to be intrinsically unfolded/disordered proteins (Hébrard *et al.*, 2009b; Satheshkumar *et al.*, 2005a). Disordered proteins, that lack a uniform structure, exist in solution in a variety of conformations, but they still have the propensity to form more rigid structures upon stabilization (Hébrard *et al.*, 2009b). RYMV VPg has been demonstrated to contain some residual structured regions while clearly containing disordered regions. An unfolded nature can also be predicted for other sobemoviral VPgs, including CfMV, SBMV and RGMoV (Hébrard *et al.*, 2009b).

Although VPg has been demonstrated to be linked to viral RNA, its role in the replication of sobemoviruses is unknown. Yet, RYMV VPg has been identified to be the virulence factor (Hébrard *et al.*, 2006). It has been proposed

and demonstrated that RYMV VPg interacts directly with the central domain of eIF(iso)4G (Hébrard *et al.*, 2008; Hébrard *et al.*, 2006; Hébrard *et al.*, 2010). It has been proposed that the central part of the VPg adopts a helical structure upon the interaction with eIF(iso)4G (Hébrard *et al.*, 2009b; Hébrard *et al.*, 2008). Mutations in eIF(iso)4G1 correspond to several described RYMV resistance alleles. However, the resistance of these alleles can be "broken" by mutations in the VPg (Hébrard *et al.*, 2006). The evolution of most resistance breaking RYMV strains resulted in the fixation of the same VPg mutations, predominantly at codon 48 (Pinel-Galzi *et al.*, 2007). The position 48, which corresponds to R in avirulent isolates, is occupied by G, I or E in virulent isolates. The virulence mutations occurred according to several different mutational pathways. Most prevalently R48 was displaced by glycine in the first step to become fixed at glutamic acid in the second step. The virulent variants have been suggested to emerge from a residual multiplication of wild-type (wt) isolates in resistant plants. The virulent isolates with the E48 mutation restored the optimal multiplication level in resistant plants, whereas the I48 and G48 mutations displayed intermediate level of accumulation (Poulicard *et al.*, 2010). Interaction analysis of VPg variants and eIF(iso)4G1 variants have demonstrated that the mutation in eIF(iso)4G1 reduces the interaction with R48 VPg, whereas the resistance breaking mutations (48G/I/E) improve (G/I) or restore (E) the interaction (Hébrard *et al.*, 2010). The interaction strength was in correlation with the level of virus accumulation. Although the exact biological role of this interaction has not been demonstrated, it has been proposed as a strategy to recruit the translation initiation complex. According to this theory the VPg substitutes the role of the cap structure in translation. It is interesting to note that the VPg of SeMV has been demonstrated to interact with P1 (Roy Chowdhury and Savithri, 2011).

P10 and P8. The function(s) of the C-terminal part of P2a polyprotein are poorly understood. Recently it was demonstrated for SeMV that the domain is processed to two small proteins, P10 and P8 (Nair and Savithri, 2010b). The P8 domain, like VPg, was found to be natively unfolded (Nair and Savithri, 2010a). In addition it was demonstrated that P8 could bind single- and double-stranded nucleic acids. Interestingly, P10–P8 fusion, but not P8 alone, was reported to exhibit Mg^{2+}-dependent ATPase activity, that was inhibited in the presence of poly(A) (Nair and Savithri, 2010a). In the absence of P8, the ATPase activity of the P10 was reduced suggesting that the natively unfolded P8 domain influenced the P10 ATPase. The P8 has also been predicted to contain a nuclear localization signal (NLS) (Nair and Savithri, 2010a). The meaning of these properties of P8 and P10 in the virus infection cycle is unknown. In addition, the P10 of SeMV has been demonstrated to interact with P1 (Roy Chowdhury and Savithri, 2011).

RdRp. The viral replicase was predicted to be a located at the C-terminal part of the P2ab by sequence comparisons (Koonin, 1991; Koonin and Dolja, 1993).

All sobemoviral sequences contain the GDD motif highly conserved among viral RdRps. Currently the properties of SeMV RdRp have been analyzed in more detail (Govind and Savithri, 2010). The SeMV RdRp domain, expressed and purified from *E. coli,* was determined to be an active RNA polymerase. The recombinant RdRp was capable of initiating minus strand RNA synthesis from the genomic RNA or sgRNA template in the absence of the putative protein primer VPg. *In vitro* the RdRp also failed to nucleotidylate the VPg. Mutation of the conserved GDD to GAA abolished synthesis activity. Similarly to other recombinant viral RdRps, the activity was dependent on Mg^{2+} ions and was enhanced several fold by Mn^{2+} ions. The method of initiation was determined to be *de novo* synthesis and not template priming. However, SeMV RdRp showed preference for structured viral RNA templates and was inactive on homopolymeric RNA templates. Furthermore, a stem-loop structure at the 3′ end of the SeMV positive strand RNA and sgRNA template was shown to be important for RNA synthesis. The mutational analysis of the required stem-loop structure showed that the structure is more important for initiation than the sequence. Further analysis showed that the SeMV RdRp was capable of recognizing stem-loop structures of various lengths and forms, which suggested that the RdRp is flexible enough to recognize different conformations. It is interesting that SeMV RdRp was also able to initiate the synthesis of the positive strand independent of VPg, but at a significantly reduced level. This suggests that *in vivo* the RNA synthesis of at least the sense strand might be primer-dependent or requiring other viral/host factors.

1.1.3 CP and the virion

Structure. The crystal structures of the virions of several sobemoviruses have been determined. These are SCPMV (Abad-Zapatero *et al.*, 1980), SeMV (Bhuvaneshwari *et al.*, 1995), RYMV (Qu *et al.*, 2000), CfMV (Tars *et al.*, 2003) and RGMoV (Plevka *et al.*, 2007). The CP primary sequences of these viruses have the identity from 12 to 30%, not taking into account two very closely related SeMV and SCPMV with 63% of identity (Plevka *et al.*, 2007). The 3D structures of their virions, however, were nearly identical, root mean square deviations between superimposed backbone C-α atoms being roughly 1.4 - 1.9 Å. The monomers of the CP have an 8 strand anti-parallel jellyroll β-sandwich topology, which is common to most nonenveloped icosahedral viruses. 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. The A subunits form pentamers around the icosahedral fivefold axis and three of each of the B and C subunits cluster into hexamers around the threefold axis. The pentamers and hexamers differ significantly in shape, hexamers being

generally planar and pentamers substantially bent. In addition, the subunit contacts between AB, BC and CA are stabilized by cation binding pockets, which contain Ca^{2+} ions.

The CP can be divided into two domains, the C-terminal S (shell) domain, which has the jellyroll β-sandwich topology, and the N-terminal R (random) domain buried within the virion. The R domain is disordered in subunits A and B, but is partially structured in subunit C. The S domain is the building block of the virion, whereas the R domain is involved in the regulation of the capsid structure and is presumably in contact with RNA. Despite very similar overall virion structures, two different structures of the R domain in the C subunit of sobemoviral virions have been found. In SCPMV, SeMV and RGMoV the N-terminus of subunit C makes a turn and extends toward the threefold axis nearest to C where it makes a β-structure together with R domains from analogous C subunits. In RYMV and CfMV there is no turn, instead the N-terminal arm of subunit C extends towards subunit B and makes a similar structure at the distal threefold axis closest to subunit B. In general, the virion of RGMoV is slightly smaller than virions of other sobemoviruses and more identical to *Tobacco necrosis virus* (TNV), a necrovirus belonging to the *Tombusviridae* family, which has a remarkably similar CP and virion structure to all sobemoviruses. Although the virion also contains the viral RNA and VPg, they are not visible in the crystal structures.

Virion formation. When the entire R domain of SCPMV or SeMV CP was removed, only T = 1 particles were formed (Lokesh *et al.*, 2002; Savithri and Erickson, 1983). In these T = 1 particles all contacts between the subunits were bent and the particle comprised of only A subunits clustering similarly to the fivefold axis of wt particle. There is further evidence suggesting that the virion assembly could be nucleated at fivefold axes since pseudo T = 2 SeMV particles comprised of AB dimers arranged similarly to the icosahedral fivefold axes of T = 3 particles (Lokesh *et al.*, 2002). Furthermore, a single point mutation at the fivefold interface of the virus arrested assembly and resulted in stable soluble AB dimers (Pappachan *et al.*, 2009).

The R domain of all sobemovirus CPs is rich in basic amino acid residues and contains an arginine-rich motif (ARM). Studies with SCPMV and SeMV CP have demonstrated that the ARM is essential for RNA encapsidation but not for particle formation (Erickson and Rossmann, 1982; Lokesh *et al.*, 2002; Satheshkumar *et al.*, 2005b). Although the presence of RNA enhanced the overall stability of capsids, RNA encapsidation was not required for particle assembly. Also, the role of the metal ion binding sites on virion assembly has been assessed (Sangita *et al.*, 2004; Satheshkumar *et al.*, 2004b). Mutational analysis of residues involved in Ca^{2+} site formation of SeMV virions revealed that the metal binding sites are not involved in assembly but rather necessary for maintaining the stability of particles. It has been proposed for sobemovirus

virions and also for the virions from other genera that the β-annulus structure formed by the R domain of the C subunit is responsible for regulating the curvature and T = 3 particle formation (Lokesh *et al.*, 2002; Plevka *et al.*, 2007 and references within). However, further analysis suggested that the β-annulus structure could form as a result of virion formation and is not required for T = 3 SeMV particle formation (Pappachan *et al.*, 2008; Satheshkumar *et al.*, 2005b). Moreover, the latter results suggested that the curvature is sufficiently maintained by contacts in the region downstream of the β-annulus and that sufficient length of the N-terminus is required for the stability of these contacts.

Virion stability. The virions of sobemoviruses are stabilized by RNA–protein, protein–protein and calcium-mediated protein–protein interactions. Studies with TRoV, SCPMV, SeMV and RYMV particles have demonstrated that the stability of the virions depends greatly on pH and the availability of Ca^{2+} ions (Brugidou *et al.*, 2002; Hull, 1977; Sangita *et al.*, 2004; Satheshkumar *et al.*, 2004b; Savithri and Erickson, 1983). Upon alkaline pH or removal of the cations the virus particles swell and become less stable. Analysis of SeMV CP has demonstrated that cation-mediated interactions are mainly needed for particle stability, while RNA interactions have a smaller role in stability (Satheshkumar *et al.*, 2005b; Satheshkumar *et al.*, 2004b). For RYMV and SeMV a model of infection based on stability and subcellular compartmentalization has been proposed (Brugidou *et al.*, 2002; Opalka *et al.*, 1998; Satheshkumar *et al.*, 2004b). After cell entry the virion is swollen by basic pH and low Ca^{2+} concentration of the cytoplasm. This is required for particle disassociation and the release of viral RNA for translation. *In vitro* evidence exists that SBMV virions remain swollen or at least partially intact during translation (Shields *et al.*, 1989). During subsequent virus accumulation new virions are assembled. Presumably at later stages of infection virions accumulate to the vacuoles and other vesicles, which have acidic pH and higher Ca^{2+} concentration and therefore are able to form stable compact virions.

Properties of CP and role in infection cycle. *In vitro* experiments with CfMV and SCPMV CPs have demonstrated that both proteins bind RNA in a sequence non-specific manner (Lee and Hacker, 2001; Tamm and Truve, 2000a). In addition, SCPMV CP has been demonstrated to possess sequence- or structure-specific viral RNA binding properties as virion-derived ribonucleoprotein (RNP) complexes were determined to be in contact with only one specific region of the genomic RNA (Hacker, 1995). The amino acid residues responsible for SCPMV CP RNA binding *in vitro* were mapped to the N-terminal region containing the ARM (Lee and Hacker, 2001). Mutation analysis has demonstrated that the ARM of SeMV CP is also responsible for RNA encapsidation (Satheshkumar *et al.*, 2005b). In addition, it has been demonstrated that the N-terminal 30 residues of SCPMV CP could interact with membranes (Lee *et al.*, 2001). These 30 residues were determined to undergo a

conformational change from a random coil to an α-helix upon interaction with membranes. However, the biological relevance of that particular feature of SCPMV CP is unknown. Based on sequence similarity it has been supposed that the N-terminal region of sobemovirus CPs contains an NLS (Mäkinen *et al.*, 1995b; Yassi *et al.*, 1994). The CP of SeMV has been demonstrated to interact with P1 (Chowdhury and Savithri, 2011).

The CPs of RYMV and SCPMV have been demonstrated to be dispensable for virus replication, but needed for virus movement (Brugidou *et al.*, 1995; Sivakumaran *et al.*, 1998). Both of these viruses, devoid of CP, were able to replicate in protoplasts. The removal of CP of SCPMV completely abolished infectivity in plants. However, RYMV CP deletion mutant was able to accumulate in inoculated leaves which indicated that the mutant RYMV was probably still capable of local cell-to-cell movement. The long distance movement of SCPMV and RYMV has been proposed to be dependent on particle formation (Fuentes and Hamilton, 1993; Hacker and Fowler, 2000; Opalka *et al.*, 1998). SCPMV systemic movement has been studied in mixed infection environments. SCPMV infection in non-systemic host has been correlated with the lack of correct virion formation and the systemic movement in the same host has been shown to be complemented by heterologous encapsidation by SBMV CP (Fuentes and Hamilton, 1993). In case of RYMV the systemic infection was detected only after virions were detected in inoculated cells (Brugidou *et al.*, 2002).

Analyses of the movement of unrelated viruses together with sobemoviral CPs have also been studied. The CP of TRoV was reported to facilitate long distance movement of red clover necrotic mosaic dianthovirus (Callaway *et al.*, 2004). The CfMV CP has been shown to complement to some extent the cell-to-cell movement of *Potato virus X* CP mutants (Fedorkin *et al.*, 2001).

The role of RYMV CP has also been assessed using transgenic plants. RYMV accumulated at higher levels in plants expressing RYMV CP or CP mutant, where the putative NLS was deleted, than in control plants (Kouassi *et al.*, 2006). These results demonstrated that the CP present *in trans* did not interfere with virus replication and even enhanced virus infection. Unfortunately, the mechanisms of enhancement have not been addressed further.

Distribution within host. Sobemovirus particles have been detected in most cell types: mesophyll, epidermal, guard and bundle sheath cells (Chamberlain and Catherall, 1976; Fuentes and Hamilton, 1993; Mohamed and Mossop, 1981; Opalka *et al.*, 1998; Rabenstein and Stanarius, 1984). Within vascular tissues sobemoviruses particles have been found both in phloem and in xylem. RYMV, BSSV, SoMV and SMAMV (Hartmann *et al.*, 1973; Lecoq *et al.*, 2011; Lombardo *et al.*, 1971; Urban *et al.*, 1989) have been observed mainly in xylem. For RYMV the virus particles have been found in xylem parenchyma and xylem vessels while at a lesser extent also in phloem parenchyma, sieve elements of the

phloem and bundle sheath cells (Brugidou *et al.*, 2002; Opalka *et al.*, 1998). However, based on the apparent higher accumulation of virus particles in xylem cells, it has been proposed that RYMV utilizes xylem as the main route of movement. Proposedly, RYMV particles accumulate in xylem cells which undergo programmed cell death to form the tracheids. During this maturation RYMV is released and distributed throughout the plant along with the solute. Contrary, CfMV, SBMV and SCPMV particles have been found mostly in the phloem of infected plants (Chamberlain and Catherall, 1976; Fuentes and Hamilton, 1993; Morales *et al.*, 1995; Schneider and Worley, 1959; Weintraub and Ragetli, 1970).

In root tissue, RYMV, BSSV and CfMV have been found only in vascular tissues (Hartmann *et al.*, 1973; Opalka *et al.*, 1998; Otsus *et al.*, manuscript). RYMV and BSSV virions were detected in xylem while CfMV was detected in phloem as well as xylem.

Subcellularly sobemovirus particles have been detected in the cytoplasm, vacuoles and other types of vesicles in infected cells, but not in the chloroplast or mitochondria (Fuentes and Hamilton, 1993; Mohamed and Mossop, 1981; Opalka *et al.*, 1998; Rabenstein and Stanarius, 1984). Majority of virus particles have been found in cytoplasm and vacuoles where the particles form paracrystalline arrays. Sobemovirus particles have also been found in the nuclei of infected cells (Chamberlain and Catherall, 1976; Fuentes and Hamilton, 1993; Mohamed and Mossop, 1981; Rabenstein and Stanarius, 1984).

1.2 The characteristics of VPg proteins

The 5' ends of single-stranded positive-sense RNA virus genomes are unmodified, capped or have a VPg. The 5' end of viral RNA of at least picornaviruses, caliciviruses, comoviruses, poleroviruses, enamoviruses, nepoviruses, potyviruses and sobemoviruses has been reported to contain VPg (Hull, 2002; Sadowy *et al.*, 2001; Salas, 1991). In addition, VPgs are found in the termini of some dsRNA viruses (Sadowy *et al.*, 2001; Salas, 1991). VPgs from different genera are highly variable in both size and composition. For instance, picornaviral VPgs have a molecular weight around 2.5 kDa while potyviral VPgs are around 25 kDa. In general the proteins do not contain common motifs or conserved regions. Being covalently linked to RNA is actually the only common feature between them. However, despite this variability, VPgs have some common roles in the infection cycle and share some properties.

The VPgs are attached to the RNA over a phosphodiester bond formed between the 5' phosphate group of the RNA and the hydroxyl group of an amino

acid residue (Ambros and Baltimore, 1978; Rothberg *et al.*, 1978). The amino acid residue to which RNA is linked to has been reported to be either tyrosine or serine (Ambros and Baltimore, 1978; Jaegle *et al.*, 1987). The chemical formula of a phosphodiester bond between tyrosine or serine and a nucleotide is shown in publication II, Figure 2. Although threonine also contains a hydroxyl group, there are no reports demonstrating its use for viral RNA attachment to VPg.

Picornaviruses were the first for which the chemical nature of the bond between RNA and VPg was determined (Ambros and Baltimore, 1978; Rothberg *et al.*, 1978). Picornaviruses utilize a highly conserved tyrosine residue, at position 3 of VPg, for RNA linking (Schein *et al.*, 2006). Potyviruses and caliciviruses have also been reported to use conserved tyrosine residue in the N-terminal part of VPg for the same purpose (Anindya *et al.*, 2005; Belliot *et al.*, 2008; Machín *et al.*, 2001; Murphy *et al.*, 1991). RNA linkage through a serine residue has been demonstrated for nepoviruses and comoviruses (Jaegle *et al.*, 1987; Zalloua *et al.*, 1996). Due to some level of conservation between VPgs of related viruses the general conclusion has been that viruses within the same genera utilize a positionally conserved serine or tyrosine. Mutation analysis of the residue involved in RNA linking has been used at least for poty-, como-, and caliciviruses (Carette *et al.*, 2001; Mitra *et al.*, 2004; Murphy *et al.*, 1996). Mutation to a residue not containing a hydroxyl group and thus preventing the formation of phosphodiester bond or replacing the residue with another hydroxyl group containing residue have both been reported to be lethal for the virus. Most likely substitution of the native residue to another hydroxyl group containing residue with different size and chemical properties is not tolerated by the mechanisms involved in the synthesis of the phosphodiester bond (see below).

The role of VPg in replication has been thoroughly characterized for picornaviruses. The VPg is first di-uridylated (VPg-pUpU) and subsequently used as a primer for both negative and positive strand viral RNA synthesis (Steil and Barton, 2009a; Steil and Barton, 2009b and references within). The initial uridylation is dependent on *cis*-active RNA element (CRE), which functions as a template for the conversion of picornaviral VPg into the di-uridylated form. The VPg uridylation reaction is a separate process from replication. The CRE, which is situated in the central part of the genome, is used for the VPg uridylation and afterwards the VPg-pUpU is used to initiate replication from genome termini. The replication of other VPg containing viruses is far less understood. For caliciviruses both primer-dependent and -independent RdRp activity has been reported (Belliot *et al.*, 2008 and references within, Rohayem *et al.*, 2006). However, for several caliciviruses the nucleotidylation of VPg by the respective RdRp has been reported, suggesting that VPg is used in replication as a primer (Belliot *et al.*, 2008; Han *et al.*, 2010). Nucleotidylation of the VPg by RdRp has been reported also for potyviruses and enamoviruses (Anindya *et al.*, 2005; Osman *et al.*, 2006; Puustinen and Mäkinen, 2004). *Potato virus A* (PVA) VPg

has been reported to contain an NTP-binding site, which is required for the nucleotidylation (Puustinen and Mäkinen, 2004). In contrast, it was recently demonstrated that SeMV VPg is not required by the RdRp for the *in vitro* synthesis of at least the negative strand of the viral RNA (Govind and Savithri, 2010).

The cap structure at the 5' end of cellular mRNAs is usually required for translation. As VPgs occupy a similar position it has long been suspected that they also participate in translation. The VPgs of different viruses have been demonstrated to interact with a variety of proteins involved in translation - eIF4E, eIF4G, eIF4A, eIF3 or the poly(A)-binding protein (Daughenbaugh *et al.*, 2003; Daughenbaugh *et al.*, 2006; Goodfellow *et al.*, 2005; Hébrard *et al.*, 2010; Khan *et al.*, 2008; Michon *et al.*, 2006; Miyoshi *et al.*, 2006). Experiments with different caliciviruses have shown that interactions between VPg and components of translation initiation complex were required for viral RNA translation (Chaudhry *et al.*, 2006; Goodfellow *et al.*, 2005). The translation of VPg-linked mRNA was blocked by the eIF4E inhibitor protein, 4E-BP1, indicating that translation requires VPg and eIF4E interaction (Goodfellow *et al.*, 2005). However, in another study translation was demonstrated to be insensitive to eIF4E depletion and eIF4G cleavage (Chaudhry *et al.*, 2006). Still, the translation of both tested caliciviruses was dependent on eIF4A (Chaudhry *et al.*, 2006). Although these interactions with eIFs have often been proposed for recruitment of translation machinery, data with potyviral VPgs suggest that the interaction is required for the inhibition of cap-dependent and the enhancement of cap-independent translation (Khan *et al.*, 2008; Miyoshi *et al.*, 2008). In this case the translation of viral RNA is probably initiated from IRES. Indeed, IRESs have been identified within the 5' leader sequences of potyviruses and nepoviruses (Karetnikov and Lehto, 2007; Zeenko and Gallie, 2005). Similarly, for picornaviruses it has been demonstrated that VPg is not required for translation as it is removed from the RNA and translation occurs in an IRES-dependent manner (Ambros *et al.*, 1978; Nomoto *et al.*, 1977).

VPgs from different virus genera have been demonstrated to be at least partially intrinsically unfolded/disordered proteins. These include the VPgs of PVA, *Potato virus Y*, *Lettuce mosaic virus,* SeMV and RYMV (Grzela *et al.*, 2008; Hébrard *et al.*, 2009b; Rantalainen *et al.*, 2008; Satheshkumar *et al.*, 2005a). The lack of a rigid 3D-structure and the ability to adopt various dynamic conformations has often been suggested to be the source of various VPg functions and proposedly a wide range of interactions. For some VPgs the structure in interaction with a partner has been experimentally characterized, whereas for others only structural interaction models exist. For instance, the crystal structure of VPg in complex with RdRp has been determined for picornaviruses (Ferrer-Orta *et al.*, 2006; Gruez *et al.*, 2008; Schein *et al.*, 2006).

Besides replication and translation potyviral VPgs have been implicated also

in other processes. The VPg of PVA seems to possess auxiliary RNAi suppressor functionality (Rajamäki and Valkonen, 2009). PVA VPg also contains two NLSs which are required for efficient replication, accumulation and movement of the virus (Rajamäki and Valkonen, 2009). In the nucleus PVA VPg interacts with fibrillarin, a nucleolar protein required for the long-distance movement of umbraviruses (Kim *et al.*, 2007a; Kim *et al.*, 2007b; Rajamäki and Valkonen, 2009).

1.3 The CP and the movement of plant viruses

As a result of transmission, plant viruses are usually introduced to one or few cells in which initial replication occurs. In order to infect the host systemically, plant viruses need to invade the whole organism rather than just remaining to the initially infected cells. The virus also has to overcome host defense mechanisms to establish infection and accumulate in the host to raise the odds for further transmission. To facilitate optimal progeny production, viruses have to be able to move between adjacent cells (cell-to-cell movement) and to other plant organs through vascular tissues (systemic movement). Whether these two distinct transport processes utilize the same mechanism(s) depends on the specific virus-host pathosystem. Some viruses and virus-host combinations also have limited virus movement and the infection is confined to the inoculated leaf.

Plant cells are interconnected by plasmodesmata (Benitez-Alfonso *et al.*, 2010 and references within). The plasmodesmata are basically channels connecting adjacent cells formed during cell division. These channels connect the plasmamembrane, endoplasmatic reticulum membrane and cytoplasms of the neighboring cells. The connections provide the route for the movement of soluble and membrane associated molecules between the cells and eventually throughout the plant. The flow of molecules between cells is carefully regulated by modulating the permeability of the plasmodesmata. Specifically, the movement of larger molecules such as proteins and nucleic acids does not occur by simple diffusion.

Virally encoded MPs are the key elements which enable viruses to utilize the intracellular trafficking pathway for virus distribution (reviewed by Benitez-Alfonso *et al.*, 2010; Lucas, 2006; Scholthof, 2005). Although the mechanisms how the MPs achieve this are not entirely understood and clearly vary even between related viruses, a few commonly recognized concepts exist:

(a) Viruses move through plasmodesmata as virions or RNPs;

(b) the MPs usually increase the size exclusion limit of plasmodesmata to allow larger complexes to pass through;

(c) movement involves intracellular targeting of the movement complexes to plasmodesmata;

(d) elements of cytoskeleton are involved in the inter- and intracellular movement.

Some viruses encode a single MP but others have the same functionality divided between several different proteins.

In addition to the MP and the host factors most viruses also require the CP for virus movement (reviewed by Callaway *et al.*, 2001; Lucas, 2006; Scholthof, 2005). Based on the requirement of CP for transport, plant viruses can be divided into three major categories. Viruses in the first group only require MP, and CP is dispensable for movement. Members of tobamovirus, dianthovirus, umbravirus, carmovirus, pomovirus, tombusvirus and hordeivirus generas have been assigned to this group (Callaway *et al.*, 2001; Lucas, 2006; Scholthof, 2005). Viruses forming the second group require both, MP and CP for trafficking. Representatives of the second group belong to potyviruses, potexviruses and cucumoviruses. A third group, represented by members of closteroviruses, nepoviruses and comoviruses, comprises viruses that also require CP because they move in the form of virions. Such classification provides an overview of the various movement strategies employed by plant viruses in regard of the CP. However, as further data about movement accumulates this type of classification is becoming oversimplified. Due to the ability of viruses to employ different non-exclusive movement strategies, in the end the requirement of viral proteins for trafficking depends on the particular virus-host pathosystem regardless of the virus taxonomy.

As an example, different strains of a virus can utilize different movement strategies in the same host. Some isolates of *Brome mosaic virus* (BMV) require CP for the movement whereas others do not (Takeda *et al.*, 2005). BMV was previously considered to move as virions since encapsidation competent CP was reported to be needed for the cell-to-cell movement (Callaway *et al.*, 2001). Takeda *et al.* (2005) demonstrated that the BMV-M1 strain required CP for movement whereas BMV-M2 did not. Mutational analysis of known BMV isolates and the isolation of spontaneous BMV mutants exhibiting CP-independent movement further demonstrated that a single amino acid residue difference in the C-terminus of MP determined the requirement of CP. *Cucumber mosaic virus* (CMV), also from the family *Bromoviridae*, is another example where the necessity of CP is determined by MP, as a C-terminal deletion in the MP renders CMV movement CP-independent (Nagano *et al.*, 2001).

There are also reports suggesting that the same virus can use alternative modes of trafficking in different hosts. In the absence of CP *Barley stripe mosaic virus* is able to systemically infect barley (*Hordeum vulgare*) but not *N. benthamiana* (Petty and Jackson, 1990). The systemic movement of

Cymbidium ringspot tombusvirus (CyRSV) has been suggested to take place either as virions or in a nonvirion form depending on the host (Dalmay *et al.*, 1992). CyRSV CP mutant unable to form virus particles infected *N. benthamiana* systemically, whereas it was confined to the inoculated leaves in *N. clevelandii*. The CP of *Tomato bushy stunt virus* (TBSV) has been reported to be dispensable for systemic infection in *N. benthamiana* and *N. clevelandii* (Russo *et al.*, 1994), but CP was indispensable in pepper (Turina *et al.*, 2003).

Over the past decade examples of cell-to-cell and long-distance movement of endogenous RNA species has been accumulating (reviewed by Kehr and Buhtz, 2008). These non-cell-autonomous RNA species include messenger RNAs, micro-RNAs and small interfering RNAs. This shows that the inter cellular transport of RNAs presumably as RNP complexes is an important host mechanism. The RNAs of viroids are also transported entirely by the host machinery as their genome does not encode any proteins. Evidence about the movement of BMV RNA independently of viral proteins has been reported (Gopinath and Kao, 2007). The RNA3 of BMV was demonstrated to move systemically without the CP or even MP, however all BMV RNAs (RNA1, RNA2, RNA3) moved more efficiently when expressed together with the MP. Currently it has not been demonstrated that any RNA virus would depend entirely on the host movement system for the trafficking of its genome.

2. AIM OF THE STUDY

The aim of the current study was to gain further knowledge about the biological functions of sobemoviral structural proteins and about the molecular biology of sobemoviruses in general.

For this purpose the following tasks were carried out:

1. Identification of the C-termini of CfMV, RYMV, SBMV and RGMoV VPgs.
2. Characterization of the RNA-linking and phosphorylation of these VPgs.
3. Determination of the subcellular localization of CfMV CP.
4. Analysis of the role of CfMV CP in movement and transmission.

3. MATERIALS AND METHODS

Materials and methods are described in detail within the corresponding sections of publications and manuscript. Standard procedures were carried out according to Sambrook and Russell (2001) or according to manufacturer's instructions. In the course of this study the following methods were used:

- Cloning and mutagenesis (Publication I, Manuscript)
- Virus inoculation (Publications II-III, Manuscript)
- RNA isolation (Publications II-III, Manuscript)
- RT-PCR (Publication I, Manuscript)
- Northern blotting (Manuscript)
- Western blotting (Manuscript)
- Expression and visualization of tagged proteins (Publication I, Manuscript)
- RNA labelling and microinjection (Publication I)
- Virus purification and VPg extraction (Publications II-III)
- Protein analysis by mass spectrometry (Publication II-III)
- Bioinformatic analysis (Publication II)

4. RESULTS AND DISCUSSION

4.1 Sobemoviral VPg processing (publications II and III)

The specificity of the sobemoviral protease has been proposed and demonstrated to be E/T,S,N (Gorbalenya *et al.*, 1988; Mäkinen *et al.*, 2000; Nair and Savithri, 2010b; van der Wilk *et al.*, 1998). Based on that many different cleavage sites can be predicted for sobemoviral polyproteins. For several sobemoviruses: CfMV, RYMV, SBMV and SeMV - the N-terminus of VPg has been mapped (Hébrard *et al.*, 2008; Mäkinen *et al.*, 2000; Nair and Savithri, 2010b; van der Wilk *et al.*, 1998) while the C-terminus of VPg has been experimentally proven only for SeMV (Nair and Savithri, 2010b). Sobemoviruses also deploy -1 PRF for the expression of the polyprotein and VPg occupies a position in the polyprotein close to the -1 PRF signal. Therefore it has been proposed that at least CfMV might express its VPg through -1 PRF mechanism and as a result even encode VPg-s with different C-termini (Mäkinen *et al.*, 2000). An approximate mass of CfMV VPg was observed on SDS-PAGE to be around 12 kDa (Mäkinen *et al.*, 2000). Based on the observed mass, the C-terminus of VPg was proposed to situate downstream of the -1 PRF signal and previously predicted processing sites. Indeed, several putative processing sites downstream of -1 PRF consensus sequence exist also in the polyproteins of RYMV, RGMoV and SBMV.

We have analyzed the polyprotein processing by identifying the VPg termini of several sobemoviruses. The VPgs, purified from virions of CfMV, RYMV, SBMV and RGMoV, and subjected to mass spectrometric analysis, revealed that the mature VPgs are 78, 79, 77 and 79 amino acid residues in length, respectively (publication II, Figure 1; publication III, Figure 1-2). The lengths of the VPgs were in accordance with the previously experimentally demonstrated length of the VPg of SeMV, which was 77 residues (Nair and Savithri, 2010b). The C-termini of VPgs of all these sobemoviruses were determined to be cleaved between E/T residues upstream of the putative and proven -1 PRF signals. This ruled out the possibility of the proposed production of VPgs with different C-termini and involvement of -1 PRF in VPg expression. In addition to determining the C-termini of these VPgs we independently corroborated the previously described N-termini of CfMV, RYMV and SBMV VPgs. The N-termini of CfMV, RYMV and SBMV VPgs were cleaved between E/N, E/S and E/T, respectively, as described earlier (Hébrard *et al.*, 2008; Mäkinen *et al.*, 2000; Nair and Savithri, 2010b).

The previously described molecular weight of 12 kDa was most likely due to a shift in mobility in SDS-PAGE caused by the acidic nature of the VPg protein (CfMV VPg predicted and experimental pI is ~4; Olspert, unpublished

results). The mobility of disordered and/or acidic proteins during denaturating electrophoresis has been previously reported to be different from the actual mass (Receveur-Bréchot *et al.*, 2006). Other factors that may influence mobility are the PTMs of VPgs that we determined (see below).

We also discovered that the previously predicted cleavage site of the N-terminus of RGMoV VPg (Accession NP_736586) was incorrect. The mature VPg of RGMoV was actually cleaved three residues upstream of the predicted site and contained additional three residues, SSE, in the N-terminus.

All experimentally proven protease cleavage sites for sobemovirus polyproteins are either E/T or E/S, with one exception, the E/N site of CfMV (Figure 2). Unfortunately, it is not known which sequence elements or other factors establish which site is actually used and which is not. As mentioned above, additional putative cleavage sites can be found in sobemoviral polyproteins. All determined cleavage sites of SBMV, some sites of SeMV and one site of CfMV are QE/T(S) (Nair and Savithri, 2010b), which indicates that residues further upstream of the actual cleavage may also be important. No recognizable pattern can be found in the vicinity of all the proven VPg cleavage sites (Figure 2). Most probably the cleavage sites are determined through complex interactions between the protease and substrate. In addition, the proteases of each species probably have different substrate recognition capacities. Disorder of the substrate may also be important for these interactions. Indeed, increased disorder near the termini of VPgs of sobemoviruses and other genera has been predicted (Hébrard *et al.*, 2009a).

Taken together we have determined and confirmed the termini and processing sites of four sobemoviral VPgs. The data provides additional evidence about the diversity of sobemoviral protease substrates. There seems to be less constraints on amino acid residues surrounding the actual cleavage sites than previously proposed, based on SeMV (Nair and Savithri, 2010b).

	N-terminal	C-terminal
CfMV	VRKFVTV**E** - **N**SELYPDQ	SDDEDTQ**E** - **T**AIRPLNL
RYMV	FKKEVVV**E** - **S**PFEIYGK	SGEDVDI**E** - **T**SHPVAPS
RGMoV	SIFHANC**E** - **S**SENGEQG	DDWDARE**E** - **S**TGNDIPL
SBMV	NYLLRSQ**E** - **T**LPPELSV	MVWSSAQ**E** - **T**VAEPLNY
SeMV	NYLLRSN**E** - **T**LPPELSI	LVWENAQ**E** - **T**VAVENLN

Figure 2. Overview of the determined VPg processing sites
The cleavage site position in relation to VPg is indicated on top. The cleavage takes place between amino acid residues indicated in bold.

4.2 Characterization of VPg-RNA linking of sobemoviruses (publications II and III)

The VPgs are covalently linked to the 5' end of viral RNA of several different virus genera. Previously it has been assumed that viruses within one genus use the same residue to link RNA. For instance, picorna-, poty- and caliciviruses use a tyrosine (Ambros and Baltimore, 1978; Anindya *et al.*, 2005; Belliot *et al.*, 2008; Murphy *et al.*, 1991; Rothberg *et al.*, 1978; Schein *et al.*, 2006) while nepo- and comoviruses are reported to exploit a serine residue (Jaegle *et al.*, 1987; Zalloua *et al.*, 1996). However, for sobemoviruses the only conserved region of the VPgs is the WAD/WGD/WNK motif followed by a D/E rich region (Mäkinen *et al.*, 2000; Mäkinen *et al.*, 1995b). This diversity made them a particularly interesting group for studying VPg RNA linkage.

We determined the residues to which CfMV, RYMV, SBMV and RGMoV RNA was linked to, using mass spectrometric analysis of virion derived VPgs. The purified VPgs were digested by trypsin and the RNA covalently linked to the peptides was degraded with acidic hydrolysis. The remaining modification after RNA hydrolysis was determined to be a 5',3'-diphosphate nucleotide, pNp (N denoting adenosine, cytidine, guanosine or uridine). The RNA linkage sites were determined by searching for all these possible modifications from all possible phosphodiester bond acceptor residues, which are serine, tyrosine or threonine. CfMV VPg was determined to be linked to RNA over a tyrosine residue at position 5 (Figure 3; publication II, Figure 2) and the modification was pGp. This was in accordance with the genomic sequence which starts with the G nucleotide. We determined that RYMV, SBMV and RGMoV VPgs were all linked to RNA through the residue at the first position of VPg (Figure 3; publication II, Figure 2; publication III, Figure 1-2). Both, RYMV and RGMoV VPgs have a serine at the first position. However, the first residue of SBMV VPg is threonine. For these three viruses the corresponding modification was pAp, which was also the first nucleotide of their reported genomes.

Altogether the results show that RNA linkage of sobemoviruses is species-specific and that sobemoviruses can use all available hydroxyl group containing amino acid residues for RNA linking. For CfMV, RYMV and SBMV there is additional experimental data supporting our results. For these three viruses the residue which was determined by us to be linked to RNA was not detected correctly by Edman sequencing (Hébrard *et al.*, 2008; Mäkinen *et al.*, 2000; van der Wilk *et al.*, 1998).

SBMV is currently the only virus in the world for which the use of threonine for linking RNA to VPg has been demonstrated. In addition to SBMV, some VPgs of other sobemoviruses also have a threonine at the putative N-terminus of VPg. These are SoMV, SCMoV, SeMV and SCPMV. The latter two are very

closely related to SBMV and their VPgs are conserved. With the exception of CfMV, for which tyrosine at position five is used for RNA linking, all sobemoviruses, for which sequence data is currently available, have either a serine or a threonine residue at the putative or proven N-terminus of VPg. Therefore, we propose that the threonine and serine residues at the first position of the VPgs of these viruses are most probably used for RNA linking.

It is interesting why the use of threonine for RNA linking has not been described earlier. Either sobemoviruses are truly unique or the VPg linking across all the diversity of viruses has not been studied intensively enough. One explanation would also be that the hydroxyl group of threonine is chemically less accessible and/or the stereo-chemical constraints of serine and tyrosine are more favorable which manifests in a more frequent use of the latter two. Therefore, it would be interesting to analyze VPg RNA linkage and other VPg properties of viruses from other genera at a larger scale.

The VPg's role in replication has been characterized for picornaviruses, where it is used as a primer for viral RNA synthesis (Steil and Barton, 2009b). In contrast, it was recently demonstrated that SeMV VPg is not required by the RdRp for the *in vitro* synthesis of the negative strand of viral RNA (Govind and Savithri, 2010). This raises a question whether sobemoviruses only use VPg as a primer for the synthesis of the positive strand RNA? However, it is also possible that the RNA synthesis initiation mechanism *in vitro* is different from the *in vivo* situation.

Altogether we have characterized the RNA linkage of four members of the *Sobemovirus* genera. The results demonstrate unprecedented diversity within one genera and warrant several conclusions on a much larger scale:

(i) the type and position of the amino acid residue involved in VPg RNA linking is not always conserved between related viruses;

(ii) all hydroxyl group containing amino acid residues can and are used for RNA linking.

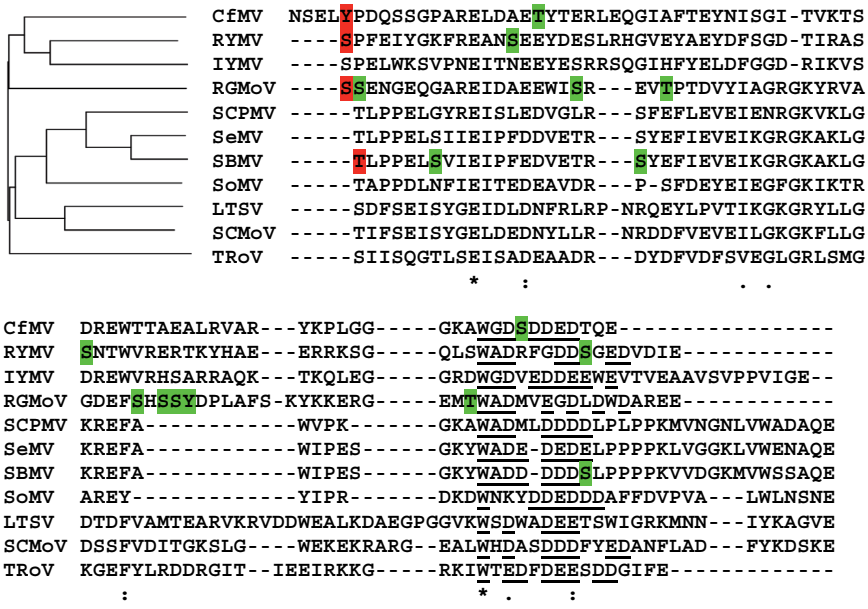


Figure 3. Sequence comparison of sobemoviruses and their VPgs
The phylogenetic tree on the left was constructed by the maximum composite likelihood mode with mid-point rooting based on the full nucleotide sequences of sobemoviruses. The VPg sequences were aligned with ClustalW. The tyrosine at position 5 of CfMV VPg, the threonine at the first position of SBMV VPg and the serines at the first positions of RYMV and RGMoV VPgs to which RNA is covalently linked to are indicated with red. The phosphorylated threonine, serine and tyrosine residues are indicated with green. The conserved motif WAD/WGD followed by multiple E/D residues is underlined. Invariant amino acid residues are indicated below the alignment by an asterisk (*), highly similar residues by a colon (:) and similar residues by a dot (.).

4.3 Sobemovirus VPgs are phosphorylated (publications II and III)

Our mass spectrometric analysis of virion purified sobemoviral VPgs demonstrated multiple phosphorylations for each protein.

CfMV VPg was found to be phosphorylated at two positions, T20 and S71 (Figure 3; publication II, Figure 2). In RYMV VPg phosphoserines were found at positions 14, 41 and 72 (Figure 3; publication II, Figure 2). With RYMV, it is possible that serines at positions 33 and 59 were also sometimes phosphorylated. The serine residues of SBMV VPg were phosphorylated at positions 7, 20 and 58 (Figure 3; publication III, Figure 1). The RGMoV VPg was detected to be phosphorylated at two serine residues at positions 2, 19, two threonine residues at positions 23, 63 and at a tyrosine residue at position 47 (Figure 3; publication III, Figure 2-3). In combination with tyrosine at position 47, one serine residue either at position 43, 45 or 46 was also phosphorylated, but the position could not be determined with certainty. For each phosphorylated position of each virus non-phosphorylated peptides were also detected, which shows that the positions are not always phosphorylated. The selection pressure expressed on the sites of the phosphorylated amino acids was analyzed for RYMV (publication II). Positions 14, 33 and 41 of RYMV VPg are conserved at the amino acid level and were determined to be under strong conservative selective pressure. Positions 59 and 72 were determined to be under neutral evolutionary selection pressure .

The comparison of phosphorylation patterns of CfMV, RYMV, SBMV and RGMoV VPgs indicates that the protein sequences as well as phosphorylation sites have little similarity. However, the data on SBMV allows us to predict the phosphorylation of the VPgs of two very closely related viruses, SeMV and SCPMV. Serines 7 and 20 are present at the same position in SeMV and serine 20 in SCPMV, which indicates that they might also be the targets of phosphorylation. The position of serine 58 is occupied in SeMV by glutamic acid and in SCPMV by aspartic acid, both chemically mimicking phosphoserine to some extent. This indicates that the negative charge at position 58 might be important for all these closely related species.

Except for S72, the VPgs of 150 RYMV isolates have no coding differences between the isolates in the phosphorylation sites at the amino acid level despite several synonymous nucleotide substitutions. In contrast, position 72 exhibits amino acid polymorphism and is neither under conservative or diversifying selection. One strain has an aspartic acid at position 72. The fitness of this strain might again be explained by the physico-chemical similarities of aspartic acid and phosphorylated serine.

For PVA the phosphorylation and disordered nature of VPg has been demonstrated (Hafrén and Mäkinen, 2008; Puustinen *et al.*, 2002; Rantalainen *et al.*, 2008). PVA VPg phosphorylation is believed to be involved in regulation of host interactions, whereas the disorder is believed to be the basis

of multifunctionality. Presumably the structural flexibility provides an ensemble of possible conformations and possible interactions. It has been experimentally demonstrated that SeMV and RYMV VPg are also disordered proteins and for other sobemoviral VPgs disordered nature has been predicted (Hébrard *et al.*, 2009b; Satheshkumar *et al.*, 2005a). However, upon secondary structure stabilization the RYMV VPg has the propensity to form structures (Hébrard *et al.*, 2009b). Stabilization and/or structure formation can occur during an interaction, which again may be dependent on phosphorylation. The regulation of folding/unfolding and interaction determination of disordered proteins by phosphorylation has been reported before (Mittag *et al.*, 2010; Stein *et al.*, 2009; Wright and Dyson, 2009). The N-termini of sobemoviral CPs are rich in positively charged amino acid residues (Tamm and Truve, 2000b) and reside inside the virion presumably in contact with the negatively charged RNA and VPg (Abad-Zapatero *et al.*, 1980; Plevka *et al.*, 2007; Rossmann *et al.*, 1983). Phosphorylation of VPg might be needed for providing additional stability to the virion through electrostatic interactions with CP. This does not rule out the possibility that phosphorylation is needed for other possible functions of VPg. For instance, the RdRp of SeMV failed to nucleotidylate *in vitro* VPg purified from bacteria where presumably phosphorylation does not occur similarly to plant cells (Govind and Savithri, 2010). One explanation would be that the nucleotidylation reaction is dependent on (host) factors missing in the *in vitro* assay. However, VPg phosphorylation could also regulate RdRp in this process.

Our results suggest that VPg phosphorylation is common at least among sobemoviruses. The conservation and physicochemical mimicry of the phosphorylation sites within one species (RYMV) and between closely related species (SBMV, SeMV and SCPMV) implies that sobemoviral VPg phosphorylation serves a biological role.

4.4 Subcellular localization of CfMV CP (publication I and manuscript)

The primary sequences of sobemovirus CPs have a moderate sequence similarity in the range of 12 to 30% between distantly related viruses. However, the CPs are structurally very similar (Plevka *et al.*, 2007). The N-terminal part of all sobemoviruses CPs is rich in basic amino acids and contains the ARM. The ARM of SeMV CP has been implicated being important for RNA encapsidation (Satheshkumar *et al.*, 2005b). *In vitro* experiments with CfMV (Tamm and Truve, 2000a) and SCPMV (Lee and Hacker, 2001) CPs have demonstrated that both proteins bind RNA in a sequence non-specific manner. The amino acid residues responsible for SCPMV CP RNA binding *in vitro* were mapped to the

ARM region (Lee and Hacker, 2001). Sobemovirus particles have been found in the nuclei of infected cells (Chamberlain and Catherall, 1976; Fuentes and Hamilton, 1993; Mohamed and Mossop, 1981; Rabenstein and Stanarius, 1984). Based on sequence similarity it has been proposed that the N-terminal region including the ARM of sobemovirus CPs contain an NLS (Mäkinen *et al.*, 1995b; Yassi *et al.*, 1994).

The experiments in plant cells with CP and EGFP fusions demonstrated that the CfMV CP indeed localized to the nucleus (publication I, Figure 1-2). The sequence directing nuclear import was determined to be within the first 33 N-terminal amino acid residues, which also contain the ARM. However, the ARM was not the only motif found to be responsible for the nuclear localization of CP. Our experiments demonstrated that in plant cells two separate regions were used for CfMV CP nuclear transport, a strong NLS1 within residues 22-33 (the ARM) and a weaker NLS2 within residues 1-22 (publication I, Figure 1). Mutating even the majority of the basic amino acid residues within the N-terminus was not sufficient to completely stop nuclear import.

Individually the full length CP fused to EGFP localized almost exclusively to the nucleus. Results obtained with infectious viruses expressing the same CP-EGFP fusion protein instead of CP, demonstrated that when CP-EGFP was expressed together with the rest of the virus-encoded proteins the fluorescence did not accumulate exclusively into the nucleus, but remained evenly distributed between cytoplasm and nucleus (manuscript). Most probably the CP was interacting with other viral and/or host proteins or with the full-length genomic RNA and therefore was not accumulating solely into the nucleus any more.

As CfMV is known to bind RNA, and was demonstrated to be transported to the nucleus, we investigated whether CP in conjunction with RNA is transported to nucleus. Microinjection studies demonstrated that the viral CP was able to transport labeled RNA to the nucleus of onion epidermal cells in sequence non-specific manner (publication I, Figure 4). This suggests also that the RNA did not interfere with importin binding.

Although positive-strand ssRNA viruses are generally believed to be mainly cytoplasmic, some of their proteins have been reported to localize to the nucleus (Hajimorad *et al.*, 1996; Lucy *et al.*, 2000; Rebelo *et al.*, 2008; Ryabov *et al.*, 2004; Schaad *et al.*, 1996). For potato leafroll virus the nuclear import of CP has also been reported (Haupt *et al.*, 2005). Thus, the nucleus has been shown to be involved in the infection of some positive-strand ssRNA viruses. For instance, 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; Rajamäki and Valkonen, 2009). The CP of CfMV may also have functions that require its transport to the nucleus. The nucleus may also be used for reducing CP concentration in the cytoplasm after initial entry and virion disassociation which may be necessary

for the regulation of virus replication.

Our results suggest that the virus particles found in the nucleus are probably also assembled there, since the N-terminus of the CP with NLS is buried within the particle (Tars *et al.*, 2003) and therefore is not accessible after virion formation. These particles may also contain viral RNA transported to the nucleus together with CP.

4.5 The CP of CfMV is dispensable for virus movement (manuscript)

The P1 proteins of sobemoviruses have been proposed as putative MPs of sobemoviruses and for RYMV, SCPMV and CfMV the P1 is indeed dispensable for replication but needed for the virus movement (Bonneau *et al.*, 1998; Meier *et al.*, 2006; Sivakumaran *et al.*, 1998). The CPs of RYMV and SCPMV are also needed for virus cell-to-cell movement but not for virus replication (Brugidou *et al.*, 1995; Sivakumaran *et al.*, 1998). In addition, the long distance movement of SCPMV and RYMV has been proposed to be dependent on particle formation (Fuentes and Hamilton, 1993; Opalka *et al.*, 1998). Recently it was shown that SeMV P1 interacts with VPg, P10 and CP (Chowdhury and Savithri, 2011; Roy Chowdhury and Savithri, 2011). Therefore, Chowdry and Savithri (2010) proposed that SeMV CP is involved in movement and that SeMV presumably moves as virions.

We analyzed the role of CfMV CP and CP arginine-rich region in viral infection cycle. The infectivity and mechanical transmission was analyzed for two different ARM mutants, R5X, with five arginines replaced by non-basic amino acid residues, and R3L, with three out of five arginines replaced by leucines (manuscript, Figure 1). In addition, a virus with the complete CP deletion was constructed. The analysis revealed that all three mutant viruses were able to replicate and accumulate in the inoculated leaves (manuscript, Figure 2). Surprisingly, all three mutants were also able to infect the host plant systemically. However, mutant R5X accumulated at lower levels in both inoculated and systemic leaves. As the mutants were infectious, it was also tested whether the mutant viruses were transmissible by mechanical inoculation. Oat (*Avena sativa*) plants were inoculated with the sap produced from the upper leaves of plants infected with the mutants and analyzed for CfMV infection (manuscript, Figure 2). Mechanical transmission of the arginine rich region mutants was successful, although in this experiment R5X failed to produce systemic infection in mechanically inoculated plants. In contrast, we were unable to detect CfMV in plants mechanically inoculated with the CP deletion mutant.

To test whether CP dispensability for systemic movement was host specific, the CP deletion mutant was also analyzed in wheat (*Triticum aestivum*) and

barley (*Hordeum vulgare*) (manuscript, Figure 3). In both hosts the deletion mutant was able to infect the wheat and barley plants systemically. This demonstrated that CfMV uses the same CP independent mechanism for movement in the majority of its known experimental hosts.

CfMV movement was further characterized with EGFP expressing CfMV. The first construct was an EGFP fusion to the C-terminus of the full length CP (CP-EGFP) (manuscript, Figure 4). Oat plants were biolistically inoculated and virus movement was analyzed by monitoring EGFP fluorescence in inoculated tissues. Two days post inoculation (dpi) the fluorescence was visible mainly in single epidermal cells and in small foci of mesophyll cells (manuscript, Figure 5). After three dpi the EGFP was detected in the neighboring epidermal cells and the infection foci in the mesophyll had expanded, which demonstrated cell-to-cell movement of the virus. From the fourth to the sixth dpi, the area of infected mesophyll cells continued to grow and presumably reached the vasculature. By that time the infected area had reached its maximum and fluorescence started to fade due to the spreading necrosis caused by the initial biolistic inoculation.

In the case of CPdelta-EGFP variant, where EGFP was inserted instead of the CP C-terminus (manuscript, Figure 4), the fluorescence was considerably weaker and strongest around the third to the fourth dpi (manuscript, Figure 5). Weak fluorescence was detected in single epidermal cells after two dpi, which sometimes spread to neighboring epidermal cells the following days. EGFP was never detected in mesophyll cells. The signal from epidermal cells usually disappeared around the fifth day.

Unfortunately no distinguishable EGFP signal was detected in the upper leaves with either of the EGFP containing viruses. However, some plants infected with CP-EGFP and CPdelta-EGFP developed CfMV infection symptoms in the upper leaves. Viral RNA was indeed detected from these leaves by RT-PCR, but the initially EGFP containing viruses had lost the EGFP sequence as well as different portions of CP/RdRp encoding sequence. We detected two different naturally occurring deletion mutant subsets (manuscript, Figure 6). In the first group majority of CP coding sequence after the overlap with RdRp gene was lost together with EGFP sequence, while the 3' UTR was fully retained. These recombinants were detected from plants inoculated with either CP-EGFP or CPdelta-EGFP. The second group comprised of viruses where only a small region coding the C-terminal part of CP and EGFP sequence were missing. These mutants were logically only detected from plants inoculated with CP-EGFP.

Previously it has been documented that sobemoviruses require P1 as well as CP for systemic movement (Brugidou *et al.*, 1995; Fuentes and Hamilton, 1993; Opalka *et al.*, 1998; Sivakumaran *et al.*, 1998). A CP deletion mutant of RYMV failed to infect rice plants systemically, but accumulated in the inoculated leaves, which was an indication of cell-to-cell movement (Brugidou *et al.*, 1995). A

SCPMV CP initiation codon mutant was undetectable even in inoculated leaves (Sivakumaran *et al.*, 1998). It is known that different strains of a virus can utilize different movement strategies in the same host and the same strain can use alternative modes of trafficking in different host (Takeda *et al.*, 2005; Nagano *et al.*, 2001). Our results with mutant viruses established that CfMV CP is dispensable for cell-to-cell as well as systemic movement in different host plants. This was further confirmed by the detection of recombinant viruses, lacking majority of the CP cistron, from plants infected with CP-EGFP and CPdelta-EGFP. Although the CP may be indispensable for CfMV movement in untested hosts, it seems that CfMV in general utilizes a different movement strategy from SCPMV and RYMV.

The CP of TRoV, another sobemovirus, has been reported to facilitate long distance movement of red clover necrotic mosaic dianthovirus (Callaway *et al.*, 2004). Similarly, it has been reported that CfMV CP can complement cell-to-cell movement of PVX CP mutants (Fedorkin *et al.*, 2001). Together with the results from SCPMV and RYMV CP mutants, it seems that at least some sobemoviral CPs have clear movement related functions. As CfMV CP was demonstrated to have no RNA silencing suppression activity (manuscript, Figure 7), the enhancement of PVX movement must involve some other function of the CP. It is also possible that CfMV CP contributes to virus spread for instance by enhancement of accumulation and/or providing further stability to viral RNP complexes. This is supported by the fact that RYMV accumulated at higher levels in plants expressing RYMV CP than in control plants, indicating that additional CP further enhanced virus infection (Kouassi *et al.*, 2006).

The viral RNA genome is usually transported from cell-to-cell and systemically through vasculature either as virions or in some other form of RNP complex. CfMV trafficking as virions can now be ruled out at least in hosts tested by us. Previously we have reported that P1 is indispensable for virus movement and accumulation in oats (Meier *et al.*, 2006). As CfMV P1 is the viral suppressor of RNA silencing (Sarmiento *et al.*, 2007), it still remains unclear whether P1 facilitates virus spread as a suppressor or as a member of the movement RNP complex.

Interestingly, we were unable to transmit the CP deletion mutant mechanically, whereas R3L and R5X were easily transmissible. This demonstrates that CP might provide additional stability to the viral RNA which is needed for an efficient transmission and that mutations in the arginine-rich region do not affect this putative viral RNA-CP complex involved in transmission. The formation of a stable virion is probably the most efficient way to ensure transmission but presumably also other RNP complexes can facilitate transmission. However, it is also possible that the CP contributes to transmission by higher accumulation, but compared to the R3L and R5X mutants, which were transmissible, the CP deletion mutants did not have a significantly lower viral

RNA load in systemic leaves.

Mutant R5X failed to produce systemic infection in mechanically inoculated plants. However, R5X has also a mutation in the RdRp and the plants inoculated with R5X usually appeared to have a lower viral load, presumably due to this mutation. Thus, we suppose that the movement of R5X was normal as transmission of the virus was also successful and the lack of systemic infection was due to defects in replication and accumulation.

In the case of CP-EGFP, the EGFP sequence was longer retained within the genome, as progeny viruses replicating in the mesophyll still expressed EGFP, whereas for CPdelta-EGFP the EGFP gene was already lost when the virus entered mesophyll cells. CfMV, like many other RNA viruses, is able to rapidly dismiss portions of its genome through recombination. It seems that when available the virus keeps as much of the CP sequence as possible. Foreign sequence seems to be the trigger for recombination because we have not encountered recombinations in plants infected with the wt virus or CP mutants.

It is fascinating that the mutants isolated from non-inoculated upper leaves, which had lost the EGFP and CP coding sequence, all contained almost the entire 3'UTR. Not all of the 3' UTR of CfMV is needed for replication, translation and accumulation in oats (Olspert, unpublished data). This leads us to speculate that the 3' UTR might contain sequences or structural elements important for transport or that there simply is a hot-spot for recombination in the beginning of the 3' UTR. These two hypotheses are not mutually exclusive.

Taken together the results clearly demonstrate that the CP is not strictly necessary neither for cell-to-cell nor long-distance movement of CfMV. In regard to other reports it seems that sobemoviruses deploy different trafficking strategies in regard of the requirement of CP. Whatever role the CP of CfMV might have in the nucleus it is not required for virus movement. The ARM of CfMV CP is dispensable for the CPs role in transmission.

CONCLUSIONS

1. Sobemoviral VPgs are cleaved from the polyprotein upstream of the -1 programmed ribosomal frameshift site.
2. Sobemoviruses use all possible hydroxyl group containing amino acid residues for linking VPg to RNA.
3. Sobemoviral VPgs are phosphoproteins.
4. The N-terminus of CfMV CP contains two nuclear localization signals.
5. The CP of CfMV can function as an RNA transporter.
6. The CP of CfMV is dispensable for cell-to-cell and systemic movement in oats, barley and wheat.
7. The CP of CfMV is involved in virus accumulation.
8. The CP of CfMV is required for efficient transmission.

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, 33-39.

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, 5263-5266.

Ambros, V., Pettersson, R.F. and Baltimore, D. (1978). An enzymatic activity in uninfected cells that cleaves the linkage between poliovirion RNA and the 5' terminal protein. Cell 15, 1439-1446.

Anindya, R., Chittori, S. and Savithri, H.S. (2005). Tyrosine 66 of *Pepper vein banding virus* genome-linked protein is uridylylated by RNA-dependent RNA polymerase. Virology 336, 154-162.

Belliot, G., Sosnovtsev, S.V., Chang, K.O., McPhie, P. and Green, K.Y. (2008). Nucleotidylylation of the VPg protein of a human norovirus by its proteinase-polymerase precursor protein. Virology 374, 33-49.

Benitez-Alfonso, Y., Faulkner, C., Ritzenthaler, C. and Maule, A.J. (2010). Plasmodesmata: gateways to local and systemic virus infection. Mol Plant Microbe Interact 23, 1403-1412.

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 Å resolution. Structure 3, 1021-1030.

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, 79-86.

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, 108-115.

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, 98-108.

Callaway, A., Giesman-Cookmeyer, D., Gillock, E.T., Sit, T.L. and Lommel, S.A. (2001). The multifunctional capsid proteins of plant RNA viruses. Annu Rev Phytopathol 39, 419-460.

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.

Carette, J.E., Kujawa, A., Gühl, K., Verver, J., Wellink, J. and Van Kammen, A. (2001). Mutational analysis of the genome-linked protein of cowpea mosaic virus. Virology 290, 21-29.

Chamberlain, J.A. and Catherall, P.L. (1976). Electron microscopy of some grasses and cereals infected with cocksfoot mottle, phleum mottle and cocksfoot mild mosaic viruses. Journal of General Virology 30, 41-50.

Chaudhry, Y., Nayak, A., Bordeleau, M., Tanaka, J., Pelletier, J., Belsham, G.J., Roberts, L.O. and Goodfellow, I.G. (2006). Caliciviruses differ in their functional requirements for eIF4F components. J Biol Chem 281, 25315-25325.

Chowdhury, S.R. and Savithri, H.S. (2011). Interaction of *Sesbania mosaic virus* movement protein with the coat protein - implications for viral spread. FEBS J 278, 257-272.

Dalmay, T., Rubino, L., Burgyán, J. and Russo, M. (1992). Replication and movement of a coat protein mutant of cymbidium ringspot tombusvirus. Mol Plant Microbe Interact 5, 379-383.

Daughenbaugh, K.F., Fraser, C.S., Hershey, J.W.B. 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, 2852-2859.

Daughenbaugh, K.F., Wobus, C.E. and Hardy, M.E. (2006). VPg of murine norovirus binds translation initiation factors in infected cells. Virol J 3, 33.

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, 128-136.

Fedorkin, O., Solovyev, A., Yelina, N., Zamyatnin, A.J., Zinovkin, R., Mäkinen, K., Schiemann, J. and Yu Morozov, S. (2001). Cell-to-cell movement of potato virus X involves distinct functions of the coat protein. J Gen Virol 82, 449-458.

Ferrer-Orta, C., Arias, A., Agudo, R., Pérez-Luque, R., Escarmís, C., Domingo, E. and Verdaguer, N. (2006). The structure of a protein primer-polymerase complex in the initiation of genome replication. EMBO J 25, 880-888.

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), 1903-1910.

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, 440-451.

Ghosh, A., Rutgers, T., Ke-Qiang, M. and Kaesberg, P. (1981). Characterization of the coat protein mRNA of southern bean mosaic virus and its relationship to the genomic RNA. J Virol 39, 87-92.

Goodfellow, I., Chaudhry, Y., Gioldasi, I., Gerondopoulos, A., Natoni, A., Labrie, L., Laliberté, J. and Roberts, L. (2005). Calicivirus translation initiation requires an interaction between VPg and eIF 4 E. EMBO Rep 6, 968-972.

Gopinath, K. and Kao, C.C. (2007). Replication-independent long-distance trafficking by viral RNAs in *Nicotiana benthamiana*. Plant Cell 19, 1179-1191.

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, 287-290.

Govind, K. and Savithri, H.S. (2010). Primer-independent initiation of RNA synthesis by SeMV recombinant RNA-dependent RNA polymerase. Virology 401, 280-292.

Gruez, A., Selisko, B., Roberts, M., Bricogne, G., Bussetta, C., Jabafi, I., Coutard, B., De Palma, A.M., Neyts, J. and Canard, B. (2008). The crystal structure of coxsackievirus B3 RNA-dependent RNA polymerase in complex with its protein primer VPg confirms the existence of a second VPg binding site on *Picornaviridae* polymerases. J Virol 82, 9577-9590.

Grzela, R., Szolajska, E., Ebel, C., Madern, D., Favier, A., Wojtal, I., Zagorski, W. and Chroboczek, J. (2008). Virulence factor of potato virus Y, genome-attached terminal protein VPg, is a highly disordered protein. J Biol Chem 283, 213-221.

Hacker, D.L. (1995). Identification of a coat protein binding site on southern bean mosaic virus RNA. Virology 207, 562-565.

Hacker, D.L. and Fowler, B.C. (2000). Complementation of the host range restriction of southern cowpea mosaic virus in bean by southern bean mosaic virus. Virology 266, 140-149.

Hacker, D.L. and Sivakumaran, K. (1997). Mapping and expression of southern bean mosaic virus genomic and subgenomic RNAs. Virology 234, 317-327.

Hafrén, A. and Mäkinen, K. (2008). Purification of viral genome-linked protein VPg from potato virus A-infected plants reveals several post-translationally modified forms of the protein. J Gen Virol 89, 1509-1518.

Hajimorad, M.R., Ding, X.S., Flasinski, S., Mahajan, S., Graff, E., Haldman-Cahill, R., Carrington, J.C. and Cassidy, B.G. (1996). Nla and Nlb of peanut stripe potyvirus are present in the nucleus of infected cells, but do not form inclusions. Virology 224, 368-379.

Han, K.R., Choi, Y., Min, B.S., Jeong, H., Cheon, D., Kim, J., Jee, Y., Shin, S. and Yang, J.M. (2010). Murine norovirus-1 3Dpol exhibits RNA-dependent RNA polymerase activity and nucleotidylylates on Tyr of the VPg. J Gen Virol 91, 1713-1722.

Hartmann, J.X., Bath, J.E. and Hooper, G.R. (1973). Electron microscopy of viruslike particles from shoestring-diseased highbush blueberry, *Vaccinium corymbosum* L. Phytopathology 63, 432-436.

Haupt, S., Stroganova, T., Ryabov, E., Kim, S.H., Fraser, G., Duncan, G., Mayo, M.A., Barker, H. and Taliansky, M. (2005). Nucleolar localization of potato leafroll virus capsid proteins. J Gen Virol 86, 2891-2896.

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

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

Hébrard, E., Pinel-Galzi, A. and Fargette, D. (2008). Virulence domain of the RYMV genome-linked viral protein VPg towards rice rymv1-2-mediated resistance. Arch Virol 153, 1161-1164.

Hébrard, E., Pinel-Galzi, A., Bersoult, A., Siré, 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, 1369-1373.

Hébrard, E., Poulicard, N., Gérard, C., Traoré, O., Wu, H., Albar, L., Fargette, D., Bessin, Y. and Vignols, F. (2010). Direct interaction between the *Rice yellow mottle virus* (RYMV) VPg and the central domain of the rice eIF(iso)4G1 factor correlates with rice susceptibility and RYMV virulence. Mol Plant Microbe Interact 23, 1506-1513.

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.

Hull, R. (2002). Matthew's plant virology. Academic Press, San Diego, USA.

Hull, R. and Fargette, D. (2005). Sobemovirus, In Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses, C. Fauquet, M.A. Mayo, J. Maniloff, U. Desselberger and L.A. Ball, ed. (Elsevier Academic Press), pp. 885-890.

Jaegle, M., Wellink, J. and Goldbach, R. (1987). The genome-linked protein of *Cowpea mosaic virus* is bound to the 5' terminus of virus RNA by phosphodiester linkage to serine. J. gen Virol. 68, 627-632.

Karetnikov, A. and Lehto, K. (2007). The RNA2 5' leader of *Blackcurrant reversion virus* mediates efficient *in vivo* translation through an internal ribosomal entry site mechanism. J Gen Virol 88, 286-297.

Kehr, J. and Buhtz, A. (2008). Long distance transport and movement of RNA through the phloem. J Exp Bot 59, 85-92.

Khan, M.A., Miyoshi, H., Gallie, D.R. and Goss, D.J. (2008). Potyvirus genome-linked protein, VPg, directly affects wheat germ *in vitro* translation: interactions with translation initiation factors eIF4F and eIFiso4F. J Biol Chem 283, 1340-1349.

Kim, S.H., MacFarlane, S., Kalinina, N.O., Rakitina, D.V., Ryabov, E.V., Gillespie, T., Haupt, S., Brown, J.W.S. 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, 11115-11120.

Kim, S.H., Ryabov, E.V., Kalinina, N.O., Rakitina, D.V., Gillespie, T., MacFarlane, S., Haupt, S., Brown, J.W.S. and Taliansky, M. (2007b). Cajal bodies and the nucleolus are required for a plant virus systemic infection. EMBO J 26, 2169-2179.

Koonin, E.V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. J Gen Virol 72 (Pt 9), 2197-2206.

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, 375-430.

Kouassi, N.K., Chen, L., Siré, C., Bangratz-Reyser, M., Beachy, R.N., Fauquet, C.M. and Brugidou, C. (2006). Expression of rice yellow mottle virus coat protein enhances virus infection in transgenic plants. Arch Virol 151, 2111-2122.

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, 371-382.

Lecoq, H., Dafalla, G., Delecolle, B., Wipf-Scheibel, C. and Desbiez, C. (2011). Snake melon asteroid mosaic virus, a tentative new member of the genus *Sobemovirus* infecting cucurbits. Plant Disease 95, 153-157.

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, 317-327.

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, 299-310.

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, 209-223.

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

Lombardo, G., Simonetta, M. and Bassi, M. (1971). An ultrastructural study of the localization of *Sowbane mosaic virus*. Caryologia 24, 323-330.

Lucas, W.J. (2006). Plant viral movement proteins: agents for cell-to-cell trafficking of viral genomes. Virology 344, 169-184.

Lucchesi, J., Mäkeläinen, K., Merits, A., Tamm, T. and Mäkinen, K. (2000). Regulation of -1 ribosomal frameshifting directed by cocksfoot mottle sobemovirus genome. Eur J Biochem 267, 3523-3529.

Lucy, A.P., Guo, H.S., Li, W.X. and Ding, S.W. (2000). Suppression of post-transcriptional gene silencing by a plant viral protein localized in the nucleus. EMBO J 19, 1672-1680.

Machín, A., Martín Alonso, J.M. and Parra, F. (2001). Identification of the amino acid residue involved in rabbit hemorrhagic disease virus VPg uridylylation. J Biol Chem 276, 27787-27792.

Mäkeläinen, K. (2006). Lost in translation: translation mechanisms inproduction of cocksfoot mottle virus proteins. PHD Thesis, University of Helsinki, Finland.

Mäkeläinen, K. and Mäkinen, K. (2005). Factors affecting translation at the programmed -1 ribosomal frameshifting site of *Cocksfoot mottle virus* RNA *in vivo*. Nucleic Acids Res 33, 2239-2247.

Mäkinen, K., Mäkeläinen, 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, 2783-2789.

Mäkinen, K., Naess, V., Tamm, T., Truve, E., Aaspõllu, A. and Saarma, M. (1995a). The putative replicase of the cocksfoot mottle sobemovirus is translated as a part of the polyprotein by -1 ribosomal frameshift. Virology 207, 566-571.

Mäkinen, K., Tamm, T., Naess, V., Truve, E., Puurand, U., Munthe, T. and Saarma, M. (1995b). Characterization of cocksfoot mottle sobemovirus genomic RNA and sequence comparison with related viruses. J Gen Virol 76, 2817-2825.

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, 264-274.

McGavin, W.J. and MacFarlane, S.A. (2009). Rubus chlorotic mottle virus, a new sobemovirus infecting raspberry and bramble. Virus Res 139, 10-13.

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

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, 321-326.

Michon, T., Estevez, Y., Walter, J., German-Retana, S. and Le Gall, O. (2006). The potyviral virus genome-linked protein VPg forms a ternary complex with the eukaryotic initiation factors eIF4E and eIF4G and reduces eIF4E affinity for a mRNA cap analogue. FEBS J 273, 1312-1322.

Mitra, T., Sosnovtsev, S.V. and Green, K.Y. (2004). Mutagenesis of tyrosine 24 in the VPg protein is lethal for feline calicivirus. J Virol 78, 4931-4935.

Mittag, T., Kay, L.E. and Forman-Kay, J.D. (2010). Protein dynamics and conformational disorder in molecular recognition. J Mol Recognit 23, 105-116.

Miyoshi, H., Okade, H., Muto, S., Suehiro, N., Nakashima, H., Tomoo, K. and Natsuaki, T. (2008). *Turnip mosaic virus* VPg interacts with *Arabidopsis thaliana* eIF(iso)4E and inhibits *in vitro* translation. Biochimie 90, 1427-1434.

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, 329-340.

Mohamed, N. and Mossop, D. (1981). *Cynosurus* and *Cocksfoot mottle viruses*: a comparison. J Gen Virol 55, 63-74.

Morales, E.J., Castano, M., Arroyave, J.A., Ospina, M.D. and Calvert, L.A. (1995). A sobemovirus hindering the utilization of *Calopogonium mucunoides* as a forage legume in the lowland tropics. Plant Disease 79, 1220-1224.

Murphy, J.F., Klein, P.G., Hunt, A.G. and Shaw, J.G. (1996). Replacement of the tyrosine residue that links a potyviral VPg to the viral RNA is lethal. Virology 220, 535-538.

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, 511-513.

Nagano, H., Mise, K., Furusawa, I. and Okuno, T. (2001). Conversion in the requirement of coat protein in cell-to-cell movement mediated by the cucumber mosaic virus movement protein. J Virol 75, 8045-8053.

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, 571-576.

Nair, S. and Savithri, H.S. (2010b). Processing of SeMV polyproteins revisited. Virology 396, 106-117.

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, 83-90.

Nomoto, A., Kitamura, N., Golini, F. and Wimmer, E. (1977). The 5'-terminal structures of poliovirion RNA and poliovirus mRNA differ only in the genome-linked protein VPg. Proc Natl Acad Sci U S A 74, 5345-5349.

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, 3323-3328.

Osman, T.A.M., Coutts, R.H.A. and Buck, K.W. (2006). *In vitro* synthesis of minus-strand RNA by an isolated cereal yellow dwarf virus RNA-dependent RNA polymerase requires VPg and a stem-loop structure at the 3' end of the virus RNA. J Virol 80, 10743-10751.

Otsus, M., Uffert, G., Sõmera, M., Paves, H., Olspert, A. and Truve, E. (manuscript). *Cocksfoot mottle virus* location during different infection stages in oats.

Pappachan, A., Chinnathambi, S., Satheshkumar, P.S., Savithri, H.S. and Murthy, M.R.N. (2009). A single point mutation disrupts the capsid assembly in *Sesbania mosaic virus* resulting in a stable isolated dimer. Virology 392, 215-221.

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 (Δ48-59) of *Sesbania mosaic virus*. Virology 375, 190-196.

Petty, I.T. and Jackson, A.O. (1990). Mutational analysis of barley stripe mosaic virus RNA beta. Virology 179, 712-718.

Pinel-Galzi, A., Rakotomalala, M., Sangu, E., Sorho, F., Kanyeka, Z., Traoré, O., Sérémé, D., Poulicard, N., Rabenantoandro, Y., Séré, Y., *et al.* (2007). Theme and variations in the evolutionary pathways to virulence of an RNA plant virus species. PLoS Pathog 3, e180.

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, 364-374.

Poulicard, N., Pinel-Galzi, A., Hebrard, E. and Fargette, D. (2010). Why *Rice yellow mottle virus*, a rapidly evolving RNA plant virus, is not efficient at breaking rymv1-2 resistance. Mol Plant Pathol 11, 145-154.

Puustinen, P. and Mäkinen, K. (2004). Uridylylation of the potyvirus VPg by viral replicase NIb correlates with the nucleotide binding capacity of VPg. J Biol Chem 279, 38103-38110.

Puustinen, P., Rajamäki, M., Ivanov, K.I., Valkonen, J.P.T. and Mäkinen, K. (2002). Detection of the potyviral genome-linked protein VPg in virions and its phosphorylation by host kinases. J Virol 76, 12703-12711.

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, 1095-1103.

Rabenstein, F. and Stanarius, A. (1984). Untersuchungen zum Knaulgrasscheckungs-Virus (cocksfoot mottle virus). Arch Phytopathol Pflanzenschutz 20, 15-31.

Rajamäki, M. and Valkonen, J.P.T. (2009). Control of nuclear and nucleolar localization of nuclear inclusion protein a of picorna-like *Potato virus A* in *Nicotiana* species. Plant Cell 21, 2485-2502.

Rantalainen, K.I., Uversky, V.N., Permi, P., Kalkkinen, N., Dunker, A.K. and Mäkinen, K. (2008). Potato virus A genome-linked protein VPg is an intrinsically disordered molten globule-like protein with a hydrophobic core. Virology 377, 280-288.

Rebelo, A.R., Niewiadomski, S., Prosser, S.W., Krell, P. and Meng, B. (2008). Subcellular localization of the triple gene block proteins encoded by a *Foveavirus* infecting grapevines. Virus Res 138, 57-69.

Receveur-Bréchot, V., Bourhis, J., Uversky, V.N., Canard, B. and Longhi, S. (2006). Assessing protein disorder and induced folding. Proteins 62, 24-45.

Rohayem, J., Robel, I., Jäger, K., Scheffler, U. and Rudolph, W. (2006). Protein-primed and *de novo* initiation of RNA synthesis by norovirus 3Dpol. J Virol 80, 7060-7069.

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

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, 4868-4872.

Roy Chowdhury, S. and Savithri, H.S. (2011). Interaction of *Sesbania mosaic virus* movement protein with VPg and P10: implication to specificity of genome recognition. PLoS One 6, e15609.

Russo, M., Burgyan, J. and Martelli, G.P. (1994). Molecular biology of tombusviridae. Adv Virus Res 44, 381-428.

Rutgers, T., Salerno-Rife, T. and Kaesberg, P. (1980). Messenger RNA for the coat protein of southern bean mosaic virus. Virology 104, 506-509.

Ryabov, E.V., Kim, S.H. and Taliansky, M. (2004). Identification of a nuclear localization signal and nuclear export signal of the umbraviral long-distance RNA movement protein. J Gen Virol 85, 1329-1333.

Sadowy, E., Milner, M. and Haenni, A.L. (2001). Proteins attached to viral genomes are multifunctional. Adv Virus Res 57, 185-262.

Salas, M. (1991). Protein-priming of DNA replication. Annu Rev Biochem 60, 39-71.

Sambrook, J., Russell, D. (2001). Molecular Cloning: A Laboratory Manual, 3rd edition. CSHL Press.

Sangita, V., Lokesh, G.L., Satheshkumar, P.S., Vijay, C.S., Saravanan, V., Savithri, H.S. and Murthy, M.R. (2004). T = 1 capsid structures of *Sesbania mosaic virus* coat protein mutants: determinants of T = 3 and T = 1 capsid assembly. J Mol Biol 342, 987-999.

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, 95-99.

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, 30291-30300.

Satheshkumar, P.S., Lokesh, G.L. and Savithri, H.S. (2004a). Polyprotein processing: cis and trans proteolytic activities of *Sesbania mosaic virus* serine protease. Virology 318, 429-438.

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, 447-458.

Satheshkumar, P.S., Lokesh, G.L., Sangita, V., Saravanan, V., Vijay, C.S., Murthy, M.R. and Savithri, H.S. (2004b). Role of metal ion-mediated interactions in the assembly and stability of *Sesbania mosaic virus* T = 3 and T = 1 capsids. J Mol Biol 342, 1001-1014.

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, 328-335.

Schaad, M.C., Haldeman-Cahill, R., Cronin, S. and Carrington, J.C. (1996). Analysis of the VPg-proteinase (NIa) encoded by tobacco etch potyvirus: effects of mutations on subcellular transport, proteolytic processing, and genome amplification. J Virol 70, 7039-7048.

Schein, C.H., Oezguen, N., Volk, D.E., Garimella, R., Paul, A. and Braun, W. (2006). NMR structure of the viral peptide linked to the genome (VPg) of poliovirus. Peptides 27, 1676-1684.

Schneider, I.R. and Worley, J.F. (1959). Upward and downward transport of infectious particles of southern bean mosaic virus through steamed portions of bean stems. Virology 8, 230-242.

Scholthof, H.B. (2005). Plant virus transport: motions of functional equivalence. Trends Plant Sci 10, 376-382.

Shields, S.A., Brisco, M.J., Wilson, T.M. and Hull, R. (1989). Southern bean mosaic virus RNA remains associated with swollen virions during translation in wheat germ cell-free extracts. Virology 171, 602-606.

Siré, 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. and Hacker, D.L. (1998). The 105-kDa polyprotein of southern bean mosaic virus is translated by scanning ribosomes. Virology 246, 34-44.

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, 376-386.

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

Steil, B.P. and Barton, D.J. (2009a). Cis-active RNA elements (CREs) and picornavirus RNA replication. Virus Res 139, 240-252.

Steil, B.P. and Barton, D.J. (2009b). Conversion of VPg into VPgpUpUOH before and during poliovirus negative-strand RNA synthesis. J Virol 83, 12660-12670.

Stein, A., Pache, R.A., Bernadó, P., Pons, M. and Aloy, P. (2009). Dynamic interactions of proteins in complex networks: a more structured view. FEBS J 276, 5390-5405.

Takeda, A., Nakamura, W., Sasaki, N., Goto, K., Kaido, M., Okuno, T. and Mise, K. (2005). Natural isolates of *Brome mosaic virus* with the ability to move from cell to cell independently of coat protein. J Gen Virol 86, 1201-1211.

Tamm, T. (2000). Cocksfoot mottle virus: the genome organization and translational strategies. PHD Thesis, University of Tartu, Estonia.

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

Tamm, T. and Truve, E. (2000b). Sobemoviruses. J Virol 74, 6231-6241.

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

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

Traoré, M.D., Traoré, V.S.E., Galzi-Pinel, A., Fargette, D., Konaté, G., Traoré, A.S. and Traoré, O. (2008). Abiotic transmission of *Rice yellow mottle virus* through soil and contact between plants. Pak J Biol Sci 11, 900-904.

Truve, E. and Fargette, D. (in press). Genus *Sobemovirus*, in Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses, A. King, E. Carstens, M. Adams and E. Lefkowitz, ed. (Elsevier Academic Press).

Turina, M., Omarov, R., Murphy, J.F., Bazaldua-Hernandez, C., Desvoyes, B. and Scholthof, H.B. (2003). A newly identified role for *Tomato bushy stunt virus* P19 in short distance spread. Mol Plant Pathol 4, 67-72.

Urban, L.A., Ramsdell, D.C., Klomparens, K.L., Lynch, T. and Hancock, J.F. (1989). Detection of *Blueberry shoestring virus* in xylem and phloem tissues of highbush blueberry. Phytophathology 79, 488-493.

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, 21-24.

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, 14147-14152.

Weintraub, M. and Ragetli, H.W. (1970). Electron microscopy of the bean and cowpea strains of southern bean mosaic virus within leaf cells. J Ultrastruct Res 32, 167-189.

Wright, P.E. and Dyson, H.J. (2009). Linking folding and binding. Curr Opin Struct Biol 19, 31-38.

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), 249-257.

Zalloua, P.A., Buzayan, J.M. and Bruening, G. (1996). Chemical cleavage of 5'-linked protein from tobacco ringspot virus genomic RNAs and characterization of the protein-RNA linkage. Virology 219, 1-8.

Zeenko, V. and Gallie, D.R. (2005). Cap-independent translation of tobacco etch virus is conferred by an RNA pseudoknot in the 5'-leader. J Biol Chem 280, 26813-26824.

ACKNOWLEDGEMENTS

This work was carried out in the Department of Gene Technology at Tallinn University of Technology, Estonia.

I wish to express my gratitude to my supervisor professor Erkki Truve for his scientific guidance, encouragement and patience over all these years. I also thank Erkki for providing a high degree of liberty in research which has been an inspiring and highly educating for me.

My appreciation goes to my first supervisor Tiina Tamm who introduced me to laboratory practices, taught me to design and conduct experiments and also to think like a molecule.

I am grateful to all the people who have been in Erkki's group during this work for creating a friendly, helpful and fun laboratory environment. Thank you Birger, Cecilia, Eve-Ly, Gabriela, Grete, Heiti, Jaanus, Jelena, Kairi, Karin, Krista, Kristel, Kristiina, Kristjan, Lenne, Liivi, Maarja, Mariliis, Marina, Merike, Raavo, Signe and Silva, working with all of you has been very pleasant. Of course, all the wonderful colleagues from other groups have my appreciation as well.

Mart Speek, Andres Veske, Richard Tamme and Marko Piirsoo deserve special nomination for creating an entertaining and relaxed atmosphere after hours.

I thank the co-authors of the publications and manuscript. I am grateful to Eugenie Hebrard and Denis Fargette at Institut de Recherche pour le Développement, for the fruitful collaboration and the opportunity to work in Montpellier. I especially value the contribution of Lauri Peil and Liisa Arike at the University of Tartu, without whom an important part of this work would not have been possible.

Last but not least, I would like to thank my mother, sister, Mailis and our children for their support, love and the joy they bring me.

PUBLICATION I

Cocksfoot mottle sobemovirus coat protein contains two nuclear localization signals

Olspert, A., Paves, H., Toomela, R., Tamm, T., Truve, E.

Virus Genes (2010), 40, 423 - 431.

Virus Genes (2010) 40:423–431
DOI 10.1007/s11262-010-0456-9

Cocksfoot mottle sobemovirus coat protein contains two nuclear localization signals

Allan Olspert · Heiti Paves · Raavo Toomela · Tiina Tamm · Erkki Truve

Received: 19 October 2009 / Accepted: 27 January 2010 / Published online: 13 February 2010

Abstract *Cocksfoot mottle virus* (CfMV) coat protein (CP) localization was studied in plant and mammalian cells. Fusion of the full-length CP with enhanced green fluorescent protein (EGFP) localized to the cell nucleus whereas similar constructs lacking the first 33 N-terminal amino acids of CP localized to the cytoplasm. CP and EGFP fusions containing mutations in the arginine-rich motif of CP localized to the cytoplasm and to the nucleus in plant cells indicating the involvement of the motif in nuclear localization. In mammalian cells, mutations in the arginine-rich region were sufficient to completely abolish nuclear transport. The analysis of deletions of amino acid residues 1–11, 1–22, and 22–33 of CP demonstrated that there were two separate nuclear localization signals (NLS) within the N-terminus—a strong NLS1 in the arginine-rich region (residues 22–33) and a weaker NLS2 within residues 1–22. Analysis of point mutants revealed that the basic amino acid residues in the region of the two NLSs were individually not sufficient to direct CP to the nucleus. Additional 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. This feature was not sequence-specific in transient assays since both CfMV and GFP mRNA were transported to the cell nucleus by CfMV CP. Together the results suggest that the nucleus may be involved in CfMV infection.

Keywords CfMV · NLS · Nuclear import of RNA · Plant virus

Introduction

Cocksfoot mottle virus (CfMV) is a positive-sense ssRNA genome plant virus of genus *Sobemovirus* (for review [1]). The viral coat protein (CP) is expressed from the 3′ proximal ORF3 via subgenomic RNA. CfMV particles are assembled according to $T = 3$ quasisymmetry, contain 180 CP molecules and are approximately 30 nm in diameter [2]. The CfMV particle is very similar to other *sobemovirus* particles with known structures: *Southern cowpea mosaic virus* (SCPMV) [3], *Sesbania mosaic virus* (SeMV) [4], *Rice yellow mottle virus* (RYMV) [5] and *Ryegrass mottle virus* (RGMoV) [6]. According to the three-dimensional structure, the sobemovirus CP is divided into two domains, the N-terminal R (random) domain and C-terminal S (shell) domain [3, 7, 8, 5, 2, 6]. The S domain is responsible for subunit–subunit contacts, whereas the R domain is believed to be involved in CP–RNA interactions [7, 8]. The R domain has also been implicated in regulation of the curvature and therefore the size of the particle [5]. The N-terminal region of CP is buried inside the virus particle and is not visible on any of the crystal structures [2, 3, 4, 5, 6]. The primary sequences of sobemovirus CPs are not conserved. However, the N-terminal part of all sobemoviruses CPs is rich in basic amino acids and contains an arginine-rich region. When the N-terminal part of SCPMV or SeMV CP

Electronic supplementary material The online version of this article (doi:10.1007/s11262-010-0456-9) contains supplementary material, which is available to authorized users.

A. Olspert · H. Paves · R. Toomela · T. Tamm · E. Truve (✉)
Department of Gene Technology, Tallinn University of Technology, Akadeemia tee 15, 12618 Tallinn, Estonia
e-mail: erkki.truve@ttu.ee

Present Address:
T. Tamm
Department of General and Microbial Biochemistry, Institute of Molecular and Cell Biology, University of Tartu, Riia 23, 51010 Tartu, Estonia

was removed, only $T = 1$ particles were formed [9, 10]. When the arginine-rich region of SeMV CP was mutated, only empty $T = 3$ particles were formed suggesting the importance of arginine residues for RNA encapsidation [11]. In vitro experiments with CfMV [12] and SCPMV [13] CPs demonstrated that both proteins bind RNA in a sequence non-specific manner. In addition, SCPMV CP possesses sequence-specific viral RNA binding properties [14]. The amino acid residues responsible for SCPMV CP RNA binding in vitro were mapped to the arginine-rich region [13]. It has also been demonstrated that the N-terminal part of SCPMV CP could interact with membranes [15], however, the biological relevance of that feature is unknown. Sobemovirus particles have been found in the nucleus [16, 17, 18]. Based on sequence similarity it has been supposed that the N-terminal region of sobemovirus CPs contains a nuclear localization signal (NLS) [19, 20]. However, as far as we know, this possible feature of sobemovirus CPs has never been addressed experimentally.

The other functions of sobemovirus CPs besides capsid formation are less characterized. It has been demonstrated that the CP is needed for virus cell-to-cell movement but not for virus replication [21, 22]. Long distance movement of SCPMV [18] and RYMV [23] has been observed to be dependent on particle formation.

This study focuses on determining and characterization of the proposed NLS of CfMV CP. The subcellular localization of CP was determined in various systems utilizing enhanced green fluorescent protein (EGFP) fusion technology. The analysis demonstrated that CfMV CP indeed localizes to the nucleus. Since the CP enters the nucleus and also binds RNA [12], it was investigated whether the CP acts as RNA transporter. Microinjection studies demonstrated that the CP was able to transport RNA to the nucleus in a sequence non-specific manner.

Methods

CfMV Norwegian isolate (CfMV-NO) RNA was extracted from purified virus particles as described by Puurand et al. [24]. The viral coat protein was isolated from particles as described by Tamm and Truve [12]. All DNA constructs used in this work were verified by sequencing.

Construction of plasmids

The PCR primers used in the study are listed in Table 1.

CfMV-NO CP cDNA fragments were produced from viral RNA by RT–PCR. The fragment corresponding to the full length CP was amplified with primers 1 and 3, whereas primers 2 and 3 were used to generate the Δ1-33CP deletion mutant. The cDNA fragments containing the coding regions were excised using *Bam*HI and *Pst*I and inserted to mammalian expression vectors pEGFP-N1 and pEGFP-C1 (Clontech), previously digested with *Bgl*II and *Pst*I, to generate plasmids pCP-EGFP, pΔ1-33CP-EGFP, pEGFP-CP and pEGFP-Δ1-33CP.

Plasmids p35S-CP-EGFP and p35S-Δ1-33CP-EGFP were constructed by inserting the *Eco*RI–*Not*I restriction fragments of pCP-EGFP and pΔ1-33CP-EGFP by blunt end ligation to the plant expression vector pANU5 [25] digested with *Ecl*136II and *Bam*HI. Plasmids p35S-EGFP-CP and p35S-EGFP-Δ1-33CP were constructed by inserting the *Xho*I–*Xba*I restriction fragments of PCR products obtained with primers 8 and 9 using pEGFP-CP and pEGFP-Δ1-33CP as templates to the plant expression vector pANU5 digested with *Xho*I and *Xba*I.

Enhanced green fluorescent protein (EGFP) hybrids containing the mutated CP, pR5X-CP-EGFP, and pR3L-CP-EGFP were generated similarly to pCP-EGFP. This portion of the CP coding region overlaps with the replicase coding region in the viral genome. Therefore, these and other point-mutations were made in the manner which disrupted the replicase the least in order to be able to study the same mutants in the whole virus context if needed. The exact mutations of these constructs and those discussed below are shown on Fig. 2. The mutations were introduced to pCP-EGFP by overlap extension PCR with mutant primers (4 and 5 for R5X, 6 and 7 for R3L) and outer primers 1 and 3. Plasmids p35S-R5X-CP-EGFP and p35S-R3L-CP-EGFP were constructed by inserting the *Eco*RI–*Xba*I restriction fragments of pR5X-CP-EGFP and pR3L-CP-EGFP to pANU5 digested with *Eco*RI and *Xba*I.

For the generation of p35S-K2X-CP-EGFP, p35S-RK3X-CP-EGFP, p35S-R3L-K2X-CP-EGFP, p35S-R3L-RK3X-CP-EGFP, p35S-R5X-RK3X-CP-EGFP, and p35S-R5X-RK3X-K2X-CP-EGFP primers containing mutations (10 and 11 for K2X, 12 and 13 for RK3X) were used in overlap PCR reactions with outer primers 3 and 18, and p35S-CP-EGFP, p35S-R3L-CP-EGFP, or p35S-R5X-CP-EGFP as a template. p35S-Δ1-11CP-EGFP and p35S-Δ1-22CP-EGFP were obtained with primers 14 or 15 and 3, respectively. For p35S-Δ22-33CP-EGFP, the deletion was inserted utilizing overlap PCR and mutated primers 16 and 17 together with outer primers 3 and 18. *Eco*RI and *Eco*47III cleaved PCR products containing mutations were inserted to p35S-CP-EGFP excised with the same enzymes. The triple mutant p35S-R5X-RK3X-K2X-CP-EGFP was generated from p35S-R5X-RK3X-CP-EGFP by additionally mutating the K2X motif, as described above.

Expression of CP and EGFP hybrids in plant cells

The epidermal cell layer of onion bulbs on medium described by Shieh et al. [26] was transformed biolistically

Table 1 Primers used for generating CfMV CP mutants

Nr	Primer sequence	Description
1	5′-CAGGATCCGAATTCATGGTGAGGAAAGGA GCAGC-3′	CP start, contains CfMV nt 3096-3115
2	5′-CAGGATCCGAATTCATGGAGCCAGTCTCT CGACC-3′	CP Δ1-33, contains CfMV nt 3191-3211
3	5′-GCCTGCAGAGATCTGGTACCCAAATTTGTA GAAGGGGAAAC-3′	CP end, complementary to CfMV nt 3834-3853
4	5′-GGCCA^[1]GCCTCCTGCGGC^[2]CCCCC-3′	R5X mutation, partially complementary to CfMV nt 3162-3180, the positions of deleted nucleotides C and GG, respectively, are indicated
5	5′-GTCCTGCCTCTACCAGCCTTCAAG-3′	Complementary to CfMV nt 3242-3265
6	5′-**T**CC**T**CAGGAGGC**T**TGGCCGGTCGATGGAG-3′	R3L mutation, partially complementary to CfMV nt 3169-3197, positions of nucleotide substitutions G to T are shown in bold
7	5′-CGACCGGCCA**A**GCCTCCTG**A**GG**A**GGCCCC-3′	R3L mutation, partially complementary to CfMV nt 3163-3191, positions of nucleotide substitutions C to A are shown in bold
8	5′-AGCTCGAGTCGCCACCATGGTGAGCAAGGG-3′	Contains *Xho*I site and pEGFP-C1 nt 604-626
9	5′-TCTCTAGACTACAAATTTGTAGAAGGGG-3′	Contains *Xba*I site and sequence complementary to CfMV nt 3838-3857
10	5′-CAACCA**GG**ACC**A**A**C**GGCTCAGCAGC-3′	K2X mutation, partially contains CfMV nt 3132-3166, positions of nucleotide substitutions AA to GG, C to A and G to C are shown in bold
11	5′-GCTGCTGAGCC**G**T**T**GGT**CC**TGGTTG-3′	K2X mutation, partially complementary to CfMV nt 3132-3166, positions of nucleotide substitutions C to G, G to T and TT to CC are shown in bold
12	5′-GA**C**GA**T**AGGAGCAGCAACGA**T**AGCCCCGC-3′	RK3X mutation, partially contains CfMV nt 3101-3128, positions of nucleotide substitutions G to C, A to T and A to T are shown in bold
13	5′-GCGGGGCT**A**TCGTTGCTGCTCCT**A**TC**G**TC-3′	RK3X mutation, partially complementary to CfMV nt 3101-3128, positions of nucleotide substitutions T to A, T to A and C to G are shown in bold
14	5′-ATGAATTCGGAGCAGCAACGAAA**ATG**C-3′	CP Δ1-11 mutation, partially contains CfMV nt 3108-3127, mutations producing the initiation codon are shown in bold
15	5′-ATGAATTCACCCAAGGCTCAGCAG**AT**GC-3′	CP Δ1-22 mutation, partially contains CfMV nt 3140-3159, mutations producing the initiation codon are shown in bold
16	5′-CCCAAGGCTCAGCAG-ATGGAGCCAGTCTC-3′	CP Δ22-33 mutation, contains CfMV nt 3141-3155 and 3192-3205, position of the deletion is indicated with a dash
17	5′-GAGACTGGCTCCAT-CTGCTGAGCCTTGGG-3′	CP Δ22-33 mutation, complementary to CfMV nt 3141-3155 and 3192-3205, position of the deletion is indicated with a dash
18	5′-GCCAGCTGCTGCAGGTCAACATGG-3′	Upstream 5′ primer for overlap PCR, complementary to pANU5 nt 2931-2947
19	5′-TCGAATTCGCCAAATTTGTAGAAGG-3′	CP end, complementary to CfMV nt 3840-3854

The restriction endonuclease sites used are underlined, base numbering refers to the CfMV-NO genome as in [20]

using PDS–100/He Biolistic device (BioRad) and 1.0 μm gold particles with 1100 or 1350 psi rupture discs or Helios gene gun (BioRad) at 200 psi. After bombardment, the cell layers were incubated in the dark for 24–48 h at room temperature. For each construct 10–50 cells expressing EGFP were analyzed.

Expression of CP and EGFP hybrids in Cos7 cells

Cos7 cells were grown on IMDM or DMEM containing 10% fetal calf serum and transfected by electroporation or using transfection reagent FuGENE6 (Roche). 1–2 μg plasmid DNA and 5×10^6 cells were taken per transfection, electroporations were carried out in the presence of 10 μg salmon sperm sonicated DNA as carrier. Afterwards the cells were incubated for 12–18 h at 37°C and 5% CO_2.

RNA labeling and microinjection

CfMV and GFP RNA were labeled with Alexa Fluor® 488-UTP (Molecular Probes) by in vitro RNA synthesis without a cap analogue and purified by standard phenol–chloroform extraction. The integrity was verified with gel electrophoresis immediately before injection. Alexa Fluor® 488-labeled RNA was mixed 15 min before the injection with CfMV CP in molar ratio of approximately 1:1.5, respectively. Fluorescent RNA (~2 pmol) or RNA/CP mixture was microinjected into onion bulb scale epidermal cells using CellTram Oil injector (Eppendorf). Viability of injected cells was determined by monitoring cytoplasmic streaming during the microinjections. In each experiment at least 10 cells were injected.

Visualization of EGFP hybrids and Alexa Fluor® 488 fluorescence

Fluorescence microscope equipped with a filter for GFP/Alexa Fluor® 488 was used to analyze the subcellular localization of protein or labeled RNA. Bombarded plant tissues were mounted into water or a mixture of water and glycerol 1:1 (v/v) for examination.

Histochemical staining of nuclei

The nuclei of Cos7 cells were stained with 0.5 μg ml^{-1} Hoechst 33342 dye (Molecular Probes) and 0.05% Triton X-100 in PBS. The nuclei and cell walls of onion epidermal cells were stained with 5 μg ml^{-1} ethidium bromide and 0.05% Triton X-100 in PEM buffer (0.1 M PIPES, 5 mM EGTA, 2 mM $MgCl_2$, pH 6.8). When onion cells were used for microinjection, the nuclei were stained with 0.5 μg ml^{-1} Hoechst 33342 dye in an aqueous solution containing 2% paraformaldehyde and 0.1% saponin.

Results

CfMV CP localizes to the nucleus

To test the hypothesis that sobemoviral CP contains a NLS, the subcellular localization of CfMV CP and EGFP fusions was determined in onion bulb epidermal cells. EGFP alone localized to the cytoplasm and to the nucleus (Fig. 1). The hybrid protein where EGFP was fused to the C-terminus of full length CP localized to the nucleus (Fig. 1). However, similar fusion with the deletion of the first 33 amino acid residues of CP (Δ1-33CP) completely stopped nuclear transport (Fig. 1). The subcellular localization of proteins where the EGFP was fused to the N-terminus of CP or Δ1-33CP was also determined. Again CP localized to the nucleus and the Δ1-33CP was detected only in the cytoplasm (Supplementary Fig. 1).

To determine the exact position of CP NLS, several previously described nuclear localization signals [27, 28] were compared with the N-terminus of CP. The region was also analyzed using PSORT WoLF program (http://psort.nibb.ac.jp). The analysis showed that most likely amino acid residues 26–30 RRRRR (the arginine-rich motif) form a single partite NLS (data not shown), the protein was also predicted to localize to the chloroplast. To verify these findings two different mutations were made in the putative NLS (Fig. 2): in protein R5X-CP CP residues 26–30 were replaced with PQEA and in R3L CP where three of the arginines were replaced with leucines: R26L, R27L, R30L. According to PSORT, such proteins should not localize to the nucleus. The subcellular localization of R5X-CP and R3L-CP in onion epidermal cells revealed that the arginine-rich region was indeed involved in determining the localization of CP. However, these mutations were not sufficient to completely abolish nuclear import as both of the proteins were detected both in the cytoplasm and in the nucleus (Fig. 1, Supplementary Fig. 2). The subcellular localization of CP EGFP fusion proteins was also determined in mammalian cells (Fig. 3). In Cos7 cells, the CfMV CP NLS was only functional when positioned at the N-terminus of the fusion protein (Fig. 3), as full length CP fused to the C-terminus of EGFP localized to the cytoplasm (Supplementary Fig. 1). Furthermore, both mutations of the arginine-rich region, R5X and R3L, were sufficient to completely disrupt nuclear transport of the protein in mammalian cells (Fig. 3).

CfMV CP N-terminus contains two NLSs

The experiments in plant cells demonstrated that the CfMV CP NLS was not exclusively located to the arginine-rich region. Hence additional deletion mutants of the N-terminus of CP were created and analyzed. The CP N-terminal region was divided into three parts and the role of amino acid residues 1–11, 1–22, and 22–33 were studied separately (Fig. 2). Deletions of 11 (Δ1-11CP) and 22 (Δ1-22CP) amino acid residues had no detectable difference in subcellular localization compared to full length CP, as majority of the signal was detected in the nucleus. Deletion of amino acid residues from 22 to 33 (Δ22-33CP), however, reduced nuclear import as the mutant was detected both in cytoplasm and nucleus (Fig. 1).

Mutants R5X and Δ22-33CP demonstrated that the arginine-rich motif was not the only motif in the N-terminus which contributes to nuclear localization. Furthermore, mutants Δ1-22CP and Δ22-33CP demonstrated that there are two separately functional NLSs and ruled out a bipartite NLS where the two basic amino acid clusters are both required for nuclear localization [29]. Therefore, it was decided to investigate the impact of all positively charged amino acid residues at the N-terminus of CP on subcellular localization. Basic amino acid residues other than those within arginine-rich region (not counting the last arginine in the region R32, which seemed to bear no significance since mutants R5X and Δ22-33CP have an identical localization) were divided into two groups: RK3X (amino acid residues 4, 5, and 10) and K2X (amino acid residues 16 and 18) (Fig. 2), which then were mutated and analyzed separately and in combination with R5X and R3L. Either group separately appeared to have no effect on nuclear transport as both mutants were localized to the nucleus (Fig. 1). This is probably due to the NLS in the arginine-rich region. When a “double” mutant R3L-K2X

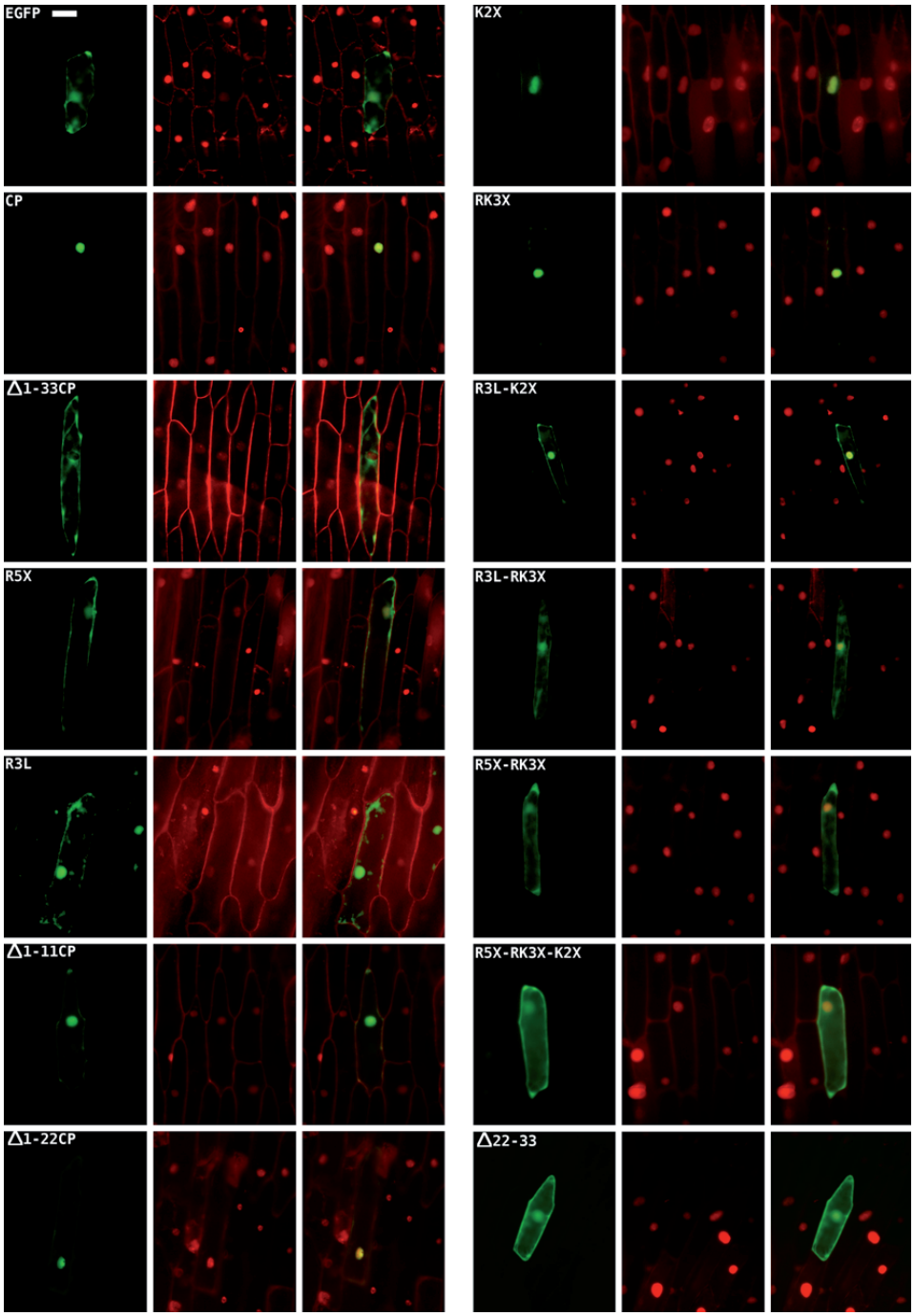


Fig. 1 The subcellular localization of CfMV CP and EGFP fusion proteins transiently expressed in onion epidermal cells. *Left panels* correspond to EGFP fluorescence (*green*), middle to nuclear staining with EtBr (*red*) and right to overlay. Cells were transfected with constructs coding the proteins indicated on the left, CP corresponds to full length protein. *Scale bar* 50 μm

WT	– MMV**RK**GAAT**K**APQQP**K**P**K**AQQQPGG**RRRRR**G**R**SM
R5X	– MMV**RK**GAAT**K**APQQP**K**P**K**AQQQPGG**PQEA**.G**R**SM
R3L	– MMV**RK**GAAT**K**APQQP**K**P**K**AQQQPGG**LLRRL**G**R**SM
K2X	– MMV**RK**GAAT**K**APQQP**G**P**T**AQQQPGG**RRRRR**G**R**SM
RK3X	– MMV**TI**GAAT**I**APQQP**K**P**K**AQQQPGG**RRRRR**G**R**SM
R3L-K2X	– MMV**RK**GAAT**K**APQQP**G**P**T**AQQQPGG**LLRRL**G**R**SM
R3L-RK3X	– MMV**TI**GAAT**I**APQQP**K**P**K**AQQQPGG**LLRRL**G**R**SM
R5X-RK3X	– MMV**TI**GAAT**I**APQQP**K**P**K**AQQQPGG**PQEA**.G**R**SM
R5X-RK3X-K2X	– MMV**TI**GAAT**I**APQQP**G**P**T**AQQQPGG**PQEA**.G**R**SM
Δ1-33CP	– .....M
Δ1-11CP	– .....**M**PQQP**K**P**K**AQQQPGG**RRRRR**G**R**SM
Δ1-22CP	– .....**M**PGG**RRRRR**G**R**SM
Δ22-33CP	– MMV**RK**GAAT**K**APQQP**K**P**K**AQQ.....M

Fig. 2 CfMV CP N-terminus and mutants used for determining the subcellular localization of EGFP hybrid proteins. The protein sequence of the first 34 residues of CfMV CP is shown on *top*, positively charged residues are shown in *bold*. Position and the exact nature of mutations are indicated

was expressed there were no distinguishable differences compared to R3L, the same being true for R3L-RK3X (Fig. 1). There was also no major dissimilarity when comparing the localization of R5X and R5X-RK3X. Analysis of a "triple" mutant R5X-RK3X-K2X also revealed localization quite similar to R5X and R5X-RK3X (Fig. 1). These data demonstrated that mutating majority of the basic amino acid residues was not sufficient to completely stop nuclear import.

CfMV CP functions as RNA transporter

Since we demonstrated that CfMV CP is transported to the cell nucleus and it had been previously shown to bind RNA [12], it was investigated whether the CP transports RNA to the cell nucleus. Series of microinjections into onion bulb scale epidermis cells were performed using mixture of fluorescent CfMV RNA or GFP mRNA and CP purified from virus particles. No nuclear entry of RNA was observed when only fluorescent RNA was injected (Fig. 4a). Injection of fluorescent CfMV RNA and non-fluorescent CP mixture showed the accumulation of fluorescent label in the nucleus of injected cell within 5 min (Fig. 4b). As the RNA binding of CfMV CP in vitro is sequence non-specific [12] it was investigated if non-viral RNA is transported to the nucleus. Fluorescent GFP mRNA also accumulated into the nucleus in the presence of CP in the injection mixture (Fig. 4c).

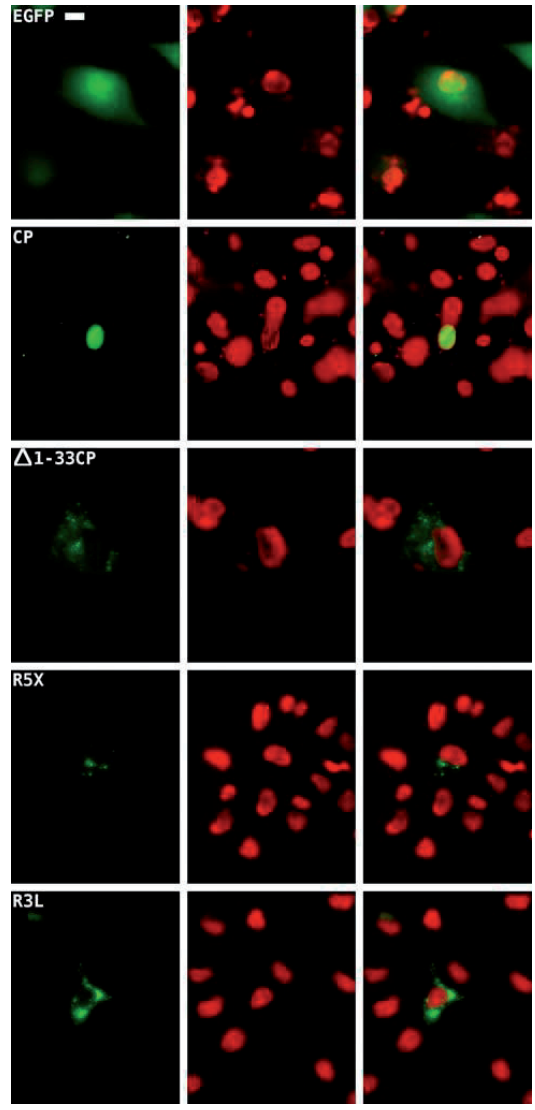


Fig. 3 The subcellular localization of CfMV CP and EGFP fusion proteins transiently expressed in Cos7 cells. *Left panels* correspond to EGFP fluorescence (*green*), *middle* to nuclear staining with Hoechst (*red*) and *right* to overlay. Cells were transfected with constructs coding the proteins indicated on the left, CP corresponds to full length protein. *Scale bar* 10 μm

Discussion

Although positive-strand RNA viruses are generally believed to be mainly cytoplasmic, some positive-strand RNA virus proteins [30, 31, 32, 33, 34] have been reported to localize to the nucleus. Nuclear transport of *Potato*

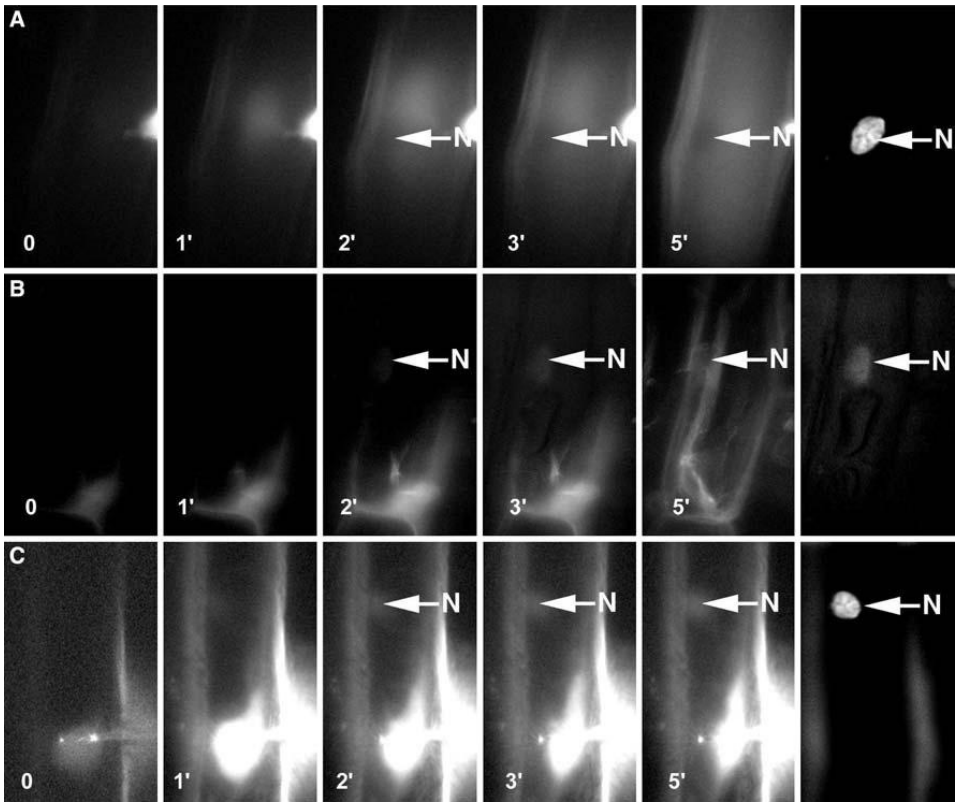


Fig. 4 Microinjection of Alexa Fluor® 488-labeled CfMV RNA into onion bulb epidermal cells. Microinjection of Alexa Fluor® 488-labeled CfMV RNA (**a**), mixture of Alexa Fluor® 488-labeled CfMV RNA and CP (**b**), and mixture of Alexa Fluor® 488-labeled GFP mRNA and CP (**c**). *N* nucleus. Position of the nucleus, determined by Hoechst staining, is shown on the *right*

leafroll virus CP has also been reported [35]. The N-terminus of CfMV and all other sobemovirus CPs are rich in basic amino acid residues and it has been suggested that all sobemoviruses might have a putative NLS in the N-terminal region of CP [19]. Furthermore, sobemovirus particles have been found in the nucleus [16, 17, 18]. The current study is the first report that provides direct evidence about the nuclear localization of CP and the importance of the N-terminal region regarding the subcellular localization. Point mutations in the arginine-rich region were sufficient to disrupt the CfMV CP NLS in mammalian cells. In onion epidermal cells the same mutations only reduced the efficiency of nuclear import but did not abolish it completely, suggesting that the arginine-rich region is not the only motif directing nuclear transport in plants. Thus, the arginine-rich region of CP is an NLS in mammalian as well as in plant cells but in the latter additional NLSs exist. This is a demonstration that, although nuclear import systems of plant and animal cells are generally very similar, subtle differences exist in case of certain proteins. According to our knowledge this is the first report on such plant virus protein.

Indeed, there are several other basic amino acid residues in the N-terminal region of CfMV CP besides arginine-rich motif and even neutral or acidic residues can be essential for nuclear localization [28, 36]. Point mutations of other basic residues in combination with the arginine-rich region mutants demonstrated that individually these residues are dispensable for the functioning of the second NLS. Nonetheless, the data allow us to conclude that in plant cells two NLSs are used for CfMV CP transport, a strong NLS1 within residues 22–33 and a weaker NLS2 within residues 1–22.

Sobemoviruses like all positive-strand RNA viruses code proteins that bind nucleic acids. It has been previously shown for CfMV that CP binds RNA in a sequence non-specific manner [12]. Therefore, we investigated whether CfMV CP enters the nucleus when bound to RNA. Microinjection studies demonstrated that the viral CP was able to transport labeled RNA to the nucleus of onion

epidermal cells in sequence non-specific manner. Previous studies have shown that the arginine-rich region is responsible for SCPMV CP RNA-binding capabilities in vitro [13] and RNA encapsidation of SeMV virus like particles [11]. In our experiments, the CP bound to RNA was still transported to nucleus which suggests that RNA did not interfere with importin binding or that RNA and importin do not interact with the same amino acid residues.

We have several hypotheses why the CP is transported to the nucleus. After entering a cell and particle disassembly, the virus may use the nucleus as a means of reducing CP concentration in the cytoplasm which may be necessary for the regulation of virus replication. However, a more likely scenario is that the CP has functions that require its transport to the nucleus. For instance 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, respectively [37, 38, 39].

Although we do not know at the moment why CfMV CP enters the nucleus, our results support previous findings, which demonstrate that sobemovirus particles are found in the nucleus. Moreover, our data suggest that the particles found in the nucleus are assembled there, since the N-terminus of the CP with NLS is buried within the particle [2] and therefore is not accessible after virion formation. Our findings that CfMV CP can act as RNA transporter also suggests that these particles may contain viral RNA.

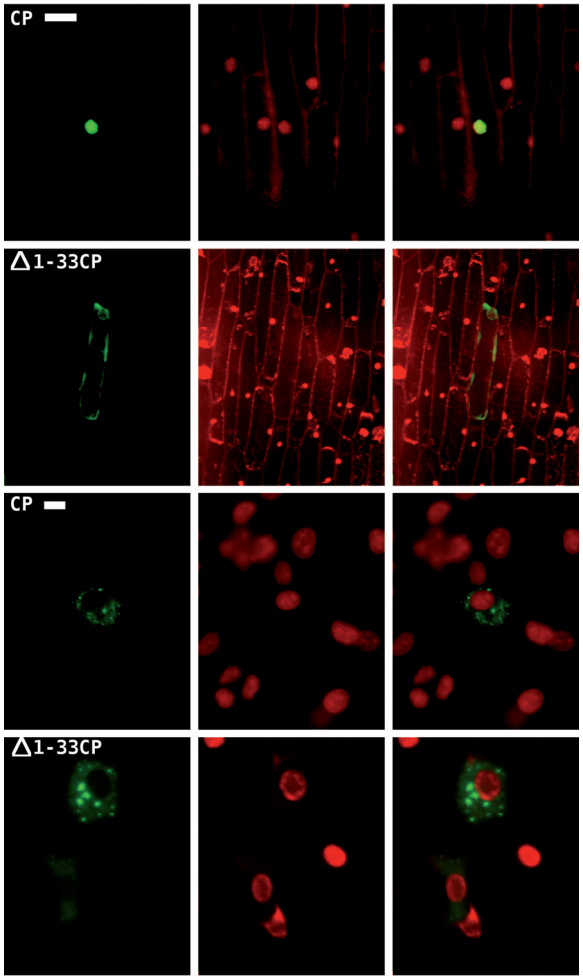
Several characteristics have been assigned to the N-terminus of sobemoviral CPs: RNA binding, particle formation and association with membranes. This study adds two more: a functional NLS along with RNA nuclear transport.

Acknowledgments We wish to thank Merike Sõmera for fruitful discussions and Ants Kurg for the initial supply of fluorescent UTP. This work was supported by Estonian Science Foundation grant no. 7363.

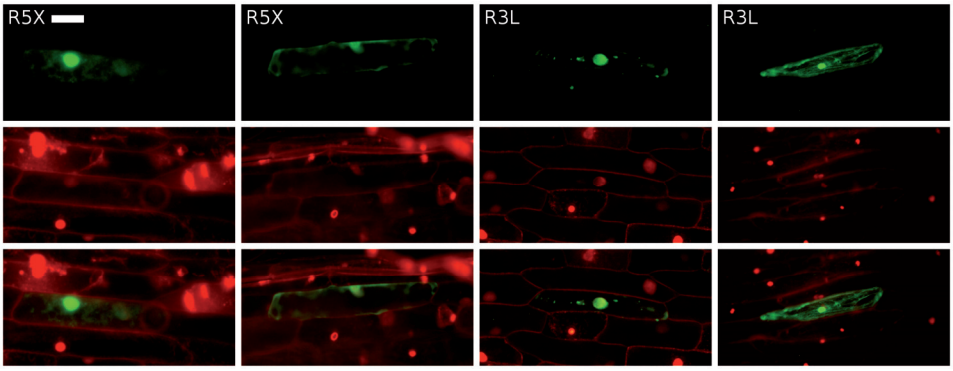
References

1. T. Tamm, E. Truve, Sobemoviruses. J. Virol. **74**, 6231–6241 (2000)
2. K. Tars, A. Zeltins, L. Liljas, The three-dimensional structure of cocksfoot mottle virus at 2.7 Å resolution. Virology **310**, 287–297 (2003)
3. C. Abad-Zapatero, S.S. Abdel-Meguid, J.E. Johnson, A.G. Leslie, I. Rayment, M.G. Rossmann, D. Suck, T. Tsukihara, Structure of southern bean mosaic virus at 2.8 Å resolution. Nature **286**, 33–39 (1980)
4. M. Bhuvaneshwari, H.S. Subramanya, K. Gopinath, H.S. Savithri, M.V. Nayudu, M.R. Murthy, Structure of sesbania mosaic virus at 3 Å resolution. Structure **3**, 1021–1030 (1995)
5. C. Qu, L. Liljas, N. Opalka, C. Brugidou, M. Yeager, R.N. Beachy, C.M. Fauquet, J.E. Johnson, T. Lin, 3D domain swapping modulates the stability of members of an icosahedral virus group. Structure **8**, 1095–1103 (2000)
6. P. Plevka, K. Tars, A. Zeltins, I. Balke, E. Truve, L. Liljas, The three-dimensional structure of ryegrass mottle virus at 2.9 Å resolution. Virology **369**, 364–374 (2007)
7. M.A. Hermodson, C. Abad-Zapatero, S.S. Abdel-Meguid, S. Pundak, M.G. Rossmann, J.H. Tremaine, Amino acid sequence of southern bean mosaic virus coat protein and its relation to the three-dimensional structure of the virus. Virology **119**, 133–149 (1982)
8. M.G. Rossmann, C. Abad-Zapatero, M.A. Hermodson, J.W. Erickson, Subunit interactions in southern bean mosaic virus. J. Mol. Biol. **166**, 37–73 (1983)
9. H.S. Savithri, J.W. Erickson, The self-assembly of the cowpea strain of southern bean mosaic virus: formation of T = 1 and T = 3 nucleoprotein particles. Virology **126**, 328–335 (1983)
10. G.L. Lokesh, T.D.S. Gowri, P.S. Satheshkumar, M.R.N. Murthy, H.S. Savithri, A molecular switch in the capsid protein controls the particle polymorphism in an icosahedral virus. Virology **292**, 211–223 (2002)
11. P.S. Satheshkumar, G.L. Lokesh, M.R. Murthy, H.S. Savithri, The role of arginine-rich motif and beta-annulus in the assembly and stability of Sesbania mosaic virus capsids. J. Mol. Biol. **353**, 447–458 (2005)
12. T. Tamm, E. Truve, RNA-binding activities of cocksfoot mottle sobemovirus proteins. Virus Res. **66**, 197–207 (2000)
13. S.K. Lee, D.L. Hacker, 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**, 317–327 (2001)
14. D.L. Hacker, Identification of a coat protein binding site on southern bean mosaic virus RNA. Virology **207**, 562–565 (1995)
15. S.K. Lee, C. Dabney-Smith, D.L. Hacker, B.D. Bruce, Membrane activity of the southern cowpea mosaic virus coat protein: the role of basic amino acids, helix-forming potential, and lipid composition. Virology **291**, 299–310 (2001)
16. F. Rabenstein, A. Stanarius, Untersuchungen zum Knaulgrasscheckungs-virus (cocksfoot mottle virus). Arch. Phytopathol. Pflanzenschutz **20**, 15–31 (1984)
17. N. Mohamed, D. Mossop, Cynosurus and cocksfoot mottle viruses: a comparison. J. Gen. Virol. **55**, 63–74 (1981)
18. A.L. Fuentes, R.I. Hamilton, 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**, 1903–1910 (1993)
19. M. Ngon A Yassi, C. Ritzenthaler, C. Brugidou, C. Fauquet, R.N. Beachy, Nucleotide sequence and genome characterization of rice yellow mottle virus RNA. J. Gen. Virol. **75**, 249–257 (1994)
20. K. Mäkinen, T. Tamm, V. Næss, E. Truve, Ü. Puurand, T. Munthe, M. Saarma, Characterization of cocksfoot mottle sobemovirus genomic RNA and sequence comparison with related viruses. J. Gen. Virol. **76**, 2817–2825 (1995)
21. C. Brugidou, C. Holt, M. Ngon A Yassi, S. Zhang, R. Beachy, C. Fauquet, Synthesis of an infectious full-length cDNA clone of rice yellow mottle virus and mutagenesis of the coat protein. Virology **206**, 108–115 (1995)
22. K. Sivakumaran, B.C. Fowler, D.L. Hacker, Identification of viral genes required for cell-to-cell movement of southern bean mosaic virus. Virology **252**, 376–386 (1998)
23. N. Opalka, C. Brugidou, C. Bonneau, M. Nicole, R.N. Beachy, M. Yeager, C. Fauquet, Movement of rice yellow mottle virus between xylem cells through pit membranes. Proc. Natl Acad. Sci. USA **95**, 3323–3328 (1998)
24. Ü. Puurand, K. Mäkinen, M. Baumann, M. Saarma, Nucleotide sequence of the 3'-terminal region of potato virus a RNA. Virus Res. **23**, 99–105 (1992)

25. K. Mäkinen, V. Næss, T. Tamm, E. Truve, A. Aaspõllu, M. Saarma, The putative replicase of the cocksfoot mottle sobemovirus is translated as a part of the polyprotein by -1 ribosomal frameshift. Virology **207**, 566–571 (1995)
26. M.W. Shieh, S.R. Wessler, N.V. Raikhel, Nuclear targeting of the maize R protein requires two nuclear localization sequences. Plant Physiol. **101**, 353–361 (1993)
27. E.A. Nigg, Nucleocytoplasmic transport: signals, mechanisms and regulation. Nature **386**, 779–787 (1997)
28. S. Kosugi, M. Hasebe, N. Matsumura, H. Takashima, E. Miyamoto-Sato, M. Tomita, H. Yanagawa, Six classes of nuclear localization signals specific to different binding grooves of importin α. J. Biol. Chem. **284**, 478–485 (2009)
29. J. Robbins, S.M. Dilworth, R.A. Laskey, C. Dingwall, Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. Cell **64**, 615–623 (1991)
30. M.C. Schaad, R. Haldeman-Cahill, S. Cronin, J.C. Carrington, Analysis of the VPg-proteinase (NIa) encoded by tobacco etch potyvirus: effects of mutations on subcellular transport, proteolytic processing, and genome amplification. J. Virol. **70**, 7039–7048 (1996)
31. M.R. Hajimorad, X.S. Ding, S. Flasinski, S. Mahajan, E. Graff, R. Haldman-Cahill, J.C. Carrington, B.G. Cassidy, NIa and NIb of peanut stripe potyvirus are present in the nucleus of infected cells, but do not form inclusions. Virology **224**, 368–379 (1996)
32. A.P. Lucy, H.S. Guo, W.X. Li, S.W. Ding, Suppression of post-transcriptional gene silencing by a plant viral protein localized in the nucleus. EMBO J. **19**, 1672–1680 (2000)
33. E.V. Ryabov, S.H. Kim, M. Taliansky, Identification of a nuclear localization signal and nuclear export signal of the umbraviral long-distance RNA movement protein. J. Gen. Virol. **85**, 1329–1333 (2004)
34. A.R. Rebelo, S. Niewiadomski, S.W. Prosser, P. Krell, B. Meng, Subcellular localization of the triple gene block proteins encoded by a foveavirus infecting grapevines. Virus Res. **138**, 57–69 (2008)
35. S. Haupt, T. Stroganova, E. Ryabov, S.H. Kim, G. Fraser, G. Duncan, M.A. Mayo, H. Barker, M. Taliansky, Nucleolar localization of potato leafroll virus capsid proteins. J. Gen. Virol. **86**, 2891–2896 (2005)
36. J.P. Makkerh, C. Dingwall, R.A. Laskey, Comparative mutagenesis of nuclear localization signals reveals the importance of neutral and acidic amino acids. Curr. Biol. **6**, 1025–1027 (1996)
37. S.H. Kim, S. Macfarlane, N.O. Kalinina, D.V. Rakitina, E.V. Ryabov, T. Gillespie, S. Haupt, J.W.S. Brown, M. Taliansky, Interaction of a plant virus-encoded protein with the major nucleolar protein fibrillarin is required for systemic virus infection. Proc. Natl Acad. Sci. USA **104**, 11115–11120 (2007)
38. S.H. Kim, E.V. Ryabov, N.O. Kalinina, D.V. Rakitina, T. Gillespie, S. MacFarlane, S. Haupt, J.W.S. Brown, M. Taliansky, Cajal bodies and the nucleolus are required for a plant virus systemic infection. EMBO J. **26**, 2169–2179 (2007)
39. M. Rajamäki, J.P.T. Valkonen, Control of nuclear and nucleolar localization of nuclear inclusion protein a of picorna-like potato virus A in *Nicotiana* species. Plant Cell **21**, 2485–2502 (2009)



Supplementary Fig. 1. The subcellular localization of CfMV CP and Δ1-33CP fused to the C-terminus of EGFP. CP and Δ1-33CP were transiently expressed in onion (upper two rows) and Cos7 (lower two rows) cells. Left panels correspond to EGFP fluorescence (green), middle to nuclear staining with Hoechst (red) and right to overlay. Cells were transfected with constructs coding the proteins indicated on the left, CP corresponds to full length protein. Scale bars represent 50 μm and 10 μm in onion and Cos7 cells, respectively.



Supplementary Fig. 2 The subcellular localization of CfMV CP mutants R3L and R5X fused to EGFP, transiently expressed in onion epidermal cells. Images from additional independent experiments revealing the consistency of the localization patterns. Upper panels correspond to EGFP fluorescence (green), middle to nuclear staining with EtBr (red) and bottom to overlay. Cells were transfected with constructs coding the proteins indicated on top. Scale bar 50 µm.

PUBLICATION II

Protein-RNA linkage and post-translational modifications of two sobemovirus VPgs

Olspert, A., Peil, L., Hébrard, E., Fargette, D., Truve, E.

Journal of General Virology (2011), 92, 445 - 452.

Journal of General Virology (2011), **92**, 445–452 DOI 10.1099/vir.0.026476-0

Protein–RNA linkage and post-translational modifications of two sobemovirus VPgs

Allan Olspert,[1] Lauri Peil,[2] Eugénie Hébrard,[3] Denis Fargette[3] and Erkki Truve[1]

Correspondence
Erkki Truve
erkki.truve@ttu.ee

[1]Department of Gene Technology, Tallinn University of Technology, Akadeemia tee 15, 12618 Tallinn, Estonia

[2]Institute of Technology, University of Tartu, Nooruse 1, 50411 Tartu, Estonia

[3]UMR, Institut de Recherche pour le Développement (IRD), BP 64501, 34394 Montpellier cedex 5, France

Received 16 August 2010
Accepted 8 November 2010

Sobemoviruses possess a viral genome-linked protein (VPg) attached to the 5′ end of viral RNA. VPg is processed from the viral polyprotein. In the current study, Cocksfoot mottle virus (CfMV) and Rice yellow mottle virus (RYMV) VPgs were purified from virions and analysed by mass spectrometry. The cleavage sites in the polyprotein and thereof the termini of VPg were experimentally proven. The lengths of the mature VPgs were determined to be 78 and 79 aa residues, respectively. The amino acid residues covalently linked to RNA in the two VPgs were, surprisingly, not conserved; it is a tyrosine at position 5 of CfMV VPg and serine at position 1 of RYMV VPg. Phosphorylations were identified in CfMV and RYMV VPgs with two positionally similar locations T20/S14 and S71/S72, respectively. RYMV VPg contains an additional phosphorylation site at S41.

INTRODUCTION

Cocksfoot mottle virus (CfMV) and *Rice yellow mottle virus* (RYMV) are members of the genus *Sobemovirus*, a group of viruses with small icosahedral virions and a positive-sense ssRNA genome of approximately 4.0–4.5 kb. Like many other genera with an RNA genome, sobemoviruses have a viral genome-linked protein (VPg) attached to the 5′ end of the genomic and subgenomic RNAs (Ghosh *et al.*, 1981; Mang *et al.*, 1982).

The VPgs of sobemoviruses are translated as part of the polyprotein and cleaved by the viral protease (Nair & Savithri, 2010; van der Wilk *et al.*, 1998). In contrast to potyviruses, the polyprotein processing and VPg maturation of sobemoviruses is poorly described. The specificity of the sobemoviral protease has been proposed as Q, E/T, S, N (Gorbalenya *et al.*, 1988; Mäkinen *et al.*, 2000; Nair & Savithri, 2010; van der Wilk *et al.*, 1998), based on the fact that many different cleavage sites can be predicted for the N and C termini of sobemovirus VPgs. For several sobemoviruses – CfMV, RYMV, Southern bean mosaic virus (SBMV) and Sesbania mosaic virus (SeMV) – the N terminus of VPg has been mapped (Hébrard *et al.*, 2008; Mäkinen *et al.*, 2000; Nair & Savithri, 2010; van der Wilk *et al.*, 1998), while the C terminus of VPg has so far been experimentally proven for only SeMV (Nair & Savithri, 2010). The determined SeMV VPg processing sites corroborate the predicted consensus cleavage sequence. However, sobemoviruses deploy −1 programmed ribosomal frameshifting (−1 PRF) for the expression of polyprotein and VPg occupies a position in the polyprotein close to the −1 PRF signal. Therefore, it has been proposed that at least CfMV might express its VPg through the −1 PRF mechanism and as a result even encode VPgs with different C termini (Mäkinen *et al.*, 2000).

The VPgs are covalently linked to the 5′ end of viral RNA (Ambros & Baltimore, 1978; Rothberg *et al.*, 1978). The VPg is attached to the RNA over a phosphodiester bond formed between the hydroxyl group of the amino acid residue and 5′ phosphate group of RNA (Ambros & Baltimore, 1978; Rothberg *et al.*, 1978). The amino acid residue involved in the linkage has been reported to be a tyrosine or a serine (Ambros & Baltimore, 1978; Jaegle *et al.*, 1987). Threonine also contains a hydroxyl group, but there is no evidence that it is used for linking with RNA. Picornaviruses use a conserved tyrosine residue situated near the N terminus of VPg for the linkage of RNA (Ambros & Baltimore, 1978; Rothberg *et al.*, 1978; Schein *et al.*, 2006). Also for potyviruses and caliciviruses the use of tyrosine has been reported (Anindya *et al.*, 2005; Belliot *et al.*, 2008; Murphy *et al.*, 1991), while nepoviruses and comoviruses are reported to exploit a serine residue (Jaegle *et al.*, 1987; Zalloua *et al.*, 1996).

Like most viral proteins, VPgs are multifunctional. They have been shown to play a role in key steps of the viral

A supplementary table is available with the online version of this paper.

cycle: replication, translation and cell-to-cell movement. These functions can be performed by mature VPgs and/or their precursors. Processing of the VPg precursors is one of the possibilities by which to regulate VPg multifunctionality. Moreover, it has been shown that VPgs can directly regulate the protease activity as SeMV protease is active *in trans* only in fusion with VPg (Satheshkumar *et al.*, 2005). To perform their various functions, VPgs establish interactions with several viral or host partners such as VPg itself, nuclear inclusion protein b, helper component protease, cylindrical inclusion protein, coat protein or eukaryotic translation initiation factors: eIF4E, eIF4G, eIF4A, eIF3 and the poly(A)-binding protein (Daughenbaugh *et al.*, 2003, 2006; Goodfellow *et al.*, 2005; Hébrard *et al.*, 2010; Khan *et al.*, 2008; Lin *et al.*, 2009; Michon *et al.*, 2006; Miyoshi *et al.*, 2006). For RYMV, an interaction of VPg with eIF(iso)4G is known to be crucial for virus infection (Albar *et al.*, 2006; Hébrard *et al.*, 2006, 2010). Recently, it was demonstrated that the VPg of another sobemovirus SeMV is not required for the negative-strand synthesis *in vitro* (Govind & Savithri, 2010).

Structural features of VPgs are also involved in their abilities to interact with several partners. For Potato virus A (PVA), Potato virus Y, Lettuce mosaic virus, SeMV and RYMV an unfolded/disordered structure of VPg has been described previously (Grzela *et al.*, 2008; Hébrard *et al.*, 2009; Rantalainen *et al.*, 2008; Satheshkumar *et al.*, 2005). VPg proteins lack a unique 3D-structure and exist as a dynamic ensemble of conformations. High-resolution structural data are limited to small VPgs of about 20 residues. The 3D structures of synthetic peptides corresponding to VPgs in complex with viral RNA-dependent RNA polymerase from members of the family *Picornaviridae* are the only ones available to date (Gruez *et al.*, 2008; Schein *et al.*, 2006). Although RYMV VPg contains disordered domains in its C-terminal half, a folding into an α-helical conformation can be induced in experimental conditions (Hébrard *et al.*, 2009). The central α-helix is involved in the interaction with eIF(iso)4G (Hébrard *et al.*, 2008, 2010). For CfMV, similar intrinsically disordered and helical domains have been predicted (Hébrard *et al.*, 2009).

In the current study, a mass spectrometry (MS)-based approach was taken to determine the C-terminal processing site of virion-purified VPgs. As a result we identified the C termini of CfMV and RYMV VPgs, determined the residues to which viral RNA was covalently linked to and discovered post-translational modifications (PTMs) of the sobemoviral VPgs.

RESULTS

Identification of the VPg C termini

Mature VPgs purified from CfMV and RYMV virions were trypsin-digested and studied with tandem MS analysis. The sequence coverage of CfMV and RYMV VPgs was 100 %, i.e. there was MS data for every tryptic peptide between the determined termini of the proteins (Supplementary Table S1, available in JGV Online). We confirmed that RYMV VPg is 79 residues in length, spanning from residue 327 to 405 in polyprotein P2a (Fig. 1). The CfMV VPg C terminus was determined to be at position 396 of P2a and the N terminus at position 318, resulting in a mature VPg protein of 78 aa residues in length. The N termini of VPgs are cleaved between E/N for CfMV and E/S for RYMV residues, as described previously (Hébrard *et al.*, 2008; Mäkinen *et al.*, 2000), and the C termini are cleaved between E/T residues.

Description of the VPg–RNA linkage

The position to which the viral RNA is linked to was deduced from the *de novo* interpretation of previously unmatched MS2 spectra. A peptide with molecular mass of 1944.7081 Da had a partially matching MS2 spectrum with the theoretical CfMV VPg N-terminal peptide with a molecular mass of 1519.6903 Da. Thus, the peptide contained a modification with the molecular mass of +425.0178 Da, corresponding to pGp (monoisotopic mass of 425.0138 Da), a product of acidic RNA degradation. As G is also the first nucleotide of the CfMV genome, these possibilities were included in the analysis parameters and the modification was pinpointed to a tyrosine at position five (Fig. 2a). For RYMV, a similar approach was taken. The theoretical mass of the N-terminal peptide of RYMV VPg is 939.4702 Da. The first nucleotide of RYMV genome is A; therefore, the corresponding modification would be pAp with monoisotopic mass of 409.0189 Da and the mass of
the N-terminal peptide with the modification would be 1348.4890 Da. A precursor peptide with that mass (within the instrument mass accuracy of 5 p.p.m.) was indeed detected, and from the fragmentation spectra the modification was assigned to the serine at position one (Fig. 2b).

Characterization of VPg phosphorylation

The MS/MS analysis provided evidence that CfMV and RYMV VPg contained a number of PTMs. At least two phosphorylation sites were found for each VPg. A threonine at position 20 and serine at position 71 (Fig. 2c, d) were found to be present in both phosphorylated as

318 / 1 NSEL**Y**^{RNA}PDQSSGPARELDAE**T**^{PHOS}YTERLEQGIAFTEYNISGITV
KTSDREWTTAEALRVARYKPLGGGKAWGD**S**^{PHOS}DDEDTQE 396 / 78

327 / 1 **S**^{RNA}PFEIYGKFREAN**S**^{PHOS}EEYDESLRHGVLYAEYDFSGDTIRAS**S**^{PHOS}
NTWVRERTKYHAEERRKSGQLSWADRFGDD**S**^{PHOS}GEDVDIE 405 / 79

Fig. 1. Mass-spectrometric detection of CfMV and RYMV VPgs. Amino acid sequence and PTMs of mature VPgs. Numeration corresponds to P2a polyprotein and VPg. Modifications are indicated in superscript: RNA, link to 5′ of viral RNA; PHOS, phosphorylation.

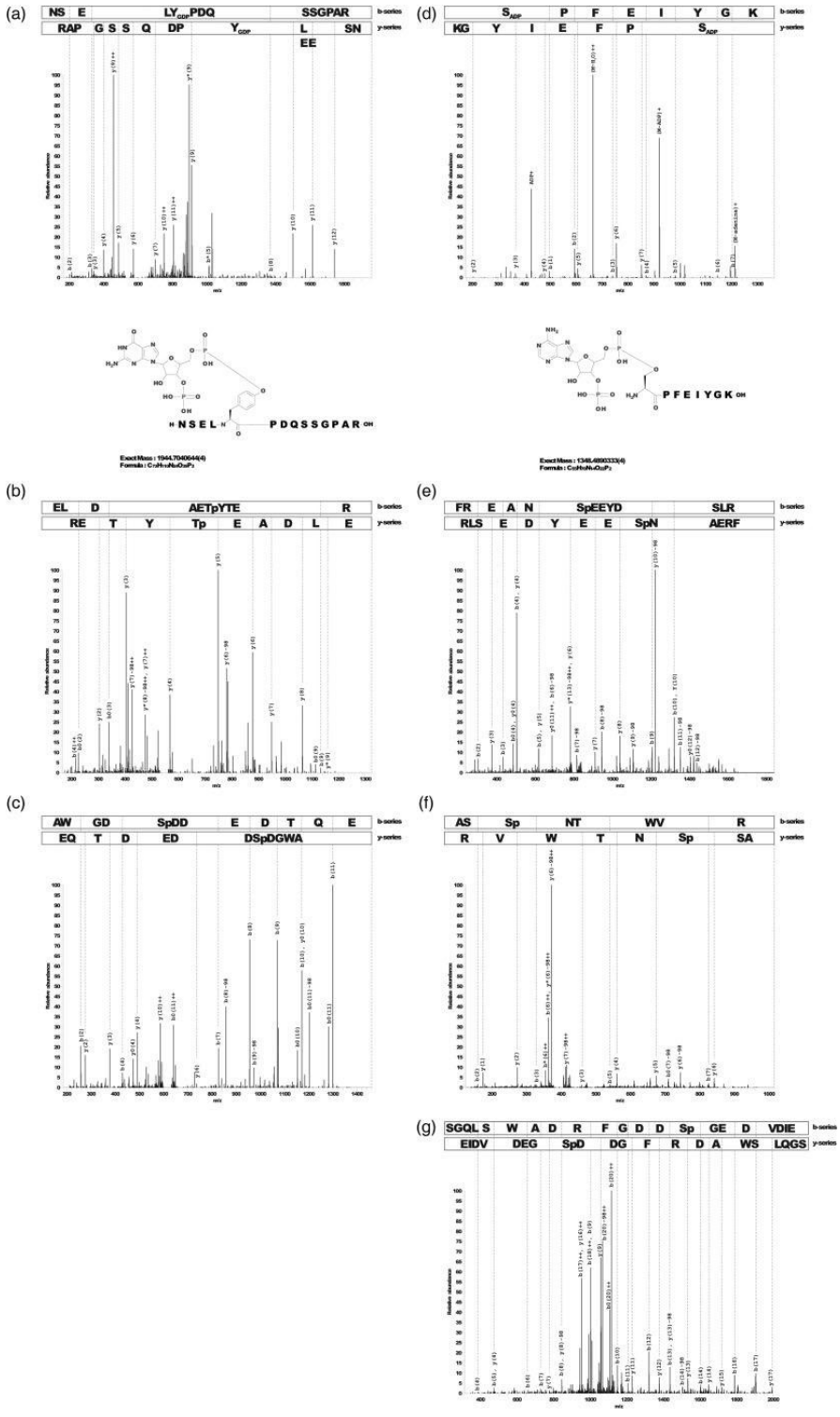

(a)
NS E LY$_{GDP}$PDQ SSGPAR
RAP G S S Q DP Y$_{GDP}$ L SN
EE
b-series
y-series
Relative abundance
m/z
NSEL PDQSSGPAR
Exact Mass : 1944.7040644(4)
(d)
S$_{ADP}$ P F E I Y G K
KG Y I E F P S$_{ADP}$
PFEIYGK
Exact Mass : 1348.4893333(4)
(b)
EL D AETpYTE R
RE T Y Tp E A D L E
(c)
AW GD SpDD E D T Q E
EQ T D ED DSpDGWA
(e)
FR E A N SpEEYD SLR
RLS E D Y E E SpN AERF
(f)
AS Sp NT WV R
R V W T N Sp SA
(g)
SGQL S W A D R F G D D Sp GE D VDIE
EIDV DEG SpD DG F R D A WS LQGS

Fig. 2. Identification of PTMs of CfMV and RYMV VPg by MS/MS analysis. Co-purified VPg linked to RNA was trypsin-digested and RNA was degraded with acidic hydrolysis. The peptides were analysed by nano-LC/MS/MS and resulting data were searched against corresponding sequence databases by MASCOT. The b and y ions represent N- and C-terminal fragment ions produced by mass spectrometry. (a) Determination of the residue covalently linked to RNA. The N-terminal CfMV VPg peptide was determined to contain a Y5-linked pGp modification, a corresponding degradation product of viral RNA. The peptide with the modification is represented below fragmentation spectrum. (b, c) Determination of phosphorylation sites within CfMV VPg. Two peptides containing phosphorylations were detected, (b) a threonine in position 20 [ELDAE(T)YTER] and (c) a serine in position 71 [AWGD(S)DDEDTQE]. (d) The N-terminal RYMV VPg peptide was determined to contain an S1-linked pAp modification. The peptide with the modification is represented below fragmentation spectrum. (e–g) Determination of phosphorylation sites within RYMV VPg. Three peptides containing phosphorylations were detected, serines in positions (e) 14 [FREAN(S)EEYDESLR], (f) 41 [AS(S)NTWVR] and (g) 72 [SGQLSWADRFGDD(S)GEDVDIE].

well as non-phosphorylated forms in CfMV VPg. No phosphothreonines were found in RYMV VPg, at the same time serines at positions 14, 41 and 72 (Fig. 2e–g) were found to be phosphorylated. Precursor peptide ratios of unmodified and phosphorylated forms varied between samples and depended greatly on preparation and handling (data not shown), making any attempt to quantitatively assess the abundance of phosphorylation futile. The analysis of another isolate of RYMV (isolate CI4 from Cote d'Ivoire) again confirmed the phosphorylation of serines at positions 14, 41 and 72 (data not shown). With RYMV, we also detected phosphorylation of serines 33 and 59 (data not shown), but with low confidence and/or from only one biological sample. Therefore, it is possible that these positions might also be phosphorylated.

It is worthwhile to note that, depending on sample preparation, peptides were detected with a +28 Da or the multiple of +28 Da modification(s) which was assigned by MASCOT as one or several random aspartate and/or glutamate ethylation(s) (data not shown), a modification that can be introduced *in vitro* during sample preparation (Xing *et al.*, 2008). The solutions used for RNA extraction with columns contained ethanol and these modifications indeed occurred only when RNA columns were used for RNA extraction. Furthermore, in some VPg batches tryptophan residues were found to be oxidized or di-oxidized (data not shown), which is also known to be an *in vitro* generated modification (Stadtman & Levine, 2003).

Estimation of the selection pressure on RYMV modification sites

The variability and evolution of the phosphorylated codons of the VPg were assessed from a dataset representative of the genetic diversity and the geographical origin of RYMV (Pinel-Galzi *et al.*, 2009). The selection pressure expressed on the sites of the phosphorylated amino acids and of the amino acid involved in protein–RNA linkage was estimated by three maximum-likelihood methods: Fixed Effect Likelihood (FEL), Internal Fixed Effect Likelihood (IFEL) and Single Likelihood Ancestor Counting (SLAC). The three methods gave similar results. Positions 1, 14, 33 and 41 are under significant negative-selection pressure with the following *P*-values (obtained with the FEL method and not substantially different from the two other methods). Site 1, $P=4\times10^{-3}$; site 14, $P=2\times10^{-5}$; site 33, $P=3\times10^{-2}$ and site 41, $P=9\times10^{-7}$. It means that not only are these positions conserved at the amino acid level but that there is evidence of strong conservative selective pressure at codon 1, 14, 33 and 41. Codon 59 is invariable at its three positions. Position 72 is the exception, being polymorphic at the amino acid level. Most RYMV isolates have a serine at position 72 of their VPg. However, one S6 strain, widely spread in Eastern Tanzania, had an aspartic acid instead of a serine at this position. The change from serine (AGU) to aspartic acid (GAU) at codon 72 involved mutations at the first and second positions of the codon, but there is no evidence that codon 72 is under diversifying selection. From the large sample examined, codon 72 was found to be under neutral evolution.

DISCUSSION

Due to their vast repertoire of functions, VPg proteins have long been in the focus of interest. When the CfMV genome was completely sequenced (Mäkinen *et al.*, 1995), several putative E/T processing sites within the polyprotein were proposed, based on previous analysis of sobemoviral proteases (Gorbalenya *et al.*, 1988). However, when the sequence of the N terminus of the CfMV VPg was determined, an approximate mass for the protein was determined with SDS-PAGE to be 12 kDa (Mäkinen *et al.*, 2000). Based on the mass observed, the C terminus of VPg was proposed to be situated downstream of the −1 PRF signal and previously predicted processing sites. Firstly, our current results confirm the previously described N terminus of CfMV VPg. The data determined the CfMV VPg C-terminal processing site E^{396}/T^{397} upstream of the −1 PRF signal. The C terminus of the CfMV VPg is in accordance with the experimentally demonstrated C terminus of the VPg of SeMV (Nair & Savithri, 2010). The previously described molecular mass of 12 kDa was more likely a result of shift in mobility in SDS-PAGE caused by the acidic nature of the VPg protein (pI ~4). Such abnormal mobility during denaturing electrophoresis of intrinsically disordered and acidic proteins has been reported previously (Receveur-Bréchot *et al.*, 2006). Our data demonstrate that the −1 PRF mechanism is not involved in the synthesis of CfMV and RYMV VPgs and that both viruses encode a single VPg as part of P2a.

The theoretical molecular masses of CfMV and RYMV VPgs are 8.6 and 9.2 kDa, respectively. Their amino acid compositions are characterized by a low proportion of hydrophobic residues (22 and 24 %) and a high proportion (32 and 39 %) of charged residues compared with globular proteins (hydrophobic 34 % and charged 23 %), respectively. CfMV and RYMV VPgs contain 17 and 20 acidic amino acids distributed along the proteins. Such a compositional bias is a characteristic of intrinsically disordered proteins.

We identified that the viral RNAs of CfMV and RYMV are attached to a tyrosine residue at position 5 and to a serine residue at position 1, respectively. Our findings are supported by the fact that previously the identity of CfMV residue 5 and RYMV residue 1 could not be confirmed by Edman sequencing (Hébrard *et al.*, 2008; Mäkinen *et al.*, 2000). This is the first characterization of VPg–RNA linkage for sobemoviruses and the first VPg–RNA linkage mapped by using the MS approach. Furthermore, to our knowledge this is also the first report about the use of a different amino acid residue for RNA linkage within one genus. CfMV and RYMV both infect monocotyledonous hosts and are genetically closely related (Fig. 3). Usually the residue is conserved within the family and cannot be substituted by another residue (Carette *et al.*, 2001; Murphy *et al.*, 1996). It appears that within the sobemovirus genera the RNA linking is species-specific. For both viruses, the residue used for RNA linking is followed by a proline. Interestingly, there is a serine at position 2 in CfMV (vs position 1 for RYMV) and a tyrosine at position 6 for RYMV (vs position 5 for CfMV). Imperata yellow mottle virus (IYMV) – which is the closest species to RYMV – also has a serine at position 1 (followed by a proline), but no tyrosine within the first 15 aa, suggesting that serine 1 is the IYMV protein–RNA linkage site. This is also supported by our preliminary results (data not shown). Due to high diversity between sobemovirus VPg sequences (Fig. 3), it is impossible to predict the linkage site for other members of the genera. Out of 11 members only five contain tyrosines within the first 16 aa residues; however, several of them contain serines. For some members, the N-terminal residue of VPg is a threonine, which allows us to hypothesize that sobemoviruses might even link RNA to threonine. It is interesting to note that the threonine in position 1 of SBMV VPg was the only residue out of the first 20 that was not detected correctly by Edman sequencing (van der Wilk *et al.*, 1998).

Several phosphorylated residues were identified in CfMV and RYMV VPgs. For PVA, the phosphorylation of VPg has also been demonstrated (Puustinen *et al.*, 2002) and it is believed to be involved in the regulation of host interactions. The mature VPgs of sobemoviruses contain PTMs with a certain degree of diversity. The sequence context of CfMV and RYMV phosphorylation sites T20/S14 and S71/S72 is not itself conserved, but the position they occupy in VPg is similar (Fig. 3). Since the VPgs are disordered proteins it is possible that the position and/or distance between the phosphorylation sites is more important than primary

Fig. 3. Sequence comparison of sobemoviruses and VPgs. CLUSTAL W was used for aligning VPg sequences. The phylogenetic tree was inferred by the maximum composite likelihood model from the full nucleotide sequences of sobemoviruses. Mid-point rooting was applied. The tyrosine no. 5 and serine no. 1 covalently linked to CfMV and RYMV RNA, respectively, are indicated by bold and italics and the phosphorylated threonine and serine residues are in bold. WAD/WGD followed by multiple E/D residues is the only conserved motif among sobemoviral VPgs and is underlined. CfMV, Cocksfoot mottle virus (accession no. NP_942019); RYMV, Rice yellow mottle virus (accession no. CAE81344); TRoV, Turnip rosette virus (accession no. NP_942000); IYMV, Imperata yellow mottle virus (accession no. CAQ48412); SCMoV, Subterranean clover mottle virus (accession no. NP_736580); LTSV, Lucerne transient streak virus (accession no. NP_736595); RGMoV, Ryegrass mottle virus (accession no. NP_736586); RuCMV, Rubus chlorotic mottle virus (accession no. CAP79624); SBMV, Southern bean mosaic virus (accession no. NP_736583); SeMV, Sesbania mosaic virus (accession no. NP_736592); SCPMV, Southern cowpea mosaic virus (accession no. NP_736598). *, Indicates invariant; ., indicates similar; :, indicates highly similar.

sequence. In summary, it is possible that these positions and their phosphorylation bear similar roles for both viruses. Both detected phosphorylation sites of CfMV VPg and the S72 of RYMV VPg correspond to the protein kinase CK2 (casein kinase II) consensus motif S/TXXD/E (reviewed by Meggio & Pinna, 2003).

VPg sequences of four CfMV isolates and 150 RYMV isolates are currently available. Except for S72 of RYMV, there are no coding differences for both viruses between the isolates in the RNA linkage and in the phosphorylation sites at the amino acid level despite several synonymous nucleotide substitutions. A strong selection pressure is expressed on the site of the amino acid involved in protein–RNA linkage and on most sites of the phosphorylated amino acids. In contrast, position 72 exhibits amino acid polymorphism and is neither under conservative or diversifying selection. One variant of S6 strain with an aspartic acid at position 72 co-exists with another variant of S6 strain with a serine in the same geographical region of

Eastern Tanzania. This indicates that a change from a phosphorylated serine to an aspartic acid is not lethal, and apparently not even disadvantageous to this strain. This result is not at variance with the hypothesis of conservation of phosphorylated amino acids as aspartic acid has physico-chemical properties close to a phosphorylated serine, which might explain the fitness of this strain. The analysis of another isolate of RYMV (isolate CI4) not only confirmed the phosphorylation of serines at positions 14, 41 and 72, but suggests that phosphorylation is independent of the genetic context. The two isolates belong to quite different strains of West Africa, CIa to S3 and CI4 to S1. For instance, the diversity between the two isolates in their VPg (and their flanking regions; a total of 540 nt) is 7.5 %.

Due to the disordered nature and the propensity to form structures upon stabilization (Hébrard *et al.*, 2009), which can occur during an interaction, dynamics of VPg function dependent on phosphorylation can be proposed here. The regulation of folding/unfolding and interaction determination of disordered proteins/disordered domains by phosphorylation has been widely reported (Mittag *et al.*, 2010; Stein *et al.*, 2009; Wright & Dyson, 2009). For example, VPg PTMs can be used to switch between the required functionality necessary at different stages of the viral replication cycle. Phosphorylations as reversible modifications are well known to regulate processes, including replication, in the viral multiplication cycle (Jakubiec & Jupin, 2007).

We conclude that the VPg is a multifunctional protein and the precise biological relevance and function(s) of each identified PTM remains to be determined in the future.

METHODS

For the infection, 2-week-old oat cv. Jaak and rice cv. IR64 plants were mechanically inoculated with CfMV (Norwegian isolate) and RYMV (isolate CIa from Cote d'Ivoire, strain S3), respectively. Complementary experiments were performed with the RYMV isolate CI4 (strain S1). After 4–5 weeks, the leaves were harvested and virus particles were purified by ultracentrifugation as described by Tars *et al.* (2003). Virions were dissociated with 1 % SDS and the RNA was isolated with RNeasy Plant RNA kit (Qiagen) or standard phenol/chloroform extraction. VPg, covalently bound to the RNA, was trypsin-digested in 50 mM ammonium bicarbonate buffer and subsequently the RNA was hydrolyzed in 10 % trifluoroacetic acid for 48 h at room temperature. For some samples, phosphatase inhibitor cocktail (Roche) was included. The samples were then dried under vacuum, purified with C18 StageTips (Rappsilber *et al.*, 2007) and analysed by LC-MS/MS using an Agilent 1200 series nanoflow system (Agilent Technologies) connected to a LTQ Orbitrap classic mass-spectrometer (Thermo Electron) equipped with a nanoelectrospray ion source (Proxeon). In short, purified peptides were dissolved in 0.5 % formic acid and loaded onto a fused silica emitter (150 mm × 0.075 mm; Proxeon) packed in-house with Repropur-Sil C18-AQ 3 μm particles (Dr Maisch, HPLC, GmbH) using a flow rate of 700 nl min^{-1}. Peptides were separated with a gradient from 3 to 40 % B (A: 0.5 % acetic acid, B: 0.5 % acetic acid/80 % acetonitrile) using a flow-rate of 200 nl min^{-1} and sprayed directly into the LTQ Orbitrap mass-spectrometer (Thermo Electron) operated at 180 °C capillary temperature and 2.4 kV spray voltage. LTQ Orbitrap was operated in the data-dependent mode with a full scan in the Orbitrap followed by up to five MS/MS scans in the LTQ part of the instrument. Precursor ion spectra (m/z 300–1900) were acquired in the Orbitrap (profile mode, resolution R=60 000, target value 1×10^6 ions); up to five data-dependent MS/MS spectra were acquired in the LTQ for each precursor ion scan (centroid mode, normalized collision energy 35 %, wideband activation enabled, target value 5000 ions). Fragment MS/MS spectra from raw files were extracted as MSM files and then merged to peak lists using Raw2MSM version 1.11 (Olsen *et al.*, 2005) selecting the top six peaks for 100 Da. MSM files were searched with the MASCOT 2.2 (Perkins *et al.*, 1999) search engine (Matrix Science) against the protein sequence database composed of VPg sequences and common contaminant proteins such as trypsin, keratins etc. To allow for the determination of VPg C termini, the VPg sequence spanning the hypothetical cleavage site was shortened by a single amino acid in its C terminus to create 20 entries of different lengths in the database. Search parameters were as follows: 5 p.p.m. precursor mass tolerance and 0.6 Da MS/MS mass tolerance, two missed trypsin cleavages plus a number of variable modifications such as oxidation (M), oxidation (HW), ethyl (DE), phospho (ST), phospho (Y), ADP (SY), GDP (SY). ADP (SY) and GDP (SY) modifications were custom-defined in MASCOT. In addition to a MASCOT search some .raw files were also *de novo* interpreted with PEAKS v4.5 (Ma *et al.*, 2003). For both viruses at least three independent biological samples were analysed.

The selection pressure expressed on the RYMV sites of the phosphorylated amino acids and of the amino acid involved in protein–RNA linkage was estimated. The ratio of non-synonymous (d_N) over synonymous (d_S) substitutions in the VPg of RYMV was calculated on a corpus of RYMV isolates representative of the geographical distribution and the genetic diversity of the virus (150 isolates from 16 countries of Africa). Three maximum-likelihood methods, FEL, IFEL and SLAC implemented in DataMonkey (http://www.datamonkey.org/) were applied (Kosakovsky Pond & Frost, 2005a, 2005b). On each codon, it determines whether the selection pressure is conservative ($d_N/d_S < 1$), diversifying ($d_N/d_S > 1$) or neutral ($d_N/d_S = 1$). The analyses were conducted with the VPg sequences (240 nt), plus its flanking regions (nt 1526–2065; 540 nt altogether) in order to increase the statistical significance of the tests.

ACKNOWLEDGEMENTS

We wish to thank Indrek Tammiste for his initial efforts involving MS, Jaanus Remme for consultations on determining a correct RNA degradation product attached to VPg, Lilian Järvekülg for consultations on virus particle purification and Signe Nõu for excellent plant care. We also wish to thank Agnes Pinel-Galzi for providing us the VPg sequences of 150 RYMV isolates for the selection pressure analysis. This work was supported by Estonian Science Foundation grant no. 7363 and PHC Parrot programme grant no. 20674ZG supporting Estonian–French scientific collaboration. Mass-spectrometric analyses were in part supported by the European Regional Development Fund through the Center of Excellence in Chemical Biology (Institute of Technology, University of Tartu).

REFERENCES

Albar, L., Bangratz-Reyser, M., Hébrard, E., Ndjiondjop, M., Jones, M. & Ghesquière, A. (2006). Mutations in the eIF(iso)4G translation initiation factor confer high resistance of rice to *Rice yellow mottle virus*. *Plant J* **47**, 417–426.

Ambros, V. & Baltimore, D. (1978). Protein is linked to the 5′ end of poliovirus RNA by a phosphodiester linkage to tyrosine. *J Biol Chem* **253**, 5263–5266.

Anindya, R., Chittori, S. & Savithri, H. S. (2005). Tyrosine 66 of *Pepper vein banding virus* genome-linked protein is uridylylated by RNA-dependent RNA polymerase. *Virology* **336**, 154–162.

Belliot, G., Sosnovtsev, S. V., Chang, K. O., McPhie, P. & Green, K. Y. (2008). Nucleotidylylation of the VPg protein of a human norovirus by its proteinase-polymerase precursor protein. *Virology* **374**, 33–49.

Carette, J. E., Kujawa, A., Gühl, K., Verver, J., Wellink, J. & Van Kammen, A. (2001). Mutational analysis of the genome-linked protein of cowpea mosaic virus. *Virology* **290**, 21–29.

Daughenbaugh, K. F., Fraser, C. S., Hershey, J. W. B. & 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**, 2852–2859.

Daughenbaugh, K. F., Wobus, C. E. & Hardy, M. E. (2006). VPg of murine norovirus binds translation initiation factors in infected cells. *Virol J* **3**, 33.

Ghosh, A., Rutgers, T., Ke-Qiang, M. & Kaesberg, P. (1981). Characterization of the coat protein mRNA of Southern bean mosaic virus and its relationship to the genomic RNA. *J Virol* **39**, 87–92.

Goodfellow, I., Chaudhry, Y., Gioldasi, I., Gerondopoulos, A., Natoni, A., Labrie, L., Laliberté, J. & Roberts, L. (2005). Calicivirus translation initiation requires an interaction between VPg and eIF4E. *EMBO Rep* **6**, 968–972.

Gorbalenya, A. E., Koonin, E. V., Blinov, V. M. & Donchenko, A. P. (1988). Sobemovirus genome appears to encode a serine protease related to cysteine proteases of picornaviruses. *FEBS Lett* **236**, 287–290.

Govind, K. & Savithri, H. S. (2010). Primer-independent initiation of RNA synthesis by SeMV recombinant RNA-dependent RNA polymerase. *Virology* **401**, 280–292.

Gruez, A., Selisko, B., Roberts, M., Bricogne, G., Bussetta, C., Jabafi, I., Coutard, B., De Palma, A. M., Neyts, J. & other authors (2008). The crystal structure of coxsackievirus B3 RNA-dependent RNA polymerase in complex with its protein primer VPg confirms the existence of a second VPg binding site on *Picornaviridae* polymerases. *J Virol* **82**, 9577–9590.

Grzela, R., Szolajska, E., Ebel, C., Madern, D., Favier, A., Wojtal, I., Zagorski, W. & Chroboczek, J. (2008). Virulence factor of potato virus Y, genome-attached terminal protein VPg, is a highly disordered protein. *J Biol Chem* **283**, 213–221.

Hébrard, E., Pinel-Galzi, A., Bersoult, A., Siré, C. & 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**, 1369–1373.

Hébrard, E., Pinel-Galzi, A. & Fargette, D. (2008). Virulence domain of the RYMV genome-linked viral protein VPg towards rice *rymv*1–2-mediated resistance. *Arch Virol* **153**, 1161–1164.

Hébrard, E., Bessin, Y., Michon, T., Longhi, S., Uversky, V. N., Delalande, F., Van Dorsselaer, A., Romero, P., Walter, J. & other authors (2009). Intrinsic disorder in viral proteins genome-linked: experimental and predictive analyses. *Virol J* **6**, 23.

Hébrard, E, Poulicard, N, Gérard, C, Traoré, O, Albar, L, Fargette, D, Bessin, Y & Vignols, F. (2010). Direct interaction between the *Rice yellow mottle virus* VPg and the central domain of the rice eIF(iso)4G1 factor correlates with rice susceptibility and RYMV virulence. *Mol Plant Microbe Interact* **23**, 1506–1513.

Jaegle, M., Wellink, J. & Goldbach, R. (1987). The genome-linked protein of *Cowpea mosaic virus* is bound to the 5′ terminus of virus RNA by phosphodiester linkage to serine. *J Gen Virol* **68**, 627–632.

Jakubiec, A. & Jupin, I. (2007). Regulation of positive-strand RNA virus replication: the emerging role of phosphorylation. *Virus Res* **129**, 73–79.

Khan, M. A., Miyoshi, H., Gallie, D. R. & Goss, D. J. (2008). Potyvirus genome-linked protein, VPg, directly affects wheat germ *in vitro* translation: interactions with translation initiation factors eIF4F and eIFiso4F. *J Biol Chem* **283**, 1340–1349.

Kosakovsky Pond, S. L. K. & Frost, S. D. W. (2005a). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**, 2531–2533.

Kosakovsky Pond, S. L. & Frost, S. D. W. (2005b). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* **22**, 1208–1222.

Lin, L., Shi, Y., Luo, Z., Lu, Y., Zheng, H., Yan, F., Chen, J., Chen, J., Adams, M. J. & other authors (2009). Protein–protein interactions in two potyviruses using the yeast two-hybrid system. *Virus Res* **142**, 36–40.

Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A. & Lajoie, G. (2003). PEAKS: powerful software for peptide *de novo* sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom* **17**, 2337–2342.

Mäkinen, K., Tamm, T., Naess, V., Truve, E., Puurand, Ü., Munthe, T. & Saarma, M. (1995). Characterization of cocksfoot mottle sobemovirus genomic RNA and sequence comparison with related viruses. *J Gen Virol* **76**, 2817–2825.

Mäkinen, K., Mäkeläinen, K., Arshava, N., Tamm, T., Merits, A., Truve, E., Zavriev, S. & Saarma, M. (2000). Characterization of VPg and the polyprotein processing of cocksfoot mottle virus (genus *Sobemovirus*). *J Gen Virol* **81**, 2783–2789.

Mang, K. Q., Ghosh, A. & Kaesberg, P. (1982). A comparative study of the cowpea and bean strains of southern bean mosaic virus. *Virology* **116**, 264–274.

Meggio, F. & Pinna, L. A. (2003). One-thousand-and-one substrates of protein kinase CK2? *FASEB J* **17**, 349–368.

Michon, T., Estevez, Y., Walter, J., German-Retana, S. & Le Gall, O. (2006). The potyviral virus genome-linked protein VPg forms a ternary complex with the eukaryotic initiation factors eIF4E and eIF4G and reduces eIF4E affinity for a mRNA cap analogue. *FEBS J* **273**, 1312–1322.

Mittag, T., Kay, L. E. & Forman-Kay, J. D. (2010). Protein dynamics and conformational disorder in molecular recognition. *J Mol Recognit* **23**, 105–116.

Miyoshi, H., Suehiro, N., Tomoo, K., Muto, S., Takahashi, T., Tsukamoto, T., Ohmori, T. & Natsuaki, T. (2006). Binding analyses for the interaction between plant virus genome-linked protein (VPg) and plant translational initiation factors. *Biochimie* **88**, 329–340.

Murphy, J. F., Rychlik, W., Rhoads, R. E., Hunt, A. G. & 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**, 511–513.

Murphy, J. F., Klein, P. G., Hunt, A. G. & Shaw, J. G. (1996). Replacement of the tyrosine residue that links a potyviral VPg to the viral RNA is lethal. *Virology* **220**, 535–538.

Nair, S. & Savithri, H. S. (2010). Processing of SeMV polyproteins revisited. *Virology* **396**, 106–117.

Olsen, J. V., de Godoy, L. M. F., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S. & other authors (2005). Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* **4**, 2010–2021.

Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567.

Pinel-Galzi, A., Mpunami, A., Sangu, E., Rakotomalala, M., Traoré, O., Sérémé, D., Sorho, F., Séré, Y., Kanyeka, Z. & other authors (2009). Recombination, selection and clock-like evolution of *Rice yellow mottle virus*. *Virology* **394**, 164–172.

Puustinen, P., Rajamäki, M., Ivanov, K. I., Valkonen, J. P. T. & Mäkinen, K. (2002). Detection of the potyviral genome-linked protein VPg in virions and its phosphorylation by host kinases. *J Virol* **76**, 12703–12711.

Rantalainen, K. I., Uversky, V. N., Permi, P., Kalkkinen, N., Dunker, A. K. & Mäkinen, K. (2008). Potato virus A genome-linked protein VPg is an intrinsically disordered molten globule-like protein with a hydrophobic core. *Virology* **377**, 280–288.

Rappsilber, J., Mann, M. & Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**, 1896–1906.

Receveur-Bréchot, V., Bourhis, J., Uversky, V. N., Canard, B. & Longhi, S. (2006). Assessing protein disorder and induced folding. *Proteins* **62**, 24–45.

Rothberg, P. G., Harris, T. J., Nomoto, A. & 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**, 4868–4872.

Satheshkumar, P. S., Gayathri, P., Prasad, K. & Savithri, H. S. (2005). Natively unfolded VPg is essential for *Sesbania mosaic virus* serine protease activity. *J Biol Chem* **280**, 30291–30300.

Schein, C. H., Oezguen, N., Volk, D. E., Garimella, R., Paul, A. & Braun, W. (2006). NMR structure of the viral peptide linked to the genome (VPg) of poliovirus. *Peptides* **27**, 1676–1684.

Stadtman, E. R. & Levine, R. L. (2003). Free radical-mediated oxidation of free amino acids and amino acid residues in proteins. *Amino Acids* **25**, 207–218.

Stein, A., Pache, R. A., Bernadó, P., Pons, M. & Aloy, P. (2009). Dynamic interactions of proteins in complex networks: a more structured view. *FEBS J* **276**, 5390–5405.

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

van der Wilk, F., Verbeek, M., Dullemans, A. & van den Heuvel, J. (1998). The genome-linked protein (VPg) of southern bean mosaic virus is encoded by the ORF2. *Virus Genes* **17**, 21–24.

Wright, P. E. & Dyson, H. J. (2009). Linking folding and binding. *Curr Opin Struct Biol* **19**, 31–38.

Xing, G., Zhang, J., Chen, Y. & Zhao, Y. (2008). Identification of four novel types of *in vitro* protein modifications. *J Proteome Res* **7**, 4603–4608.

Zalloua, P. A., Buzayan, J. M. & Bruening, G. (1996). Chemical cleavage of 5′-linked protein from tobacco ringspot virus genomic RNAs and characterization of the protein-RNA linkage. *Virology* **219**, 1–8.

PUBLICATION III

Viral RNA linked to VPg over a threonine residue

Olspert, A., Arike, L., Peil, L., Truve, E.

FEBS Letters (2011), 585, 2979 - 2985.

FEBS Letters 585 (2011) 2979–2985

FEBS Letters

journal homepage: www.FEBSLetters.org

Sobemovirus RNA linked to VPg over a threonine residue

Allan Olspert [a], Liisa Arike [b], Lauri Peil [b,c,d], Erkki Truve [a,*]

[a] Department of Gene Technology, Tallinn University of Technology, Akadeemia tee 15, 12618 Tallinn, Estonia
[b] Institute of Technology, University of Tartu, Nooruse 1, 50411 Tartu, Estonia
[c] Estonian Biocentre, Riia 23b, 51010 Tartu, Estonia
[d] Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3JR, United Kingdom

ARTICLE INFO

Article history:
Received 30 June 2011
Revised 3 August 2011
Accepted 4 August 2011
Available online 17 August 2011

Edited by Tamas Dalmay

Keywords:
Southern bean mosaic virus
Ryegrass mottle virus
RNA linkage
Phosphorylation
Mass spectrometry
Viral protein genome linked
Protein RNA covalent bond

ABSTRACT

Positive sense ssRNA virus genomes from several genera have a viral protein genome-linked (VPg) attached over a phosphodiester bond to the 5′ end of the genome. The VPgs of *Southern bean mosaic virus* (SBMV) and *Ryegrass mottle virus* (RGMoV) were purified from virions and analyzed by mass spectrometry. SBMV VPg was determined to be linked to RNA through a threonine residue at position one, whereas RGMoV VPg was linked to RNA through a serine also at the first position. In addition, we identified the termini of the corresponding VPgs and discovered three and seven phosphorylation sites in SBMV and RGMoV VPgs, respectively. This is the first report on the use of threonine for linking RNA to VPg.

1. Introduction

The 5′ ends of single-stranded positive-sense RNA virus genomes are unmodified, capped or have a viral protein genome-linked (VPg). The VPgs are attached to the RNA over a phosphodiester bond formed between the 5′ phosphate group of RNA and the hydroxyl group of an amino acid (aa) residue usually situated near the N-terminus of the protein. The aa residues involved in the VPg–RNA linkage have been reported to be either tyrosine or serine [1,2]. Picornaviruses utilize a highly conserved tyrosine residue situated near the N-terminus of VPg [1,3,4]. The use of tyrosine has also been shown for poty- and caliciviruses [5–7], whereas RNA linkage through a serine residue has been demonstrated for nepo- and comoviruses [2,8]. Recently we demonstrated that, dependent on the virus, sobemoviruses can utilize either tyrosine or serine [9]. Although threonine also contains a hydroxyl group, its use for linking RNA to VPg has not been reported.

Phosphodiester bonds exist also between DNA and proteins. Terminal proteins (TP) serve as primers for the synthesis of genomes and subsequent protective agents of genome termini of DNA viruses, mitochondrial plasmids and linear chromosomes [10,11]. Bond formation between TPs and DNA has been demonstrated to occur over the hydroxyl group of either serine, tyrosine or in that case, also threonine residues [10].

The VPgs of picornaviruses have been established mainly as primers for RNA synthesis [12], whereas potyviral VPgs are involved in the inhibition of cap-dependent and enhancement of cap-independent viral RNA translation [13].

Southern bean mosaic virus (SBMV) and Ryegrass mottle virus (RGMoV) are members of the genus Sobemovirus (reviewed in [14], genome organization revised in [15]), a group of small spherical viruses with a positive-sense single stranded RNA genome of approximately 4–4.5 kb. In sobemoviruses, the VPg is cleaved from the polyprotein by the viral protease [9,16–19] and is covalently attached to the 5′ end of genomic and subgenomic RNAs [20,21].

Sobemoviral VPgs are not conserved and unique in sense of RNA linking as they use either tyrosine or serine residues [9]. All this makes it difficult to predict the aa residue responsible for RNA linking for each virus species. The predicted RGMoV VPg does not have a suitable aa residue for linking near the N-terminus, whereas for SBMV VPg sequence comparisons suggest threonine as the most likely candidate [9]. We analyzed the VPg–RNA linkage of these two sobemoviruses. As a result we identified the true cleavage sites of RGMoV VPg, demonstrated that threonine is indeed used for linking RNA to SBMV VPg and described the post-translational modifications of these two VPgs.

* Corresponding author. Fax: +372 6204401.
E-mail address: erkki.truve@ttu.ee (E. Truve).

0014-5793/$36.00 © 2011 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.febslet.2011.08.009

2. Materials and methods

For the infection, 2 week old oat (*Avena sativa* cv. Jaak) and bean (*Phaseolus vulgaris* cv. Sonesta or Aura) plants were mechanically inoculated with RGMoV (Japanese isolate, PV-307043 obtained from MAFF GenBank) and SBMV (Colombian isolate, PV-0100 obtained from DSMZ), respectively. After 4–5 weeks the leaves were harvested, virus particles and VPg were purified as described [9]. Briefly, virions were purified by ultracentrifugation, dissociated and the RNA was isolated. VPg, covalently bound to the RNA, was trypsin digested and subsequently the RNA was hydrolyzed in 10% trifluoroacetic acid for 48 h at room temperature. The samples were then dried under vacuum, purified with StageTips [22] and analyzed by LC–MS/MS using an Agilent 1200 series nanoflow system (Agilent Technologies) connected to a LTQ Orbitrap mass-spectrometer (Thermo Electron) equipped with a nanoelectrospray ion source (Proxeon), as described before [9].

LTQ Orbitrap was operated in the data dependent mode with a full scan in the Orbitrap (mass range m/z 300–1900, resolution 60 000 at m/z 400, target value 1×10^6 ions) followed by up to five MS/MS scans in the LTQ part of the instrument (normalized collision energy 35%, wideband activation enabled, target value 5000 ions). Fragment MS/MS spectra from raw files were extracted as MSM files and then merged to peak lists using Raw2MSM version 1.11, selecting top eight peaks for each 100 Da [23]. MSM files were searched with the Mascot 2.3 search engine (Matrix Science) against the protein sequence database composed of VPg sequences and common contaminant proteins such as trypsin, keratins etc. Search parameters were as follows: 5 ppm precursor mass tolerance and 0.6 Da MS/MS mass tolerance, three missed trypsin cleavages plus a number of variable modifications such as oxidation (M), oxidation (HW), ethyl (DE), phospho (ST), phospho (Y), pAp (SYT), pGp (SYT), pCp (SYT) and pUp (SYT). For both viruses at least two independent biological samples were analyzed, each biological sample was in turn analyzed twice.

3. Results

3.1. Characterization of SBMV and RGMoV VPgs

VPgs purified from SBMV and RGMoV virions were trypsin-digested and studied with tandem MS analysis. The sequence coverage of SBMV VPg was 81–84%, identified peptides are shown in Table 1 and Fig. 1. We confirmed that the SBMV VPg is 77 aa residues in length, spanning from residue 326 to 402 in polyprotein P2a (Fig. 1A). Peptides between residues 30–38 and 65–69 were not detected, most likely due to their small size. Both N-terminal and C-terminal SBMV VPg protease cleavage sites were found to be between E/T residues.

The N-terminal peptide of RGMoV VPg was not detected when using the predicted annotated sequence as reference (accession NP_736586). Since sobemoviral proteases are known to cleave between E/T, E/S and E/N residues [16–18], we extended the N-terminal sequence to possible cleavage sites further upstream and discovered that the RGMoV VPg N-terminus is cleaved between E/S, three aa residues upstream of the previously proposed E/N site (Fig. 2A). The sequence coverage of RGMoV VPg was 91–94% and the length of the VPg was 79 aa residues. Identified peptides are shown in Table 1 and Fig. 2. The RGMoV VPg is cleaved from the P2a polyprotein between E/S residues at positions 314/315 and 393/394. Peptides between aa residue positions 33–36 of the RGMoV VPg were not detected in our study, similarly to short peptides from SBMV VPg.

3.2. SBMV and RGMoV VPg post-translational modifications

When searching for VPg–RNA linkage sites, we utilized the knowledge that the corresponding aa residue modification after RNA hydrolysis is a 5′,3′-diphosphate nucleotide, pNp (N denoting adenosine, cytidine, guanosine or uridine) [9], and expanded all these possible modifications to all possible phosphodiester bond acceptor residues (serine, tyrosine and threonine). By this approach we determined that the SBMV VPg was linked to RNA through a threonine residue at position one and the corresponding modification was pAp as assigned by modification delta mass and corresponding fragmentation spectrum (Fig. 1B). The RGMoV VPg was determined to be linked to RNA through a serine residue at position one and the modification was again assigned to be pAp (Fig. 2B). For both viruses the VPg N-terminal peptide was never detected in Mascot database search without the nucleotide modification.

In addition we identified several phosphorylation sites in both SBMV and RGMoV VPg-s (Table 1). In the SBMV VPg, serines at positions 7, 20 and 58 were found to be phosphorylated

Table 1
Examples of detected peptides identified by fragmentation spectra. The post-translational modifications are described and the modified position is in bold in the peptide sequence.

Virus	Position	Peptide	Modification	Experimental mass (Da)	Calculated mass (Da)	Mascot score
SBMV	1–19	**T**LPPELSVIEIPFEDVETR + pAp	pAp	2592.16	2592.16	53
SBMV	1–19	**T**LPPEL**S**VI**E**IPFEDVETR + pAp + P	pAp + phosphorylation + ethylation	2700.17	2700.15	60
SBMV	20–29	SYEFIEVEIK		1255.63	1255.63	72
SBMV	20–29	**S**YEFIEVEIK + P	Phosphorylation	1335.6	1335.6	65
SBMV	39–49	REFAWIPESGK		1318.67	1318.67	47
SBMV	40–49	EFAWIPESGK		1162.57	1162.57	72
SBMV	50–64	YWADDDDDSLPPPPK		1729.75	1729.75	99
SBMV	50–64	YWADDDDD**S**LPPPPK + P	Phosphorylation	1809.71	1809.71	99
SBMV	70–77	MVWSSAQE		936.4	936.4	42
RGMoV	1–10	**S**SENGEQGAR + pAp	pAp	1442.46	1442.46	28
RGMoV	1–10	**SS**ENGEQGAR + pAp + P	pAp (S1) + phos.(S2)	1522.43	1522.43	19
RGMoV	11–20	EIDAEEWISR		1246.58	1246.58	79
RGMoV	11–20	EIDAEEWI**S**R + P	Phosphorylation	1326.55	1326.55	42
RGMoV	21–32	EIDAEEWISREV**T**PTDVYIAGR + P	Phosphorylation	2628.22	2628.21	65
RGMoV	21–32	EIDAEEWI**S**REV**T**PTDVYIAGR + 2xP	Phosphorylation	2708.18	2708.18	54
RGMoV	21–32	EVTPTDVYIAGR		1319.68	1319.67	88
RGMoV	21–32	EV**T**PTDVYIAGR + P	Phosphorylation	1399.64	1399.64	56
RGMoV	37–54	VAGDEFSHSSYDPLAFSK		1955.9	1955.89	68
RGMoV	55–59	YKKER		722.41	722.41	22
RGMoV	58–79	ERGEMTWADMVEGDLDWDAREE		2639.09	2639.09	71
RGMoV	60–79	GEMTWADMVEGDLDWDAREE		2353.95	2353.95	91
RGMoV	60–79	GEM**T**WA**D**MVEGDLDWDAREE + P	Phosphorylation + ethylation	2461.94	2461.94	45

(Fig. 1B–D). In the RGMoV VPg on the other hand, we found phosphorylations on multiple serines, threonines and also on one tyrosine residue (Table 1). Of these listed modification sites, only phosphorylations at positions Ser2, Ser19, Thr23 and Thr63 were determined unambiguously (Fig. 2B–E). Serines at positions 43, 45 and 46 were identified by Mascot to be phosphorylated, but since the same MS2 spectrum had similar high ion scores for each of the potential phosphorylation sites (Table 2, Supplementary data), exact modification site assignments remain ambiguous. One of the possible phosphorylated serine assignments, Ser45 (best-scoring for this scan) is shown in Fig. 3A in bold, with other potentially phosphorylated serines underlined. While it is possible that all of these serine residues are phosphorylated, this result can also be explained by the known phenomenon of gas-phase rearrangements of phosphate groups. Namely, it has been shown that phosphate groups can be transferred to neighboring unmodified hydroxyl-containing amino acid residues upon collisional induced dissociation in linear ion traps, rendering many of the phosphorylation site assignments uncertain [24].

Interestingly, when the same peptide was detected as a doubly phosphorylated peptide, one of the site assignments, Tyr47, became unambiguous throughout Mascot searches (Fig. 2F, Table 2, Supplementary data), whereas it was still not possible to pinpoint confidently which of the serines was phosphorylated (Fig. 3B, Table 2, Supplementary data). In Fig. 3B, unambiguous phospho-Tyr47 is displayed in bold, best-scoring phospho-serine (Ser43) for this scan in underlined italic and other potentially phosphorylated serines are underlined. Therefore, we can conclude that Tyr47 is phosphorylated in combination with any of the three neighboring serines (Ser43, Ser45 or Ser46) being phosphorylated at the same time. For both SBMV and RGMoV, all phosphorylated peptides were also detected without the phosphorylations.

In addition to biologically relevant modifications, we detected random aspartate and/or glutamate ethylation(s), together with methionine and tryptophan oxidations (data not shown). These modifications are known to be generated in vitro during sample preparation [25,26] and were therefore not considered to be of biological relevance. All the VPg peptides usually detected in course of

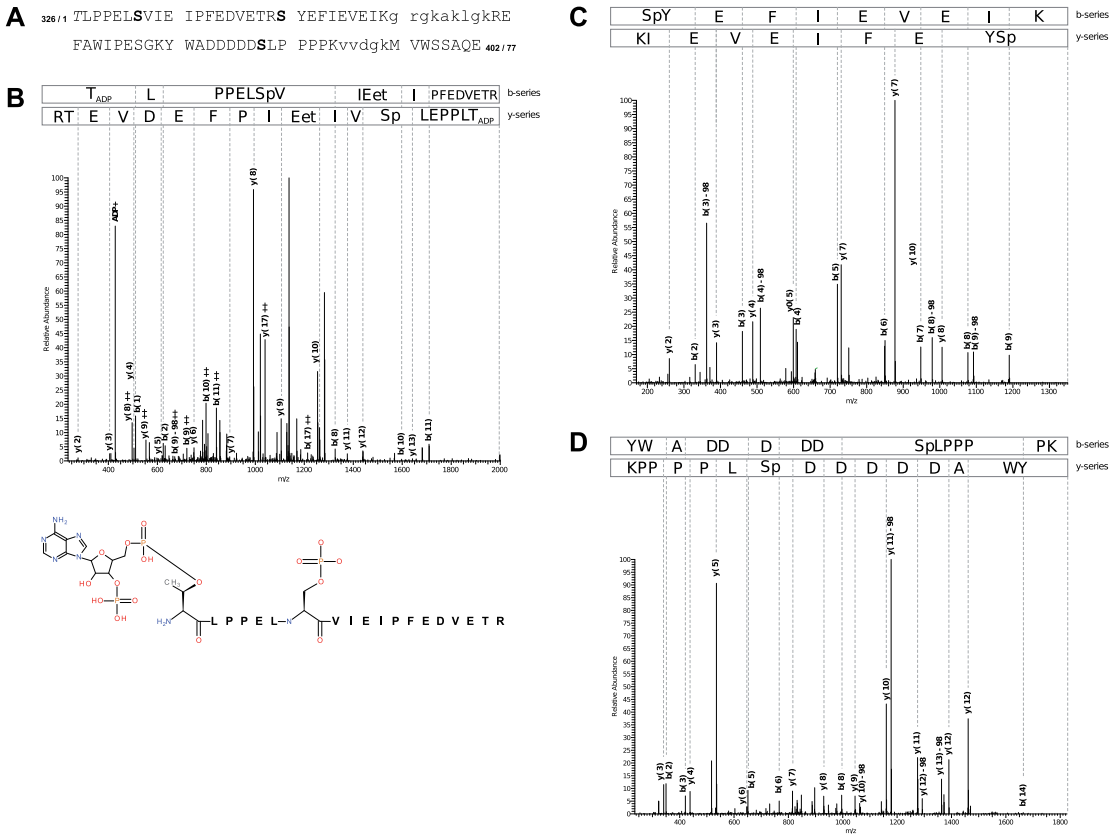


Fig. 1. Mass-spectrometrical characterization of SBMV VPg. (A) Representation of SBMV VPg. Numeration, indicated in subscript, corresponds to P2a polyprotein and VPg, respectively. The amino acid residue linked to RNA is in italics, phosphorylated residues are in bold and the region not detected is in lowercase. (B–D) Identification of post-translational modifications of SBMV VPg by MS/MS analysis. Co-purified VPg linked to RNA was trypsin-digested and RNA was degraded with acidic hydrolysis. The peptides were analyzed by nano-LC/MS/MS and resulting data was searched against corresponding sequence databases by MASCOT. The b and y ions represent peptide N- and C-terminal fragment ions produced by collision-induced dissociation in the mass spectrometer. (B) Determination of the residue covalently linked to RNA. The N-terminal SBMV VPg peptide, *T*LPPEL**S**VIE IPFEDVETR, was determined to contain a T1 linked pAp modification, a corresponding degradation product of viral RNA, and an additional phosphorylation at S6. The peptide sequence with the modification structure is represented below fragmentation spectrum. (C–D) Peptides **S**YEFIEVEIK and YWADDDDD**S**LP PPPK were detected to contain phosphorylations at respective positions.

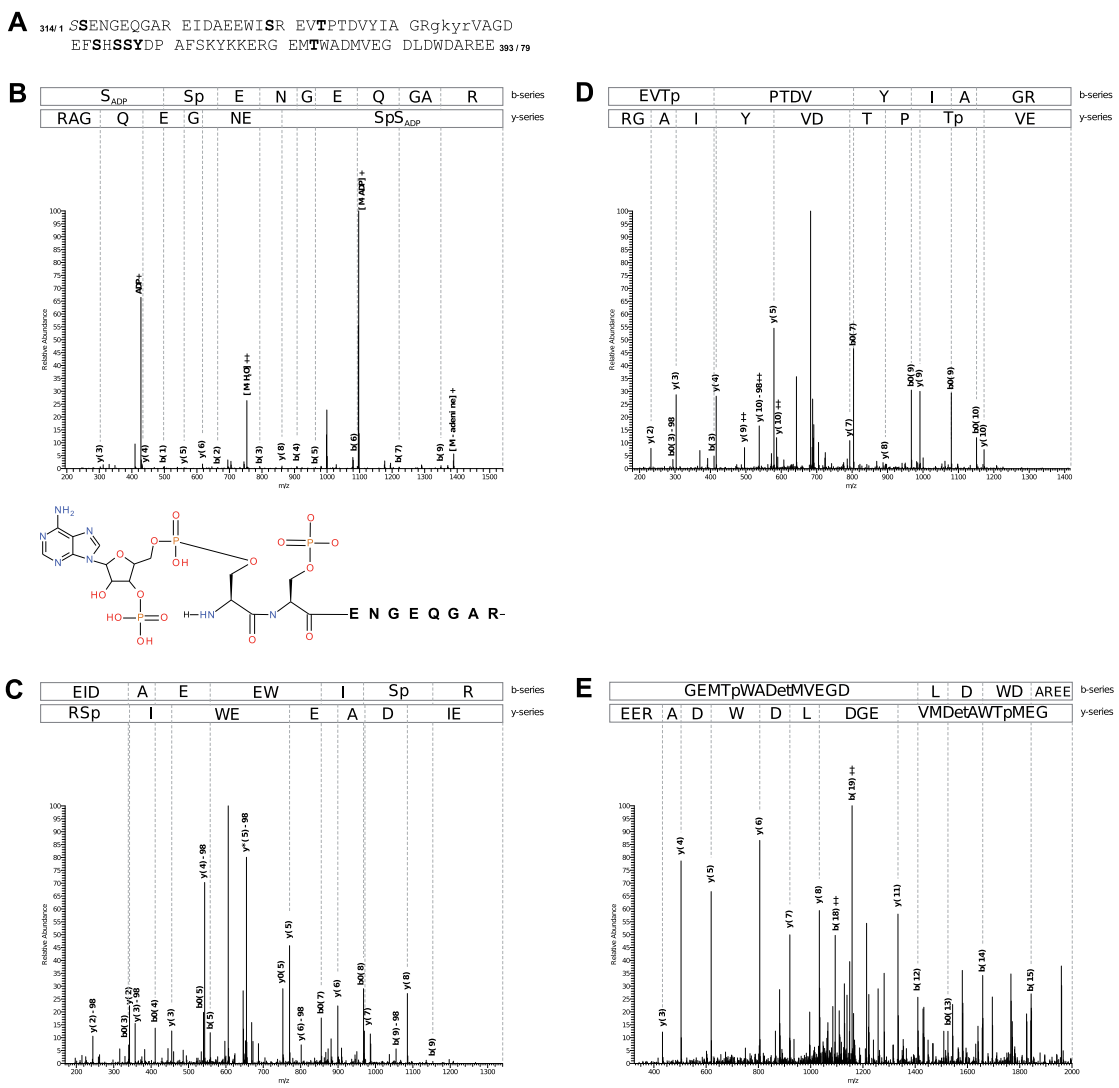


Fig. 2. Mass-spectrometrical characterization of RGMoV VPg. (A) Representation of RGMoV VPg. Numeration, indicated in subscript, corresponds to P2a polyprotein and VPg, respectively. The amino acid residue linked to RNA is in italics, phosphorylated residues are in bold and the region not detected is in lowercase. (B–E) Identification of post-translational modifications of RGMoV VPg by MS/MS analysis. Co-purified VPg linked to RNA was trypsin-digested and RNA was degraded with acidic hydrolysis. The peptides were analyzed by nano-LC/MS/MS and resulting data was searched against corresponding sequence databases by MASCOT. The b and y ions represent peptide N- and C-terminal fragment ions produced by collision-induced dissociation in the mass spectrometer. (B) The N-terminal RGMoV VPg peptide, S**S**ENGEQGAR, was determined to contain a S1 linked pAp modification and an additional phosphorylation at S2. The peptide sequence with the modification structure is represented below fragmentation spectrum. (C–E) Peptides EIDAEEWI**S**R, EV**T**PTDVYIAGR and GEM**T**WADMVEGDLDWDAR were detected to contain phosphorylations at respective positions. Peptide GEM**T**WADMVEGDLDWDAR also contained a in vitro induced ethylation (et) at position D6.

these analyses (except for oxidized and ethylated forms) are shown in Tables 1–2.

4. Discussion

The VPg aa residue reported to be linked to RNA has so far been either serine or tyrosine. In addition to these two aa residues, threonine also contains a hydroxyl group and has been shown to be used for phosphodiester bond formation with DNA [10]. Furthermore, there are reports about the use of threonine in RNA linking of cellular non-viral proteins [27]. Sobemoviral VPg sequences are not conserved at genus level and we recently demonstrated that individual sobemoviruses can somewhat surprisingly use different aa residues at different positions for RNA linking [9]. Sequence comparisons of sobemoviral VPgs as well as experimental data suggested that sobemoviruses might also use threonine for RNA linking as the threonine in position 1 of SBMV VPg was the only residue out of the first 20 which was not detected correctly by Edman sequencing [16]. Here we report for the first time that threonine can indeed be utilized for linking RNA to VPg. This

Table 2
Overview of RGMoV VPg peptide VAGDEFSHSSYDPAFSK phosphorylation annotations. For each listed MS2 scan, up to five highest-scoring Mascot assignments are shown. For singly phosphorylated peptide, the exact modification site is uncertain between different scans whereas for doubly phosphorylated peptide Tyr47 is always the best-scoring one. Full Mascot results for individual scans are shown in detail in the Supplementary data.

Virus	Position	Experiment/ injection	Scan number	Peptide	Modification	Experimental mass (Da)	Calculated mass (Da)	Mass error (ppm)	Mascot score
RGMoV	37–54	1/1	5766	VAGDEFSHSSYDPLAFSK	S9 phosphorylation	2035.8583	2035.8564	0.92	47
					S7 phosphorylation				47
					S10 phosphorylation				44
					Y11 phosphorylation				37
RGMoV	37–54	1/2	5580	VAGDEFSHSSYDPLAFSK	S7 phosphorylation	2035.8597	2035.8564	1.6	32
					S9 phosphorylation				26
					S10 phosphorylation				21
					Y11 phosphorylation				16
RGMoV	37–54	1/1	6141	VAGDEFSHSSYDPLAFSK	S9 phosphorylation	2035.86	2035.8564	1.75	62
					S10 phosphorylation				61
					S7 phosphorylation				51
					Y11 phosphorylation				46
					S17 phosphorylation				40
RGMoV	37–54	1/2	5911	VAGDEFSHSSYDPLAFSK	S10 phosphorylation	2035.8607	2035.8564	2.11	58
					S7 phosphorylation				56
					Y11 phosphorylation				50
					S9 phosphorylation				48
RGMoV	37–54	1/1	6259	VAGDEFSHSSYDPLAFSK	S9 + Y11 phosphorylation	2115.8269	2115.8228	1.94	62
					S7 + Y11 phosphorylation				62
					S10 + Y11 phosphorylation				62
					S9 + S10 phosphorylation				55
					S7 + S10 phosphorylation				53
RGMoV	37–54	1/2	6010	VAGDEFSHSSYDPLAFSK	S7 + Y11 phosphorylation	2115.8278	2115.8228	2.36	36
					S9 + Y11 phosphorylation				33
					S10 + Y11 phosphorylation				27
					S7 + S10 phosphorylation				27
					S9 + S10 phosphorylation				24
RGMoV	37–54	1/2	6011	VAGDEFSHSSYDPLAFSK	S7 + Y11 phosphorylation	2115.8295	2115.8228	3.2	43
					S7 + S10 phosphorylation				41
					S7 + S17 phosphorylation				35
					S10 + Y11 phosphorylation				34
					S9 + Y11 phosphorylation				34
RGMoV	37–54	2/2	5791	VAGDEFSHSSYDPLAFSK	S10 phosphorylation	2035.8554	2035.8564	−0.51	54
					S9 phosphorylation				47
					Y11 phosphorylation				41
					S7 phosphorylation				37
RGMoV	37–54	2/2	5797	VAGDEFSHSSYDPLAFSK	S7 phosphorylation	2035.8566	2035.8564	0.075	47
					S9 phosphorylation				39
					S10 phosphorylation				36
RGMoV	37–54	2/1	5851	VAGDEFSHSSYDPLAFSK	S10 phosphorylation	2035.8586	2035.8564	1.05	58
					S9 phosphorylation				56
					S7 phosphorylation				50
					S17 phosphorylation				45
					Y11 phosphorylation				40
RGMoV	37–54	2/2	5889	VAGDEFSHSSYDPLAFSK	S7 + Y11 phosphorylation	2115.8262	2115.8228	1.63	26
					S7 + S10 phosphorylation				22
					S9 + Y11 phosphorylation				20
					S7 + S9 phosphorylation				19
					S10 + Y11 phosphorylation				19
RGMoV	37–54	2/2	5892	VAGDEFSHSSYDPLAFSK	S7 + Y11 phosphorylation	2115.8287	2115.8228	2.81	45
					S7 + S10 phosphorylation				35
					S7 + S9 phosphorylation				35
					S10 + Y11 phosphorylation				35
					S9 + Y11 phosphorylation				35
RGMoV	37–54	2-1	6086	VAGDEFSHSSYDPLAFSK	S7 + Y11 phosphorylation, D4 ethylation	2143.8579	2143.8541	1.78	49
					S9 + Y11 phosphorylation, D4 ethylation				42
					S7 + Y11 phosphorylation, E5 ethylation				41
					S9 + Y11 phosphorylation, E5 ethylation				40

demonstrates that sobemoviruses are even more diverse than expected and can use all available hydroxyl group containing aa residues for RNA linking. Furthermore, this finding also shows that threonine as an RNA linking residue cannot be ruled out for viruses from other genera. It is surprising that the use of threonine had not been described before, indicating that sobemoviruses are either truly unique in their variability or that the VPg linking as such has not been studied intensively enough. In addition to SBMV, a few VPgs of other sobemoviruses also have a threonine at their N-terminus – *Sowbane mosaic virus*, *Subterranean clover mottle virus,* SeMV and *Southern cowpea mosaic virus* (SCPMV). The latter two are very closely related to SBMV and their VPgs are conserved (65% identity and 79% similarity within the VPgs of the three viruses). With the exception of *Cocksfoot mottle virus* (CfMV) for which tyrosine at position five is used for RNA linking, the rest of sequenced sobemoviruses all have a serine residue at the putative first position of VPg. Furthermore, for SBMV, RGMoV and RYMV it has now been experimentally proven that it is the first VPg residue

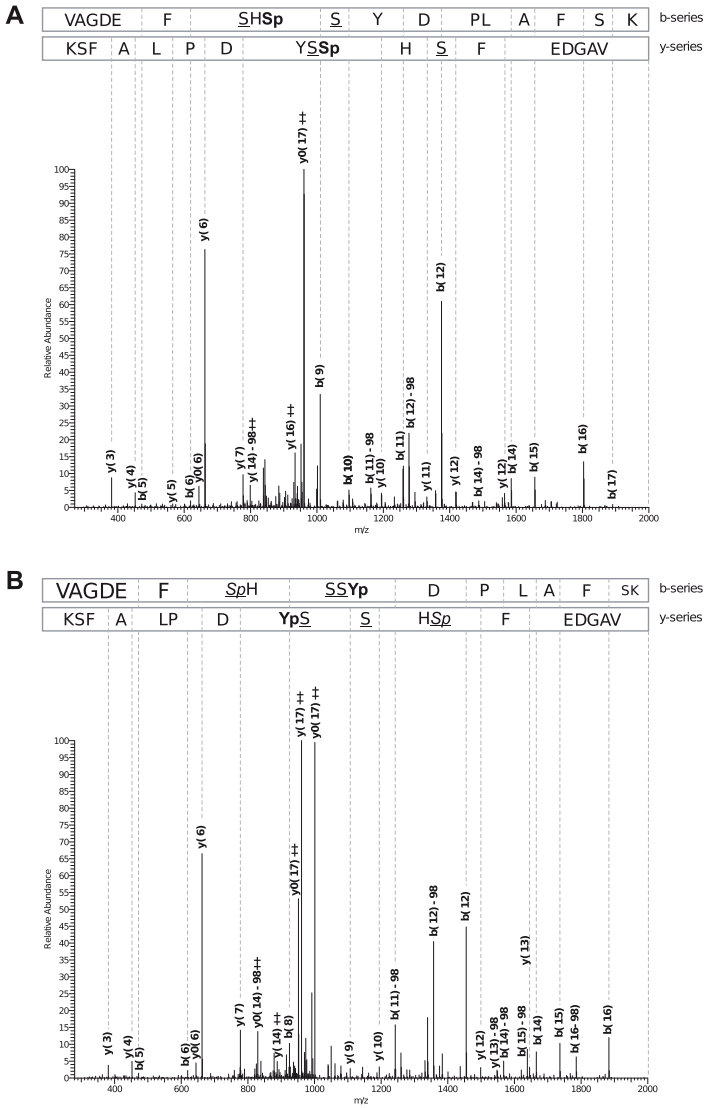


Fig. 3. Phosphorylation site assignments for RGMoV VPg peptide VAGDEFSHSSYDPAFSK. (A) Annotated fragmentation spectrum for singly phosphorylated peptide. Best-scoring phosphorylated serine assignment (for this scan) is shown in bold, other potentially phosphorylated positions are underlined. (B) Annotated fragmentation spectrum for doubly phosphorylated peptide. Unambiguous phospho-Tyr47 is displayed in bold, best-scoring phospho-serine (Ser43) for this scan in underlined italic and other potentially phosphorylated positions are underlined.

that is indeed used for RNA linking. Therefore, we conclude that threonine and serine residues at the first position of the VPgs of these viruses are most probably used for RNA linking.

The remainder of RNA degradation attached to the SBMV VPg was detected to be pAp, establishing that the first nucleotide of SBMV genomic RNA must be A. Intriguingly, all available SBMV (and SCPMV) genome sequences have the 5′ sequence starting with CACAAA. However, it has been demonstrated that at least for some isolates of SBMV and SCPMV the 5′ terminal C is erroneous and that the true 5′ sequence is ACAAA [28]. Our results independently confirm that the first nucleotide of the SBMV genome should indeed be A. Since CACAAA is reported to be the beginning of the genome of another very closely related species, SeMV, it would be interesting to see whether these viruses actually vary at the 5′ end of the genome or not.

We also confirmed that the 5′ sequence of RGMoV genome starts with an A nucleotide [29]. At the same time we discovered that putative N-terminus of RGMoV VPg was annotated incorrectly – mature RGMoV VPg has additional three amino acid residues, SSE, in its N-terminus. This discovery provides us with further information about the protease specificity of sobemoviruses. The experimentally proven protease cleavage sites for all sobemoviruses are E/T and E/S, with the only exception being CfMV that utilizes also E/N site for the cleavage of its VPg. For SeMV the proposed

specificity of the protease is (N or Q)–E/(T or S)–X, where X is an aliphatic residue [18]. While this proposed consensus applies to SeMV, to one CfMV site and also to SBMV (closely related to SeMV), it certainly does not apply for the whole genera.

Previously the phosphorylation of potyviral and sobemoviral VPg has been reported [9,30,31]. Comparing the phosphorylation sites of VPgs of CfMV, RYMV, SBMV and RGMoV only indicates that the protein sequences as well as their phosphorylation patterns are diverse. However, based on VPg sequence similarity, the data on SBMV allows us to make predictions about the phosphorylation of SeMV and SCPMV VPgs. Serines 7 and 20 are present at the same positions in SeMV and serine 20 in SCPMV, which indicates that these residues might also be the targets for phosphorylation. The position corresponding to serine 58 of SBMV VPg is occupied by glutamic acid in SeMV VPg and by aspartic acid in SCPMV VPg, both of them chemically mimic phosphoserine to some extent. This indicates that the negative charge at VPg position 58 might be important for all three viruses. Altogether, sobemoviral VPgs, which are anyhow rich in negatively charged aa residues seem to require an additional overall negative charge, achieved through multiple phosphorylations. In contrast, the N-termini of sobemoviral coat proteins are rich in positively charged aa residues [14] and reside inside the virion, presumably in contact with negatively charged RNA and VPg [32–34]. Therefore the phosphorylation of VPg might be needed for providing additional stability to the virion through electrostatic interactions. Nevertheless, one also cannot rule out other possible functions of VPg phosphorylation throughout the virus infection cycle. For instance, in the in vitro assay RdRp of SeMV failed to nucleotidylate VPg purified from bacteria where presumably phosphorylation does not occur similarly to plant cells [35]. The authors propose that the nucleotidylation reaction could be dependent on (host) factors missing in the in vitro assay. VPg phosphorylation could be one factor regulating RdRp VPg interactions required for this process.

Acknowledgments

We wish to thank Signe Nõu for excellent plant care. This work was supported by Estonian Science Foundation grant no. 7363. Mass spectrometric analyses were in part supported by the European Regional Development Fund through the Center of Excellence in Chemical Biology (Institute of Technology, University of Tartu) and by the ECOGENE project (EC grant number 205419 to Estonian Biocentre).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2011.08.009.

References

[1] 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, 5263–5266.
[2] Jaegle, M., Wellink, J. and Goldbach, R. (1987) The genome-linked protein of *Cowpea mosaic virus* is bound to the 5′ terminus of virus RNA by phosphodiester linkage to serine. J. Gen. Virol. 68, 627–632.
[3] 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, 4868–4872.
[4] Schein, C.H., Oezguen, N., Volk, D.E., Garimella, R., Paul, A. and Braun, W. (2006) NMR structure of the viral peptide linked to the genome (VPg) of poliovirus. Peptides 27, 1676–1684.
[5] Anindya, R., Chittori, S. and Savithri, H.S. (2005) Tyrosine 66 of *Pepper vein banding virus* genome-linked protein is uridylylated by RNA-dependent RNA polymerase. Virology 336, 154–162.
[6] Belliot, G., Sosnovtsev, S.V., Chang, K.O., McPhie, P. and Green, K.Y. (2008) Nucleotidylylation of the VPg protein of a human norovirus by its proteinase-polymerase precursor protein. Virology 374, 33–49.
[7] 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, 511–513.
[8] Zalloua, P.A., Buzayan, J.M. and Bruening, G. (1996) Chemical cleavage of 5′-linked protein from tobacco ringspot virus genomic RNAs and characterization of the protein–RNA linkage. Virology 219, 1–8.
[9] Olspert, A., Peil, L., Hébrard, E., Fargette, D. and Truve, E. (2011) Protein–RNA linkage and post-translational modifications of two sobemovirus VPgs. J. Gen. Virol. 92, 445–452.
[10] Salas, M. (1991) Protein-priming of DNA replication. Annu. Rev. Biochem. 60, 39–71.
[11] Fricova, D., Valach, M., Farkas, Z., Pfeiffer, I., Kucsera, J., Tomaska, L. and Nosek, J. (2010) The mitochondrial genome of the pathogenic yeast *Candida subhashii*: GC-rich linear DNA with a protein covalently attached to the 5′ termini. Microbiology 156, 2153–2163.
[12] Steil, B.P. and Barton, D.J. (2009) Conversion of VPg into VPg-pUpU-OH before and during poliovirus negative-strand RNA synthesis. J. Virol. 83, 12660–12670.
[13] Khan, M.A., Miyoshi, H., Gallie, D.R. and Goss, D.J. (2008) Potyvirus genome-linked protein, VPg, directly affects wheat germ in vitro translation: interactions with translation initiation factors eIF4F and eIFiso4F. J. Biol. Chem. 283, 1340–1349.
[14] Tamm, T. and Truve, E. (2000) Sobemoviruses. J. Virol. 74, 6231–6241.
[15] Meier, M. and Truve, E. (2007) Sobemoviruses possess a common CfMV-like genomic organization. Arch. Virol. 152, 635–640.
[16] 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, 21–24.
[17] Mäkinen, K., Mäkeläinen, 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, 2783–2789.
[18] Nair, S. and Savithri, H.S. (2010) Processing of SeMV polyproteins revisited. Virology 396, 106–117.
[19] Hébrard, E., Pinel-Galzi, A. and Fargette, D. (2008) Virulence domain of the RYMV genome-linked viral protein VPg towards rice rymv1–2-mediated resistance. Arch. Virol. 153, 1161–1164.
[20] Ghosh, A., Rutgers, T., Ke-Qiang, M. and Kaesberg, P. (1981) Characterization of the coat protein mRNA of southern bean mosaic virus and its relationship to the genomic RNA. J. Virol. 39, 87–92.
[21] 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, 264–274.
[22] Rappsilber, J., Mann, M. and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using stagetips. Nat. Protoc. 2, 1896–1906.
[23] Olsen, J.V., de Godoy, L.M.F., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S. and Mann, M. (2005) Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a c-trap. Mol. Cell Proteomics 4, 2010–2021.
[24] Palumbo, A.M. and Reid, G.E. (2008) Evaluation of gas-phase rearrangement and competing fragmentation reactions on protein phosphorylation site assignment using collision induced dissociation-ms/ms and ms3. Anal. Chem. 80, 9735–9747.
[25] Xing, G., Zhang, J., Chen, Y. and Zhao, Y. (2008) Identification of four novel types of *in vitro* protein modifications. J. Proteome Res. 7, 4603–4608.
[26] Stadtman, E.R. and Levine, R.L. (2003) Free radical-mediated oxidation of free amino acids and amino acid residues in proteins. Amino Acids 25, 207–218.
[27] Thomas, L., Pfeifle, J. and Anderer, F.A. (1987) Human SS-b/La autoantigen contains a covalent protein–RNA linkage. Biochim. Biophys. Acta 909, 173–182.
[28] Hacker, D.L. and Sivakumaran, K. (1997) Mapping and expression of southern bean mosaic virus genomic and subgenomic RNAs. Virology 234, 317–327.
[29] 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, 395–398.
[30] Puustinen, P., Rajamäki, M., Ivanov, K.I., Valkonen, J.P.T. and Mäkinen, K. (2002) Detection of the potyviral genome-linked protein VPg in virions and its phosphorylation by host kinases. J. Virol. 76, 12703–12711.
[31] Hafrén, A. and Mäkinen, K. (2008) Purification of viral genome-linked protein VPg from potato virus a-infected plants reveals several post-translationally modified forms of the protein. J. Gen. Virol. 89, 1509–1518.
[32] 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 Å resolution. Virology 369, 364–374.
[33] 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 Å resolution. Nature 286, 33–39.
[34] Rossmann, M.G., Abad-Zapatero, C., Erickson, J.W. and Savithri, H.S. (1983) RNA–protein interactions in some small plant viruses. J. Biomol. Struct. Dyn. 1, 565–579.
[35] Govind, K. and Savithri, H.S. (2010) Primer-independent initiation of RNA synthesis by SeMV recombinant RNA-dependent RNA polymerase. Virology 401, 280–292.

MANUSCRIPT

The CP of *Cocksfoot mottle virus* is dispensable for movement

Olspert, A., Kamsol, K., Sarmiento, C., Truve, E.

Cocksfoot mottle virus coat protein is dispensable for the systemic infection

Allan Olspert, Kristjan Kamsol, Cecilia Sarmiento, Erkki Truve*

Department of Gene Technology, Tallinn University of Technology, Akadeemia tee 15, 12618 Tallinn, Estonia

*Corresponding author:

Erkki Truve

Department of Gene Technology

Tallinn University of Technology

Akadeemia tee 15

12618 Tallinn

Estonia

Tel.: +372 620 4422 (office)

Fax: +372 620 4401

Summary

The way how a virus moves inside a plant depends on the specific virus-host pathosystem. In the case of sobemoviruses, several studies have shown the involvement of the coat protein (CP) in viral movement. Cocksfoot mottle sobemovirus (CfMV) has not yet been analyzed in that aspect. In this study we analyzed the movement of different CfMV CP substitution and deletion mutants. All mutant viruses were able to move and infect plants systemically, although usually the virus accumulation was reduced when compared to wild type CfMV. Thus, the CP was shown to be dispensable for systemic movement in all three tested hosts, oat, wheat and barley. Movement of CfMV was also characterized with mutant viruses expressing EGFP fused to the C-terminus of CP or replacing CP C-terminal region. EGFP expression was detected up to one week days post inoculation (dpi) in epidermal and mesophyll cells. Although EGFP fluorescence was not detected in upper leaves, some of the plants infected with these EGFP containing viruses displayed CfMV symptoms. Analysis of the upper leaves revealed that due to recombination the viruses had lost the EGFP sequence and sometimes also most of the CP sequence.

In addition we demonstrated that CP does not have RNA silencing suppressor activity nor does it influence RNA silencing suppression of P1. In conclusion we demonstrated that the CP of CfMV is dispensable for systemic movement but required for efficient transmission and that CP might have functions enhancing virus accumulation unrelated to RNA silencing.

Introduction

In order to infect the host efficiently, plant viruses need to invade the whole organism rather than just remain in the initially infected cell(s). To facilitate this, viruses have to be able to move between adjacent cells (cell-to-cell movement) and to other plant organs through vascular tissues (systemic movement). Whether these two distinct transport processes utilize the same mechanisms depends on the specific virus-host pathosystem. Virally encoded movement proteins (MP) are the key elements which enable virus movement through various host interactions (reviewed by (Benitez-Alfonso et al., 2010; Lucas, 2006; Scholthof, 2005)). In addition to MP most viruses also require the capsid protein (CP) for virus movement (reviewed by (Callaway et al., 2001; Lucas, 2006; Scholthof, 2005). Based on the requirement of CP for transport, plant viruses can be divided into three major categories. Viruses in the first group only relay on the MP and CP is dispensable for movement. Members of *Tobamovirus, Dianthovirus, Umbravirus, Carmovirus, Pomovirus, Tombusvirus* and *Hordeivirus* genera have been assigned to this group (Lucas, 2006; Scholthof, 2005). Viruses forming the second group require both, MP and CP for trafficking. Representatives of the second group belong to potyviruses, potexviruses and cucumoviruses. A third group, represented by members of closteroviruses, nepoviruses and comoviruses, comprises viruses that also require CP because they move as virions. However, in the end the requirement of viral proteins for movement depends on the particular virus-host pathosystem. Even different strains of a virus can utilize different movement strategies in the same host and the same strain can use alternative modes of trafficking in different host. For instance, some isolates of *Brome mosaic virus* (BMV) require CP for movement whereas others do not (Takeda et al., 2005). The transport mode, i.e. requirement of CP, of different isolates of BMV was determined solely by the MP. *Cucumber mosaic virus* (CMV) is another example where the necessity of CP is determined by MP, as a deletion in the MP renders CMV movement CP-independent (Nagano et al., 2001). In the absence of CP *Barley stripe mosaic virus* is able to systemically infect barley but not *Nicotiana benthamiana* (Petty and Jackson, 1990). The systemic movement of cymbidium ringspot tombusvirus has been suggested to take place either as virions or in a non-virion form, depending on the host (Dalmay et al., 1992).

Over the past decade examples of cell-to-cell and long-distance movement of endogenous RNA species have been accumulating (reviewed by Kehr and Buhtz, 2008). These non-cell-autonomous RNA species include messenger RNAs, microRNAs and small interfering RNA (siRNA). This shows that the intercellular transport of RNAs presumably as ribonucleoprotein complexes (RNP) is an important host mechanism. The RNAs of viroids are also transported entirely by the host machinery as their genome does not encode any

proteins. Evidence about the movement of BMV RNA independently of viral proteins has been reported (Gopinath and Kao, 2007), but it has not been demonstrated that the transport of any viral RNA genome depends entirely on the host movement system. Viruses need an effective replication, quick movement and accumulation in the initial infection site in order to overcome host defenses and therefore presumably cannot allow the host be solely in charge of their movement.

Cocksfoot mottle virus (CfMV) is a plant sobemovirus with a monopartite positive-sense ssRNA genome (Figure1.A)(for review Tamm and Truve, 2000b, genome organization revised in Meier and Truve, 2007). Sobemoviruses have a viral protein genome linked (VPg) covalently attached to the 5' end of genomic and subgenomic RNAs. The central part of the genome encodes the viral polyprotein through a -1 programmed frameshift mechanism (Mäkinen et al., 1995). The 5' proximal ORF1 encodes the P1 protein, while the viral CP is expressed from the 3' proximal ORF3 via subgenomic RNA. The P1 proteins of sobemoviruses, which lack sequence similarity with other viral or non-viral proteins, have long been proposed as the putative MPs. Indeed, the P1s of *Rice yellow mottle virus* (RYMV), *Southern cowpea mosaic virus* (SCPMV) and CfMV are known to be dispensable for replication but needed for virus movement (Bonneau et al., 1998; Meier et al., 2006; Sivakumaran et al., 1998). In addition, the P1 proteins of *Rice yellow mottle virus* (RYMV) and CfMV also act as suppressors of RNA silencing and for the latter ssRNA binding in a sequence-nonspecific manner has been reported (Lacombe et al., 2010; Sarmiento et al., 2007; Tamm and Truve, 2000a; Voinnet et al., 1999). Recently it was shown that *Sesbania mosaic virus* (SeMV) P1 interacts with VPg, P10 and CP (Chowdhury and Savithri, 2011; Roy Chowdhury and Savithri, 2011). This indicates a complex interplay of these proteins during infection. Indeed, the capsid protein is another sobemoviral protein known to be involved in the spread of some sobemoviruses. The CPs of RYMV and SCPMV have been demonstrated to be needed for virus cell-to-cell movement but not for virus replication (Brugidou et al., 1995; Sivakumaran et al., 1998). In addition, the long distance movement of SCPMV and RYMV has been proposed to be dependent on particle formation (Fuentes and Hamilton, 1993; Opalka et al., 1998). The N-terminal part of all sobemoviruses CPs is rich in basic amino acids and contains an arginine-rich region. This terminal part seems to be implicated in CP-RNA interactions and RNA encapsidation (Lee and Hacker, 2001; Satheshkumar et al., 2005) as well as in the regulation of T=3 particle formation (Lokesh et al., 2002; Savithri and Erickson, 1983). It has also been shown that the N-terminal part of SCPMV CP interacts with membranes (Lee et al., 2001) and that the N-terminus of CfMV CP contains functional nuclear localization signals (Olspert et al., 2010).

Although the movement in the form of virus particles has been proposed for

some sobemoviruses, it has not been demonstrated. The infectivity of two CP arginine-rich region CfMV mutants and of a virus incapable of producing CP was tested *in planta*. In addition, CfMV expressing different CP EGFP fusions were analyzed for cell-to-cell and systemic movement. To our surprise, the experiments demonstrated that CP is dispensable for cell-to-cell as well as systemic movement of CfMV in oats, wheat and barley.

Results

Infectivity of CP mutants

In the current study we wanted to analyze the role of CfMV CP and in particular CP arginine-rich region in viral infection cycle. As the region encoding the arginine-rich region of CfMV CP overlaps with the coding part of RdRp, two different arginine-rich region mutants were generated. In mutant R5X, five arginines were replaced by four non-basic amino acid residues, which also caused the introduction of mutations into the RdRp gene (Figure 1A). In mutant R3L, three out of five arginines were replaced by leucines and no mutations were introduced to RdRp. A full CP knockout virus, titled noCP, was created by mutating the CP initiation sequence AUGAUG to ACGACG and by introducing a stop codon into the CP reading frame after its overlap with RdRp gene (Figure 1B).

Oat plants were biolistically inoculated with the mutant viruses and tested for infection by Northern hybridization, RT-PCR and Western blot (Figure 2). The analysis of inoculated leaves revealed that all three mutants were able to replicate and accumulate in the inoculated leaves. To our surprise all three mutants were also able to infect the host plant systemically. Sequencing of the fragments obtained by RT-PCR proved that the viral RNA detected from upper leaves still carried all the mutations (data not shown). As expected we were not able to detect CP in the case of mutant noCP (Figure 2A, W.blot). Indeed, CP was undetectable for this mutant even after enrichment for CP by immunoprecipitation (data not shown).

In comparison, R5X had the lowest RNA levels of all tested viruses while R3L usually had a similar or slightly lower titer than wild-type (wt) CfMV (Figure 2 and our unpublished observations). Mutant noCP usually had reduced accumulation when compared to wt CfMV (Figure 2). Sometimes the sub-genomic RNA levels of mutant noCP were significantly higher than genomic RNA levels (data not shown).

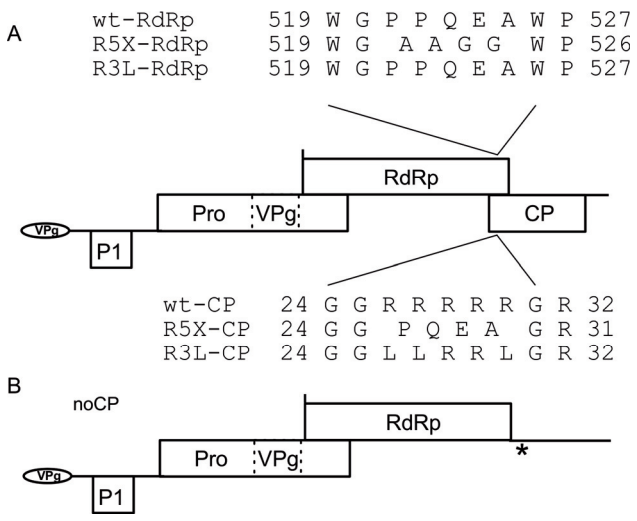


Figure 1. **Overview of the mutations introduced into the CfMV genome.** (A) Mutations in the arginine rich region of CP and the corresponding mutations in RdRp. The RdRp is translated in –1 reading frame in relation to CP. Wild type (wt-RdRp and wt-CP) and mutated (R5X-RdRp, R5X-CP, R3L-RdRp and R3L-CP) amino acid sequences of the changed regions are indicated in single letter code. CfMV genome: Pro, protease domain, VPg, VPg domain. (B) Representation of CfMV mutant noCP, in which the initiation codon of CP is mutated and an additional stop codon, indicated with an asterisk, is introduced into the CP reading frame after RdRp sequence.

Mechanical transmission of CP mutants

As CP is essential for the formation of virus particles, it was interesting to see whether the mutant viruses would be transmissible by mechanical inoculation. Oat plants were inoculated with sap obtained from the upper leaves of plants infected with the mutants and analyzed for CfMV infection as described above (Figure 2B). Mechanical transmission of mutant R3L occurred in a similar way to wt virus, as CfMV infection was detected in both inoculated and upper leaves. However, in the case of mutant R5X the transmission was successful in the inoculated leaves, but the mutant was unable to infect the plants systemically. For both mutants the presence of the mutations in the viral RNA was verified by sequencing (data not shown). Interestingly, we were unable to detect infection in plants inoculated with noCP. Again R5X accumulated poorly while R3L and wt CfMV had roughly similar titers (Figure 2).

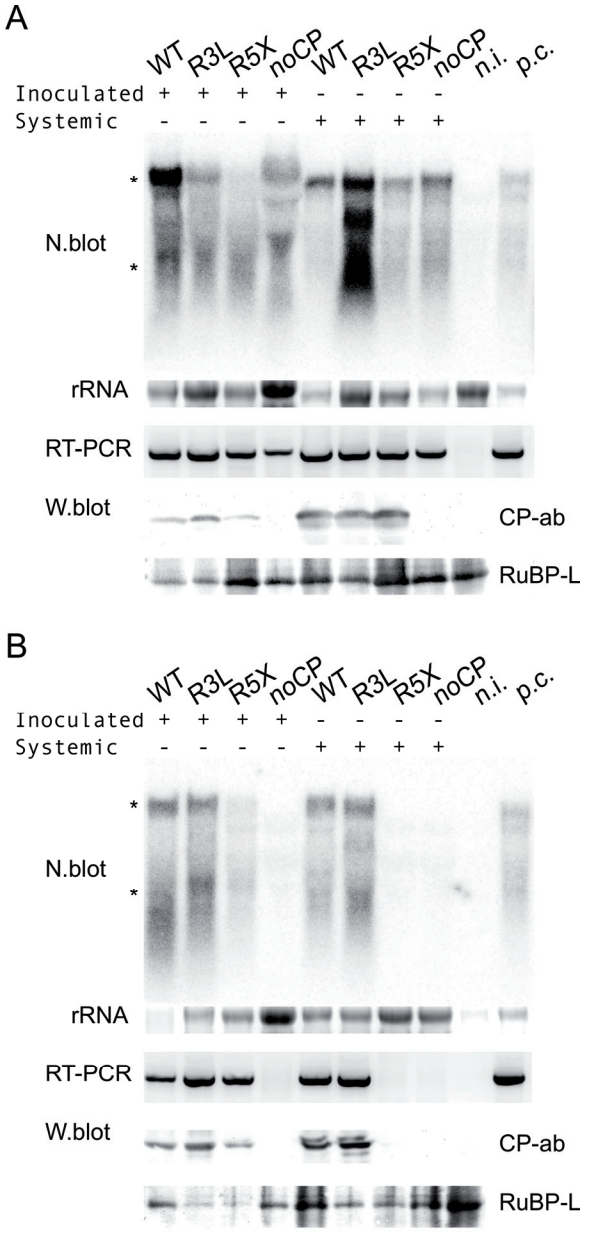


Figure 2. **Detection of CfMV infection in oat plants inoculated with CfMV**. (A) N.blot - Northern hybridization analysis of total RNA extracted from plants inoculated with CfMV (WT) and CfMV mutants (R3L, R5X, noCP). RNA from a non-inoculated plant (n.i.) was used as negative control and RNA from a plant previously known to be infected was used as positive control (p.c.). Samples were collected from the inoculated leaves (Inoculated) at 14 dpi and from upper leaves (Systemic) at 21 dpi. The

hybridization was carried out with a radioactive probe detecting CfMV CP coding region. Upper asterisk shows genomic RNA and the lower one corresponds to subgenomic RNA, ethidium-bromide staining of 28S ribosomal RNA (rRNA) serves as loading control. RT-PCR - analysis of the same RNA by reverse transcription PCR using primers amplifying the region of RdRp and CP genes. W.blot - Western blot analysis of plant total protein extracts with polyclonal anti CfMV CP antibody (CP-ab). For W.blot an additional positive control was not used. Ribulose-1,5-bisphosphate carboxylase oxygenase large subunit (RuBP-L) was visualized with Poucean S stain as loading control. (B) Detection of CfMV infection in sap-inoculated plants. Oat plants were sap-inoculated with CfMV and its mutants. Analysis and annotation as described in A.

Infectivity of mutant noCP in other hosts

In order to analyze whether the CP of CfMV is dispensable for infection and systemic movement also in other host plants, wheat and barley plants were biolistically inoculated with CfMV and the mutant noCP. Infection and systemic movement of noCP occurred in both host species similarly to what was observed in oat (Figure 3). The mutant noCP usually accumulated to lower levels than wt CfMV and displayed an increased sub-genomic RNA to genomic RNA ratio. The latter was most clearly seen in wheat (Figure 3), although the tendency was present in all tested hosts (Figure 2-3). In wheat, the mutant noCP produced strong systemic symptoms similarly to wt CfMV (data not shown).

Movement of CP-EGFP and CPdelta-EGFP viruses

Since it was demonstrated that the CP was dispensable for cell-to-cell and long distance movement of CfMV, it was decided to further characterize CfMV movement. Therefore, two EGFP expressing CfMV viruses were constructed (Figure 4). EGFP was fused to the C-terminus of the full length CP (CP-EGFP) or placed instead of the CP C-terminus (CPdelta-EGFP).

Oat plants were biolistically inoculated and virus movement was analyzed by monitoring EGFP fluorescence in inoculated tissues. At two days post-inoculation (dpi) the fluorescence was visible mainly in single epidermal cells and in small foci of mesophyll cells (Figure 5). The CP-EGFP fusion protein localized to the nucleus and cytoplasm of epidermal cells (data not shown). At three dpi the EGFP was detected in the neighboring epidermal cells and the infection foci in the mesophyll were expanding. From the fourth to the sixth dpi the area of infected mesophyll cells continued to grow and presumably reached the vasculature. Beyond that time point the infected area stopped to grow and fluorescence started to fade due to the spreading necrosis caused by the initial biolistic inoculation.

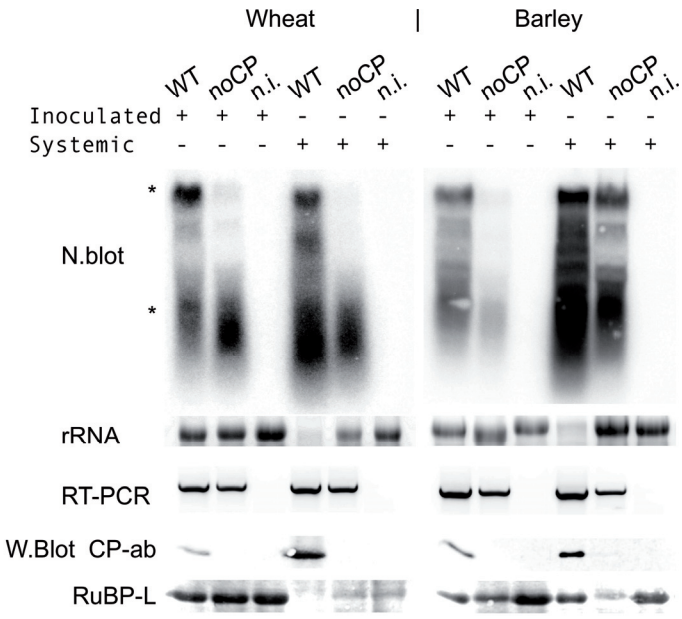


Figure 3. **Detection of CfMV and mutant noCP in biolistically inoculated wheat and barley plants**. The host plant is indicated above the panels. N.blot - Northern hybridization analysis of total RNA extracted from plants inoculated with CfMV (WT) and CfMV CP deletion mutant (noCP). Samples from a non-inoculated plant (n.i.) were used as negative control. Samples were collected from the inoculated leaves (Inoculated) at 14 dpi and from upper leaves (Systemic) at 21 dpi. The hybridization was carried out with a radioactive probe detecting CfMV CP coding region. Upper asterisk shows genomic RNA and the lower one corresponds to subgenomic RNA, ethidium-bromide staining of 28S ribosomal RNA (rRNA) serves as loading control. RT-PCR - analysis of the same RNA by reverse transcription PCR using primers amplifying the region of RdRp and CP genes. W.blot - Western blot analysis of plant total protein extracts with polyclonal anti CfMV CP antibody (CP-ab). Ribulose-1,5-bisphosphate carboxylase oxygenase large subunit (RuBP-L) was visualized with Poucean S stain as loading control.

In the case of CPdelta-EGFP the fluorescence was considerably weaker of what was observed with CP-EGFP. The fluorescence appeared strongest around the third to the fourth dpi with CPdelta-EGFP (Figure 5). Weak fluorescence was detected in single epidermal cells after two dpi, which sometimes spread to neighboring epidermal cells the following days. EGFP was not detected in mesophyll cells. The signal from epidermal cells usually disappeared around the fifth day (Figure 5).

No distinguishable EGFP signal was detected in the upper leaves with either of the EGFP containing viruses.

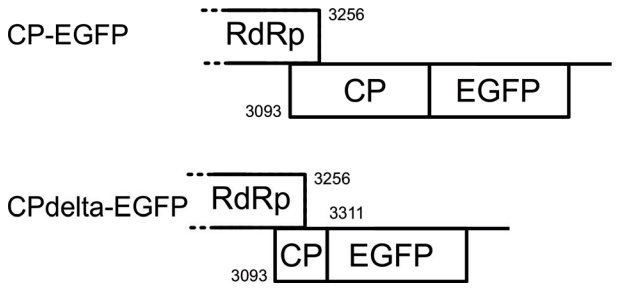


Figure 4. **Schematic representation of EGFP expressing CfMV viruses.** CP-EGFP – EGFP is fused to the C-terminus of full-length CP. CPdelta-EGFP – EGFP is replacing CP C-terminus starting from the nucleotide position 3311. Both viruses have full-length 3' UTR.

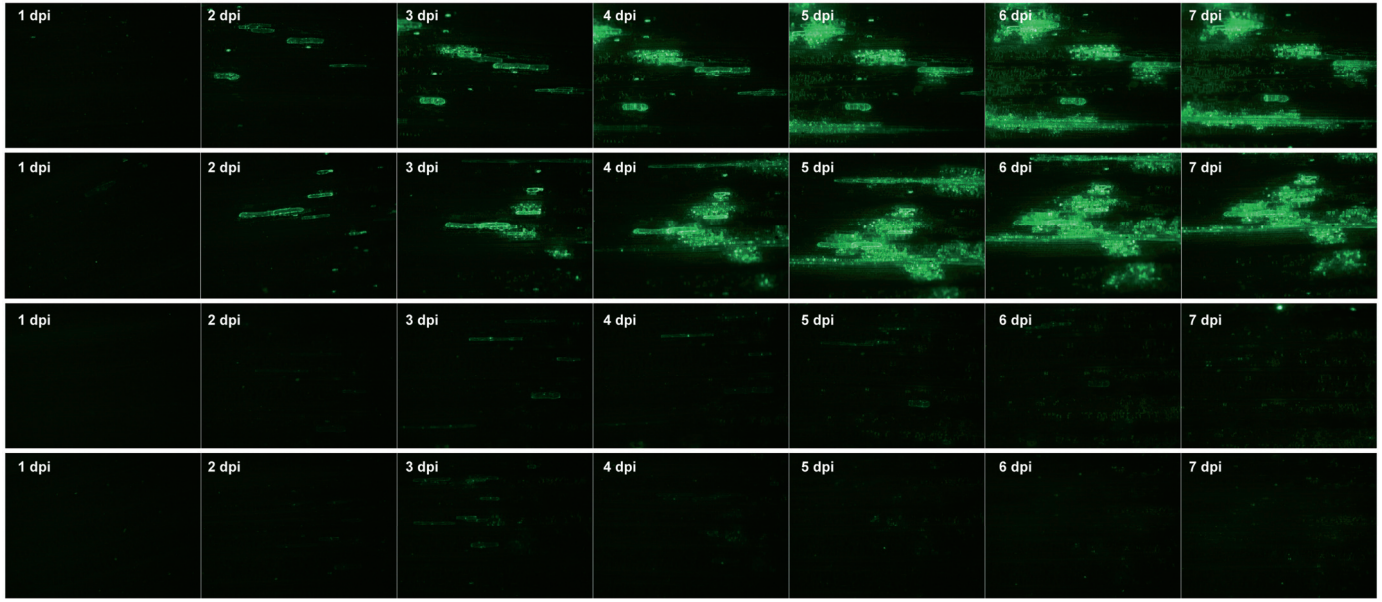


Figure 5. **1-7 days post inoculation of CP-EGFP and CPdelta-EGFP CfMV movement in inoculated oat leaves**. Upper two panels correspond to CP-EGFP and lower two to CPdelta-EGFP. Oat leaves were biolistically inoculated with mutant CfMV clones expressing EGFP fused to either the full length CP (CP-EGFP) or in place of the CP C-terminal region (CPdelta-EGFP). Virus localization was monitored by EGFP fluorescence.

Detection of CfMV recombination mutants

Although no EGFP was detectable in the upper leaves of plants infected with CP-EGFP or CPdelta-EGFP, some of these plants developed CfMV infection symptoms in the upper leaves (data not shown). This prompted for the further analysis of the upper leaves. Indeed, viral RNA was detected from the upper leaves of inoculated plants by RT-PCR but the fragments had considerably lower molecular weight than expected (data not shown). Sequencing revealed that the initially EGFP containing viruses had lost the EGFP sequence as well as different portions of CP/RdRp coding sequence (Figure 6). CfMV RdRp coding sequence ends at position 3256 and nucleotides (nt) 3093-3857 code the CP. Two different naturally occurring deletion mutant subsets were identified. In the first group nt from around 3244-3253 to 3864 of the RdRP and CP genes (numbering corresponds to wt CfMV) were deleted. Usually around 5 nt were present at the junction site, that could not be matched to CfMV sequence, but in some cases up to 65 nt of EGFP sequence were nested inside the CfMV sequence (data not shown). Deletion mutants belonging to the first group were detected in plants inoculated with either CP-EGFP or CPdelta-EGFP. The second group was comprised of viruses where nt between 3831-3854 to 3864-3868 at the end of the CP gene were missing. These mutants were obviously only detected in plants inoculated with CP-EGFP.

CP and RNA silencing

According to the results obtained in this study, CP is not needed for the movement of the virus but it could help viral spread facilitating its accumulation. As enhancement of viral accumulation can occur due to suppression of RNA silencing, we tested if CP of CfMV is such a suppressor. Using the *Agrobacterium*-mediated transient expression method we infiltrated *N. benthamiana* 16c line, expressing GFP, with *Agrobacterium* carrying the RNA silencing inducer GFP gene together with *Agrobacterium* containing CfMV CP gene. As controls we infiltrated in a similar way GFP together with the empty vector or together with CfMV P1, known as RNA silencing suppressor (Sarmiento et al., 2007). To test if P1 suppressor's activity could be influenced by CP, we also infiltrated a mixture of *Agrobacterium* carrying both genes together with GFP.

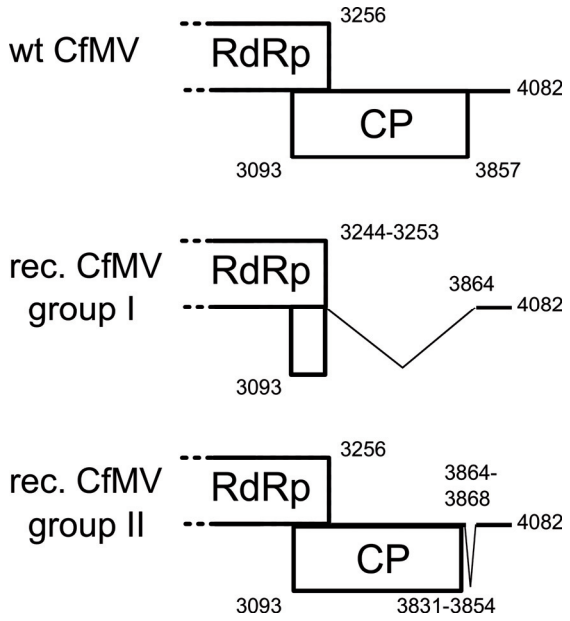


Figure 6. **Overview of spontaneous CfMV deletion mutants detected from non-inoculated upper leaves of oat plants inoculated with CP-EGFP and CPdelta-EGFP.** Wt CfMV – the 3' end organization of CfMV genome. Rec. CfMV group I – group of recombinant viruses which had lost CfMV nt from 3244-3253 to 3864 of the RdRp and CP genes along with the EGFP sequence. Rec. group II – graphic description of recombinants which had lost CfMV nt from 3831-3854 to 3864-3868 of the CP cistron together with EGFP sequence.. The numbering corresponds to the nucleotides in wt CfMV genome.

At 5 dpi, when RNA silencing had just started, the amount of GFP fluorescence inside the infiltrated patch was less in the case of leaves inoculated with CP or with the empty vector (pBin61), compared to the leaves infiltrated with P1 or with the mixture containing CP and P1 (Figure 7 A). The molecular analysis of GFP siRNAs in the infiltrated patches of these leaves confirmed our results: in the case of pBin61 the amount of 21-nt, 22-nt and 24-nt siRNAs was the same as for CP, whereas in the case of P1 or of CP+P1 the amount of siRNAs was considerably lower and only the 21-nt siRNAs were detectable (Figure 7 B). We followed the spread of the RNA silencing signal until 21 dpi. During this time, there was a clear difference between plants infiltrated with pBin61 or CP and the ones infiltrated with P1 or with CP+P1. The amount of red (silenced) tissue was considerably higher in the first group compared to the latter, where the silencing did not reach the upper leaves (data not shown).

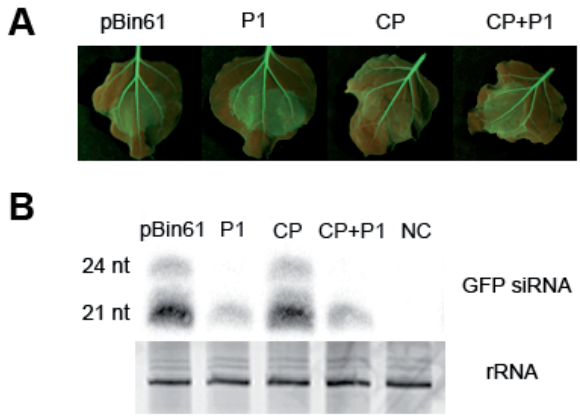


Figure 7**. CP of CfMV is not a suppressor of RNA silencing.** (A) *N. benthamiana* 16c leaves infiltrated with *A. tumefaciens* harboring the constructs shown on the upper part of the panels (P1 of CfMV, CP of CfMV, empty vector pBin61) together with *A. tumefaciens* harboring GFP. Pictures were taken 5 dpi under UV light. (B) Northern blot of RNA isolated from the infiltrated patches of *N. benthamiana* 16c leaves infiltrated as indicated in (A). The upper part shows the radioactive detection of GFP siRNAs (24 nt, 22 nt and 21 nt long). The lower part shows ethidium bromide staining of rRNA as loading control.

Discussion

The arginine-rich region of CP of sobemoviruses and other virus genera has been studied before. The CP of mutants R3L and R5X localizes to cytoplasm and nucleus whereas the wt CP localizes almost exclusively to the nucleus (Olspert et al., 2010). Results obtained here, with viruses containing the same mutations in CP, demonstrate that these mutations have no deleterious effect on virus cell-to-cell and systemic movement as well as on mechanical transmission in oats. The majority of CP-EGFP fusion protein localizes to the cell nucleus when expressed independently from the rest of the virus genome (Olspert et al., 2010). Here we observed that when CP-EGFP was expressed together with the rest of the viral proteins, the fluorescence did not accumulate in the nucleus, but remained evenly distributed between cytoplasm and nucleus. Most probably the CP was interacting with other viral and/or host proteins or with full-length genomic RNA and therefore was not accumulating in the nucleus any more.

Despite these possible interactions, CfMV CP null mutant (noCP) was capable of cell-to-cell as well as systemic movement in all three tested hosts. The fact that CfMV CP is not strictly needed for the infection was further demonstrated by experiments with viruses expressing CP-EGFP and CPdelta-EGFP. As expected, the virus replicated in single epidermal cells and migrated to the mesophyll, beneath these epidermal cells, from where it presumably entered vascular tissue. Spontaneous CfMV recombination mutants detected in plants inoculated with mutants CP-EGFP/CPdelta-EGFP that lacked most of the CP cistron (group I, deleted nt 3254-3864), produced systemic infection as well as symptoms.

In contrast to the results presented in the current study, it has been previously documented that sobemoviruses require P1 as well as CP for systemic movement (Brugidou et al., 1995; Fuentes and Hamilton, 1993; Opalka et al., 1998; Sivakumaran et al., 1998). A RYMV CP mutant failed to infect rice plants systemically, but accumulated in the inoculated leaves, indicating cell-to-cell movement (Brugidou et al., 1995), whereas SCPMV CP initiation codon mutant, analogous to CfMV noCP, was undetectable even in inoculated leaves, but replicated in protoplasts (Sivakumaran et al., 1998). The CP of *Turnip rosette virus*, another sobemovirus, has been reported to facilitate long distance movement of red clover necrotic mosaic dianthovirus (Callaway et al., 2004). Based on protein interaction studies of P1 with either CP or native virions of SeMV, which is very closely related to SCPMV, Chowdry and Savithri (2010) have proposed a model for CP involvement in SeMV movement. We conclude that CfMV in general utilizes a different movement strategy from SCPMV and RYMV and that distinct sobemoviruses have different movement strategies.

The viral RNA genome is usually transported from cell-to-cell and systemically through the vasculature either as virions or as some other form of RNP complex. Trafficking as virions can be now ruled out for CfMV, at least in the tested host plants. Previously we have reported that P1 is indispensable for virus movement and accumulation in oats (Meier et al., 2006). As CfMV P1 is the viral suppressor of RNA silencing (Sarmiento et al., 2007), it still remains unclear whether P1 facilitates virus spread as a suppressor or as a member of the movement RNP complex. Here we demonstrate that CP of CfMV does not possess RNA silencing suppressor activity in 16c *N. benthamiana*, nor is it influencing the suppression activity of P1. It is also possible that CfMV CP contributes to virus spread through enhancement of accumulation. Such claim is supported by experiments with RYMV, which accumulated to higher levels in transgenic plants expressing RYMV CP compared to control plants, indicating that transgenic CP further enhanced virus infection and accumulation (Kouassi et al., 2006). The authors speculated that CP may enhance viral accumulation by influencing replication or host susceptibility or, alternatively, by suppressing RNA silencing. For CfMV, at least this latter possibility seems not to be true.

In our experiments the mutant noCP had reduced accumulation compared to wt CfMV in both inoculated and systemically infected leaves of all three challenged hosts. In the case of wt CfMV infection, the genomic and subgenomic RNAs were usually detected at the same level and if there was a difference, then sgRNA was detected at a lower level (Figure 2-3). However, in the case of noCP, the sgRNA was generally detected at higher levels than genomic RNA, suggesting an involvement of CP in the regulation of sgRNA and/or genomic RNA synthesis. CP accumulation can be the signal to switch from sgRNA production to the synthesis of genomic RNA, which is then encapsidated by CP.

Interestingly, we were unable to transmit noCP mechanically, whereas R3L and R5X were transmittable. This demonstrates that CP is necessary for an efficient transmission and that mutations in the arginine-rich region do not affect the viral RNA-CP complex involved in this transmission. The mutant noCP should be, in theory, transmissible as well, since viral RNA is all what is needed for initiating the infection. Nevertheless, we have observed that mechanical inoculation with *in vitro* synthesized CfMV RNA is less efficient than biolistic inoculation with the same RNA (our unpublished observations). When compared to the R3L and R5X (that were transmissible), mutant noCPdid not have significantly lower viral RNA levels in systemically infected leaves (Figure 2). Most likely the formation of virus particles is necessary for the efficient CfMV transmission. However based on the data currently available, we cannot rule out that in wild-type situation CfMV is moving as some other type of RNP complex, instead of particles.

Mutant R5X failed to produce systemic infection in mechanically inoculated plants. Compared to plants biolistically inoculated with wt CfMV or R3L, the plants inoculated with R5X usually appeared to have a lower viral load in systemically infected leaves (Figure 2 and our unpublished observations). Since R5X has also a mutation in the RdRp, we propose that this deficiency is due to the mutation in RdRp. Thus, we assume that the movement of R5X itself was normal, as transmission of the virus was also successful.

In the case of CP-EGFP, the EGFP sequence was longer retained within the genome, as progeny viruses replicating in the mesophyll still expressed EGFP, whereas for CPdelta-EGFP the EGFP gene was already lost when the virus entered the mesophyll cells. The recombinant viruses we detected could be divided into two groups based on how much of the CP cistron was retained. Furthermore, it seems that, if available, the virus usually maintains as much of the CP sequence as possible. Foreign EGFP sequence seems to be the trigger for recombination because we have not encountered recombinants in plants infected with noCP mutant.

It is somewhat surprising that the mutants isolated from non-inoculated upper leaves, which had lost the EGFP and CP coding sequences, all contained almost the entire3'UTR. The 3' UTR starts at nt 3858 and the recombinant mutants had retained the sequence starting from nt 3865 or 3869. This leads us to speculate that the 3' UTR might contain sequences or structural elements important for the transport, as the full-length 3' UTR is not needed for replication, translation and accumulation in oats (Olspert, unpublished results). Another explanation is that nt 3865-3869 simply contain a hot-spot for recombination. These two hypotheses are, of course, not mutually exclusive.

Altogether we have demonstrated that in different pathosystems individual sobemoviruses can exploit alternative cell-to-cell and long distance movement strategies. It would be interesting to determine whether the CfMV P1 RNA silencing suppressor activity can be uncoupled from the movement function and to determine the composition of the movement RNP complex if such exists at all. The mechanism by which CP contributes to higher virus accumulation of CfMV is an interesting object of further investigation as well.

Acknowledgements

We wish to thank Signe Nõu for excellent plant care, Heiti Paves for assistance in microscopy and Jaanus Suurväli for his earlier efforts with EGFP containing viruses. This work was supported by Estonian Science Foundation grant no. 7363. We thank Professor Sir David Baulcombe for *N. benthamiana* 16cline, pBin61 and GFP in pBin61, supplied by PBL.

Experimental procedures

Construction of CfMV mutants

Base numbering of constructs generated for this study corresponds to CfMV Norwegian isolate (Mäkinen1995). All CfMV cDNA clones were created by modifying the original cDNA clone (Meier2006). It was decided to simplify plant inoculation by removing the necessity for *in vitro* RNA synthesis before plant inoculation. To achieve this, the CfMV genome was cloned downstream of CaMV 35S promoter and *Hepatitis delta virus* ribozyme together with nopaline synthetase terminator were introduced after the genome in order to maintain the exact 3' end of the genome after transcription. The exchange of the promoter and the addition of the ribozyme coupled with a terminator were carried out using overlap-extention-PCR with appropriate primers and standard cloning techniques.

cDNA clones of CfMV containing mutations were also generated using overlap-PCR. Mutations of R5X and R3L were introduced by primers described in Olspert *et al.* (2010) and in mutant noCP the CP initiation sequence ATGATG was mutated to ACGACG. Using primers containing the mutations, the CfMV fragments containing nt 3096-3853 and 1604-3162 were produced and merged in a following round of PCR, so that a fragment corresponding to CfMV nt 1604-3853 was obtained. The latter fragment was used to introduce the mutations to the CfMV cDNA by employing *Nco*I sites at positions 2508 and 3619. In the case of virus mutant noCP, in addition to the CP initiation codon mutation a stop codon was introduced to CP reading frame after the overlap with RdRp gene (Figure 1B). This was achieved by cleaving the plasmid with *XmaJ*I restrictase at CfMV position 3311, filling in the termini and re-ligating the plasmid. This produced a reading frame switch starting from position 3311 and an in frame stop codon at positon 3343.

Virus clones expressing CP EGFP fusions were generated by replacing the CfMV sequence between *XmaJ*I and *Pst*I restriction sites at positions 3311 and 3869, respectively. In the case of CP-EGFP, the fusion sequence was obtained by PCR using a plasmid expressing the fusion protein (Olspert2010) as the template. For CPdelta-EGFP EGFP primers with aforementioned restriction sites were used to generate the appropriate fragment.

All DNA constructs used in this work were verified by sequencing.

Plant inoculation and virus detection

12-14 days old oat (cv. Jaak) plants were inoculated biolistically (Helios,

BioRad) with CfMV constructs according to manufacturer's instructions. Samples from the inoculated leaves were collected at 14 dpi and from upper leaves at 21 dpi. Each experiment was repeated at least twice and with a minimum of 8 plants per construct. Follow-up analysis in wheat (cv. Zebra) and barley (cv. Kymppi) were conducted once with 16 plants per construct. Plant tissue was homogenized with TissueLyzer (QIAgen) and total RNA was extracted from samples according to (Logemann et al., 1987) and subjected to standard Northern blot analysis. The RNA was electrophoretically separated on 1% agarose/formaldehyde gel and transferred to Hybond N+ membrane (GE Healthcare). Viral RNAs were detected by hybridization with [α-^{32} P]dCTP-labeled probe specific for the CfMV CP gene (nt 3093–3857). The analysis of RNA by RT-PCR was carried out with primers detecting nt 2749-4082 of the positive strand of viral RNA The obtained RT-PCR fragments were purified from the agarose gel and the region containing the mutations was sequenced. Protein samples from the same material were obtained in parallel from the cell debris collected after the first centrifugation of RNA extraction. The pellet was suspended in PBS-Tween buffer and total protein was precipitated from the supernatant with TCA. Subsequently, the protein extract was analyzed on 12.5% SDS-PAGE, blotted onto Hybond C membrane (GE Healthcare) and probed with rabbit anti-CP polyserum (Tamm1999). Goat anti-rabbit HRP conjugate was used for detection. Ribulose-1,5-bisphosphate carboxylase oxygenase large subunit was visualized with Poucean S stain for calibration.

For the analysis of mechanical transmission, the upper leaves of infected plants were ground in liquid nitrogen and the homogenate suspended in 10 volumes (w/v) of 100 mM potassium phosphate buffer (pH 7.0) containing 0.5% Celite. This suspension was used to mechanically inoculate plants and the following analysis of virus infection was done as in the case of bombarded plants.

Agroinfiltration assay and siRNA analysis

CfMV CP coding sequence nt 3096–4082 was amplified with primers containing the appropriate restriction sites, excised with *Bam*HI and *Fsp*AI and cloned into pBin61 between the 35S promoter and Nos terminator to give pBin61-CP. 35S-CP refers to *Agrobacterium tumefaciens* containing pBin61-CP.

The *A. tumefaciens* strain used throughout the experiment was C58C1. Equal volumes of 35S-CP and 35S-GFP (*A. tumefaciens* carrying GFP, kindly provided by D. Baulcombe), as well as of 35S-P1 [*A. tumefaciens* containing CfMV P1, (Sarmiento et al., 2007)] and 35S-GFP or 35S-P1 and 35S-CP together with 35S-GFP (total volume divided in three parts) were mixed and co-infiltrated (OD_{600} = 1) to *N. benthamiana* line 16c (kind gift of D. Baulcombe) leaves of 4-week-old

plants, as described previously (Hamilton et al., 2002). As a control, *A. tumefaciens* carrying the empty binary vector pBin61 was infiltrated together with 35S-GFP. Three independent experiments were carried out, each including at least 5 infiltrated plants for each mixture (35S-GFP+35S-CP, 35S-GFP+35S-P1, 35S-GFP+35S-CP+35S-P1 and 35S-GFP+pBin61). Infiltrated plants were kept in a plant chamber at 22 °C under a 16-h photoperiod. GFP fluorescence was monitored using a hand-held 100 W, long-wave UV lamp (Black-Ray B-100AP, Ultraviolet Products) until 21 dpi. Plants were photographed with Pentax K200D digital camera and pictures were processed with Adobe Photoshop CS2 (version 9.0.2).

Total RNA was extracted from the infiltrated patches 5 dpi using TRIzol reagent (Invitrogen) according to the manufacturer's instructions. 15 µg of total RNA were denatured and loaded on 15 % polyacrylamide gel (19:1 ratio of acrylamide to bis-acrylamide, 8 M urea). The gel was run at 400 V for 3,5 h and then cut in the middle. The lower half of the gel was transferred to Hybond N+-membrane by electroblotting in 0.5 X TBE buffer at 10 V overnight. ULTRAhyb-Oligo buffer (Ambion) was used for overnight hybridization at 42 °C. As a radioactive probe DNA oligo containing a sequence complementary to GFP (5'-CTCTTGAAGAAGTCGTGCCGCTTCATATGA-3') was end-labeled with ^{32}P by T4 polynucleotide kinase (Fermentas) and purified through NICK Sephadex G-50 columns (GE Healthcare) according to manufacturers' protocols. The membrane was finally washed twice with 2 X SSC, 0.1% SDS for 30 min at 42 °C. Radioactive signal was detected after one hour exposure using Personal Molecular Imager FX (BioRad). As a reference marker we used a 30-nt [^{32}P]-end labeled DNA oligo.

References

Benitez-Alfonso, Y., Faulkner, C., Ritzenthaler, C. and Maule, A.J. (2010). Plasmodesmata: gateways to local and systemic virus infection. Mol Plant Microbe Interact *23*, 1403-1412.
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*, 79-86.
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*, 108-115.
Callaway, A., Giesman-Cookmeyer, D., Gillock, E.T., Sit, T.L. and Lommel, S.A. (2001). The multifunctional capsid proteins of plant RNA viruses. Annu Rev Phytopathol *39*, 419-460.
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.
Chowdhury, S.R. and Savithri, H.S. (2011). Interaction of *Sesbania mosaic virus* movement protein with the coat protein--implications for viral spread. FEBS J *278*, 257-272.
Dalmay, T., Rubino, L., Burgyán, J. and Russo, M. (1992). Replication and movement of a coat protein mutant of cymbidium ringspot tombusvirus. Mol Plant Microbe Interact *5*, 379-383.
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)*, 1903-1910.
Gopinath, K. and Kao, C.C. (2007). Replication-independent long-distance trafficking by viral RNAs in *Nicotiana benthamiana*. Plant Cell *19*, 1179-1191.
Hamilton, A., Voinnet, O., Chappell, L. and Baulcombe, D. (2002). Two classes of short interfering RNA in RNA silencing. EMBO J *21*, 4671-4679.
Kehr, J. and Buhtz, A. (2008). Long distance transport and movement of RNA through the phloem. J Exp Bot *59*, 85-92.
Kouassi, N.K., Chen, L., Siré, C., Bangratz-Reyser, M., Beachy, R.N., Fauquet, C.M. and Brugidou, C. (2006). Expression of rice yellow mottle virus coat protein enhances virus infection in transgenic plants. Arch Virol *151*, 2111-2122.
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*, 371-382.
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*, 317-327.

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*, 299-310.
Logemann, J., Schell, J. and Willmitzer, L. (1987). Improved method for the isolation of RNA from plant tissues. Anal Biochem *163*, 16-20.
Lokesh, G.L., Gowri, T.D.S., Satheshkumar, P.S., Murthy, M.R.N. and Savithri, H.S. (2002). A molecular switch in the capsid protein controls the particle polymorphism in an icosahedral virus. Virology *292*, 211-223.
Lucas, W.J. (2006). Plant viral movement proteins: agents for cell-to-cell trafficking of viral genomes. Virology *344*, 169-184.
Mäkinen, K., Naess, V., Tamm, T., Truve, E., Aaspõllu, A. and Saarma, M. (1995). The putative replicase of the cocksfoot mottle sobemovirus is translated as a part of the polyprotein by -1 ribosomal frameshift. Virology *207*, 566-571.
Meier, M. and Truve, E. (2007). Sobemoviruses possess a common CfMV-like genomic organization. Arch Virol *152*, 635-640.
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*, 321-326.
Nagano, H., Mise, K., Furusawa, I. and Okuno, T. (2001). Conversion in the requirement of coat protein in cell-to-cell movement mediated by the cucumber mosaic virus movement protein. J Virol *75*, 8045-8053.
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*, 423-431.
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*, 3323-3328.
Petty, I.T. and Jackson, A.O. (1990). Mutational analysis of barley stripe mosaic virus RNA beta. Virology *179*, 712-718.
Roy Chowdhury, S. and Savithri, H.S. (2011). Interaction of *Sesbania mosaic virus* movement protein with VPg and P10: implication to specificity of genome recognition. PLoS One *6*, e15609.
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*, 95-99.
Satheshkumar, P.S., Lokesh, G.L., Murthy, M.R. and Savithri, H.S. (2005). The role of arginine-rich motif and beta-annulus in the assembly and stability of *Sesbania mosaic virus* capsids. J Mol Biol *353*, 447-458.
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*, 328-335.

Scholthof, H.B. (2005). Plant virus transport: motions of functional equivalence. Trends Plant Sci *10*, 376-382.
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*, 376-386.
Takeda, A., Nakamura, W., Sasaki, N., Goto, K., Kaido, M., Okuno, T. and Mise, K. (2005). Natural isolates of *Brome mosaic virus* with the ability to move from cell to cell independently of coat protein. J Gen Virol *86*, 1201-1211.
Tamm, T. and Truve, E. (2000a). RNA-binding activities of cocksfoot mottle sobemovirus proteins. Virus Res *66*, 197-207.
Tamm, T. and Truve, E. (2000b). Sobemoviruses. J Virol *74*, 6231-6241.
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, 14147-14152.

ABSTRACT

The genus *Sobemovirus* comprises of small spherical plant viruses with a single positive-sense ssRNA genome of approximately 4 – 4.5 kb. Sobemoviruses have a viral protein genome-linked (VPg) covalently attached to the 5' end of genomic and subgenomic RNAs. The genome contains four open reading frames. The 5' proximal encodes P1, which is the RNA silencing suppressor. The viral coat protein (CP) is expressed from the 3' proximal ORF3 via subgenomic RNA. The central part of the genome encodes the viral polyprotein, which is processed to at least the following domains: protease, VPg and RNA-dependent RNA polymerase (RdRp). The RdRp is expressed utilizing -1 programmed frameshift mechanism. The icosahedral virions are assembled according to T = 3 quasisymmetry, contain 180 CP molecules and are approximately 30 nm in diameter.

In addition to sobemoviruses, at least picornaviruses, caliciviruses, comoviruses, poleroviruses, enamoviruses, nepoviruses and potyviruses have been reported to contain VPgs. The VPgs from these genera have been demonstrated to be linked to RNA over a phosphodiester bond formed between the 5' end of RNA and the hydroxyl group of either serine or tyrosine. Besides being linked to RNA the VPgs from different genera have little in common. They vary considerably in size and are not homologous. For some viruses the main biological role of VPg is to serve as the primer for viral RNA synthesis while for others VPg requirement for viral RNA translation has been more clearly established.

Currently available sobemoviral VPg sequences share only one common short motif and lack a suitable conserved amino acid residue that could be used universally for RNA linking. Therefore, in the current study it was decided to analyze the VPg-RNA linkage of several sobemoviruses. We determined that *Cocksfoot mottle virus* (CfMV) VPg is linked to RNA through a tyrosine residue at position five, while *Rice yellow mottle virus* (RYMV) and *Ryegrass mottle virus* (RGMoV) utilize a serine residue at the first position of VPg for RNA linking. Moreover, we demonstrated that *Southern bean mosaic virus* (SBMV) VPg is attached to RNA over a threonine residue at the first position. This demonstrates unprecedented positional and chemical versatility regarding RNA linking within one genus. Furthermore, VPg linking to RNA over a threonine residue has not been described before. In addition, we identified the termini of the corresponding VPgs and discovered several phosphorylation sites in each VPg.

The biological aspects of sobemoviral CPs besides virion formation are not well defined. The sobemoviral CP is an RNA binding protein, particularly the N-terminal part is rich in basic amino acid residues and contains an arginine-rich motif (ARM). For some members of the genus the CP has been demonstrated to

be dispensable for virus replication but indispensable for movement.

In this thesis CfMV CP subcellular localization and CP involvement in virus movement were addressed. The analysis of CfMV CP localization by transiently expressed enhanced green fluorescent protein (EGFP) and CP hybrids revealed that the protein is localized to the nucleus. In conclusion, we determined that CfMV CP N-terminus contains two nuclear localization signals. Furthermore, microinjection studies demonstrated that the wild-type CP was capable of transporting RNA to the nucleus in a sequence-nonspecific manner. These results suggest that the nucleus may be involved in CfMV infection. However, the same CP-EGFP hybrid displayed both nuclear and cytoplasmic distribution in CfMV infected cells. Further investigation of CfMV movement revealed that the CP is dispensable for cell-to-cell as well as systemic movement in different hosts. However, the accumulation of the CP-deficient virus was usually reduced when compared to wild type CfMV. In addition we demonstrated that CP does not have RNA silencing suppressor activity nor does it modulate the RNA suppression activity of P1. Taken together this suggests that CfMV can move in the infected plants without the help of the CP and that CfMV CP ability to enhance virus infection is not realized through the suppression of RNA silencing.

In conclusion, this thesis reveals novel aspects of the biological properties of sobemoviral structural proteins and, hopefully, provides the basis for further investigations.

KOKKUVÕTE

Sobemoviirused on taimeviiruste perekond, mille liikmetel on väikesed sfäärilised virionid ning nende genoomiks on üheahelaline positiivse orientatsiooniga RNA pikkusega 4 - 4,5 kb. Sobemoviiruste genoomse ja subgenoomse RNA 5' otsadesse on kovalentselt seotud viiruse genoomiseoseline valk VPg. Sobemoviiruste genoomid sisaldavad nelja avatud lugemisraami. Kõige 5' poolsemalt lugemisraamilt sünteesitakse valk P1, mis on geenivaigistamise supressor. Viiruse kattevalku (CP) kodeeriv lugemisraam asub genoomi 3' otsa lähedal ning valgusüntees toimub sellelt subgenoomse RNA vahendusel. Sobemoviiruste genoomide keskosas asuvad kaks teineteisega kattuvat lugemisraami, milledelt transleeritakse polüproteiini, mis protsessitakse hiljem vähemalt proteaasiks, VPg-ks ja RNA-sõltuvaks RNA polümeraasiks (RdRp). RdRp sünteesiks kasutatakse -1 ribosomaalse raaminihke mehanismi. Sobemoviiruste virionid on umbes 30 nm läbimõõduga ning sisaldavad 180 CP molekuli vastavalt T = 3 sümmeetriale.

Lisaks sobemoviirustele on VPg olemasolu täheldatud ka vähemalt picorna-, calici-, como-, polero-, enamo-, nepo- ja potyviirustel. Nende perekondade viirustel on näidatud, et VPg on RNA-ga seotud fosfodiester sidemega, mis on moodustunud RNA 5' otsa ja seriini või türosiini hüdroksüülrühma vahel. Erinevate perekondade VPg-d pole järjestuselt omavahel sarnased ning ka nende suurused on küllaltki erinevad. Osadel viirustel kasutatakse VPg-d praimeriks viraalse RNA sünteesil, samas teiste puhul on näidatud VPg rolli viraalse RNA translatsioonis.

Nende sobemoviiruste, millede genoomid on sekveneeritud, VPg primaarjärjestustes esineb ainult üks kõigile omane motiiv ning ei leidu ühtegi sobivat konserveerunud aminohappe jääki, mille külge võidaks siduda RNA. Seetõttu otsustati antud töös määrata mitme erineva sobemoviiruse puhul VPg aminohappe jääk, mille külge seotakse viiruse RNA. Keraheina laiguviiruse (CfMV) puhul osutus selleks türosiin VPg viiendas positsioonis, samas kui riisi kollalaiksuse viiruse (RYMV) ja raiheina laiguviiruse (RGMoV) puhul leiti, et RNA on seotud VPg N-terminaalse seriini külge. Seevastu selgus, et põldoa mosaiigiviiruse (SBMV) RNA on VPg külge seotud üle N-terminaalse treoniini. Need tulemused näitavad, et sobemoviiruste perekonnas on VPg ja RNA sidumiseks kasutatavate aminohapete positsioon ja füüsikalis-keemilised omadused üllatavalt mitmekesised. RNA sidumist treoniini jäägi kaudu pole varem kirjeldatud ühegi teise viiruse puhul. Täiendavalt määrati katsete käigus eelpool nimetatud viiruste VPg-de terminused ning leiti iga VPg puhul mitu fosforüleeritud aminohappe jääki.

Hetkel pole teada, kas lisaks virioni moodustamisele on sobemoviiruste CP-del täiendavaid bioloogilisi funktsioone. On teada, et sobemoviiruste CP-d seovad RNA-d, nende N-terminaalsed regioonid on rikkad aluseliste

aminohappe jääkide poolest ning seal asub kõigile perekonna liikmetele omane arginiinirikas motiiv. Mõnede viiruste puhul on näidatud, et CP ei ole vajalik replikatisoonil, kuid osaleb viiruse liikumises peremeestaimes.

Käesolevas töös uuriti CfMV CP rakusisest lokalisatsiooni ning osalemist viiruse liikumises. Transientselt ekspresseeritud CP ja roheliselt fluorestseeruva valgu (EGFP) hübriidid lokaliseerusid raku tuuma. Selgus, et CfMV CP N-terminaalses alas paikneb kaks tuumalokalisatsiooni signaali. Täiendavad mikroinjektsiooni katsed näitasid, et CfMV CP suudab endaga koos transportida rakutuuma ka RNA-d sõltumata selle järjestusest. Need tulemused viitavad sellele, et raku tuum ning selle komponendid võivad osaleda CfMV infektsioonitsüklis. Samas hilisemates katsetes, kus jälgiti samade CP-EGFP hübriide rakusisest lokalisatsiooni CfMV-ga nakatunud rakkudes, täheldati hübriidvalgu lokalisatsiooni ühtlaselt nii tuumas kui ka tsütoplasmas. Viiruse liikumise analüüsil erinevates peremeestaimedes selgus, et CfMV CP pole vajalik ei rakust rakku ega süsteemseks liikumiseks. Samas akumuleerus ilma CP-ta viirus madalamal tasemel kui metsikut tüüpi CfMV. Lisakatsed näitasid, et CfMV CP ei ole geenivaigistamise suppressor ega mõjuta CfMV valgu P1 poolt teostatavat vaigistamise suppressiooni. Antud tulemuste põhjal võib järeldada, et CfMV on võimeline liikuma peremeestaimes ilma CP-ta ning viimase positiivne mõju viiruse akumulatsioonile ei ole seotud geenivaigistamise suppressiooniga.

Kokkuvõtvalt kirjeldatakse käesolevas töös sobemoviiruste struktuursete valkude seni teadmata bioloogilisi omadusi, mis võimaldab tulevikus paremini mõista ja uurida nende valkude funktsioone.

CURRICULUM VITAE

Personal data

Name:	Allan Olspert
Date and place of birth:	15.12.1979, Tallinn, Estonia
Citizenship:	Estonian

Contact information

Aadress: of Tallinn	Department of Gene Technology, Tallinn University Technology (TUT), Akadeemia tee 15, 12618
Phone:	+3726204421
e-mail:	allan.olspert@ttu.ee

Education

2004 – 2011 TUT Faculty of Science, Department of Gene Technology, PhD

2002 – 2004 TUT Faculty of Science, Department of Gene Technology, MSc

1998 – 2002 TUT Faculty of Chemistry, Centre for Gene Techn., BSc cum laude

Employment

2005 – TUT, Faculty of Science, Department of Gene Technology, researcher

2009 – 2011 Competence Center for Cancer Research, researcher

2002 – 2005 TUT, Department of Gene Technology, research assistant

Supervised theses

Kristjan Kamsol, MSc, 2010, (sup) Erkki Truve, Allan Olspert, Mutational analysis of the phosphorylation sites of *Cocksfoot mottle virus* VPg protein, TUT, Faculty of Science, Department of Gene Technology

Jaanus Suurväli, MSc, 2008, (sup) Erkki Truve, Allan Olspert, Mutational analysis of the *Cocksfoot mottle virus* ribosomal frameshifting signal, TUT, Faculty of Science, Department of Gene Technology

Kristjan Kamsol, BSc, 2008, (sup) Erkki Truve, Allan Olspert, Testing of silencing-vectors based on cocksfoot mottle virus, TUT, Faculty of Science, Department of Gene Technology

Kristiina Talts, BSc, 2008, (sup) Erkki Truve, Allan Olspert, Construction of cocksfoot mottle virus-based gene silencing vectors, TUT, Faculty of Science, Department of Gene Technology

Jaanus Suurväli, BSc, 2005, (sup) Erkki Truve, Allan Olspert, Analysis of cocksfoot mottle virus cDNA clone containing GFP fused to the coat protein, TUT, Faculty of Science, Department of Gene Technology

Publications

Olspert, A.; Arike, L.; Peil, L.; Truve, E. (2011). Viral RNA linked to VPg over a threonine residue. FEBS Letters, 585, 2979 - 2985.

Olspert, A.; Peil, L.; Hébrard, E.; Fargette, D.; Truve, E. (2011). Protein-RNA linkage and post-translational modifications of two sobemovirus VPgs. Journal of General Virology, 92, 445 - 452.

Olspert, A.; Paves, H.; Toomela, R.; Tamm, T.; Truve, E. (2010). Cocksfoot mottle sobemovirus coat protein contains two nuclear localization signals. Virus Genes, 40, 423 - 431.

Tamm, T.; Suurväli, J.; Lucchesi, J.; **Olspert, A**.; Truve, E. (2009). Stem-loop structure of *Cocksfoot mottle virus* RNA is indispensable for programmed −1 ribosomal frameshifting . Virus Research, 146, 73 - 80.

Meier, M.; **Olspert, A**.; Sarmiento, C.; Truve, E. (2008). Sobemoviruses. Mahy, Brian; Van Regenmortel, Marc (Editors). Encyclopedia of Virology, 3rd Edition (644 - 652). Oxford: Elsevier

Meier, M.; Paves, H.; **Olspert, A**.; Tamm, T.; Truve, E. (2006). P1 protein of *Cocksfoot mottle virus* is indispensable for the systemic spread of the virus. Virus Genes, 32, 321 - 326.

Nigul, L.; **Olspert, A.**; Meier, M.; Paves, H.; Talpsep, T.; Truve, E (2004). New plant vectors for protein tagging with E2 epitope. Plant Molecular Biology Reporter, 22, 399 - 407.

ELULULOOKIRJELDUS

Isikuandmed	
Nimi:	Allan Olspert
Sünniaeg ja -koht:	15.12.1979, Tallinn, Eesti
Kodakondsus:	Eesti
Kontaktandmed	
Aadress:	Geenitehnoloogia Instituut, Tallinna Tehnikaülikool, Akadeemia tee 15, 12618 Tallinn
Telefon:	+3726204421
e-post:	allan.olspert@ttu.ee

Hariduskäik

2004 – 2011 TTÜ Matemaatika-loodusteaduskond, geenitehnoloogia doktorant

2002 – 2004 TTÜ Matemaatika-loodusteaduskond, geenitehnoloogia magister

1998 – 2002 TTÜ Keemiateaduskond, cum laude geenitehnoloogia bakalaureus

Teenistuskäik

2009 – Vähiuuringute Tehnoloogia Arenduskeskus AS, teadur

2005 – Tallinna Tehnikaülikool, Matemaatika-loodusteaduskond, Geenitehnoloogia instituut, Geenitehnoloogia õppetool, teadur

2002 – 2005 Tallinna Tehnikaülikool, Geenitehnoloogia instituut, insener

Juhendatud lõputööd

Kristjan Kamsol, magistrikraad, 2010, (juh) Erkki Truve, Allan Olspert, Keraheina laiguviiruse VPg valgu fosforülatsioonisaitide mutatsioonanalüüs, TTÜ, Matemaatika-loodusteaduskond, Geenitehnoloogia instituut

Jaanus Suurväli, magistrikraad, 2008, (juh) Erkki Truve, Allan Olspert, Keraheina laiguviiruse ribosomaalse raaminihke signaaljärjestuse mutageneesanalüüs, TTÜ, Matemaatika-loodusteaduskond, Geenitehnoloogia instituut

Kristjan Kamsol, bakalaureusekraad, 2008, (juh) Erkki Truve, Allan Olspert, Keraheina laiguviiruse põhjal konstrueeritud vaigistamisvektorite testimine, TTÜ, Matemaatika-loodusteaduskond, Geenitehnoloogia instituut

Kristiina Talts, bakalaureusekraad, 2008, (juh) Erkki Truve, Allan Olspert, Geenivaigistamise vektorite konstrueerimine keraheina laiguviiruse põhjal, TTÜ, Matemaatika-loodusteaduskond, Geenitehnoloogia instituut

Jaanus Suurväli, bakalaureusekraad, 2005, (juh) Erkki Truve, Allan Olspert, GFP – kattevalgu liitvalku sisaldava keraheina laiguviiruse cDNA klooni infektsioonilisuse analüüs, TTÜ, Matemaatika-loodusteaduskond, Geenitehnoloogia instituut

Publikatsioonid

Olspert, A.; Arike, L.; Peil, L.; Truve, E. (2011). Viral RNA linked to VPg over a threonine residue. FEBS Letters, 585, 2979 - 2985.

Olspert, A.; Peil, L.; Hébrard, E.; Fargette, D.; Truve, E. (2011). Protein-RNA linkage and post-translational modifications of two sobemovirus VPgs. Journal of General Virology, 92, 445 - 452.

Olspert, A.; Paves, H.; Toomela, R.; Tamm, T.; Truve, E. (2010). Cocksfoot mottle sobemovirus coat protein contains two nuclear localization signals. Virus Genes, 40, 423 - 431.

Tamm, T.; Suurväli, J.; Lucchesi, J.; **Olspert, A**.; Truve, E. (2009). Stem-loop structure of *Cocksfoot mottle virus* RNA is indispensable for programmed −1 ribosomal frameshifting . Virus Research, 146, 73 - 80.

Meier, M.; **Olspert, A**.; Sarmiento, C.; Truve, E. (2008). Sobemoviruses. Mahy, Brian; Van Regenmortel, Marc (Toim.). Encyclopedia of Virology, 3rd Edition (644 - 652). Oxford: Elsevier

Meier, M.; Paves, H.; **Olspert, A**.; Tamm, T.; Truve, E. (2006). P1 protein of *Cocksfoot mottle virus* is indispensable for the systemic spread of the virus. Virus Genes, 32, 321 - 326.

Nigul, L.; **Olspert, A.**; Meier, M.; Paves, H.; Talpsep, T.; Truve, E (2004). New plant vectors for protein tagging with E2 epitope. Plant Molecular Biology Reporter, 22, 399 - 407.

DISSERTATIONS DEFENDED AT TALLINN UNIVERSITY OF TECHNOLOGY ON *NATURAL AND EXACT SCIENCES*

1. **Olav Kongas**. Nonlinear Dynamics in Modeling Cardiac Arrhytmias. 1998.

2. **Kalju Vanatalu**. Optimization of Processes of Microbial Biosynthesis of Isotopically Labeled Biomolecules and Their Complexes. 1999.

3. **Ahto Buldas**. An Algebraic Approach to the Structure of Graphs. 1999.

4. **Monika Drews**. A Metabolic Study of Insect Cells in Batch and Continuous Culture: Application of Chemostat and Turbidostat to the Production of Recombinant Proteins. 1999.

5. **Eola Valdre**. Endothelial-Specific Regulation of Vessel Formation: Role of Receptor Tyrosine Kinases. 2000.

6. **Kalju Lott**. Doping and Defect Thermodynamic Equilibrium in ZnS. 2000.

7. **Reet Koljak**. Novel Fatty Acid Dioxygenases from the Corals *Plexaura homomalla* and *Gersemia fruticosa.* 2001.

8. **Anne Paju**. Asymmetric oxidation of Prochiral and Racemic Ketones by Using Sharpless Catalyst. 2001.

9. **Marko Vendelin**. Cardiac Mechanoenergetics *in silico*. 2001.

10. **Pearu Peterson**. Multi-Soliton Interactions and the Inverse Problem of Wave Crest. 2001.

11. **Anne Menert**. Microcalorimetry of Anaerobic Digestion. 2001.

12. **Toomas Tiivel**. The Role of the Mitochondrial Outer Membrane in *in vivo* Regulation of Respiration in Normal Heart and Skeletal Muscle Cell. 2002.

13. **Olle Hints**. Ordovician Scolecodonts of Estonia and Neighbouring Areas: Taxonomy, Distribution, Palaeoecology, and Application. 2002.

14. **Jaak Nõlvak**. Chitinozoan Biostratigrapy in the Ordovician of Baltoscandia. 2002.

15. **Liivi Kluge**. On Algebraic Structure of Pre-Operad. 2002.

16. **Jaanus Lass**. Biosignal Interpretation: Study of Cardiac Arrhytmias and Electromagnetic Field Effects on Human Nervous System. 2002.

17. **Janek Peterson**. Synthesis, Structural Characterization and Modification of PAMAM Dendrimers. 2002.

18. **Merike Vaher**. Room Temperature Ionic Liquids as Background Electrolyte Additives in Capillary Electrophoresis. 2002.

19. **Valdek Mikli**. Electron Microscopy and Image Analysis Study of Powdered Hardmetal Materials and Optoelectronic Thin Films. 2003.

20. **Mart Viljus**. The Microstructure and Properties of Fine-Grained Cermets. 2003.

21. **Signe Kask**. Identification and Characterization of Dairy-Related *Lactobacillus.* 2003

22. **Tiiu-Mai Laht**. Influence of Microstructure of the Curd on Enzymatic and Microbiological Processes in Swiss-Type Cheese. 2003.

23. **Anne Kuusksalu**. 2–5A Synthetase in the Marine Sponge *Geodia cydonium*. 2003.

24. **Sergei Bereznev**. Solar Cells Based on Polycristalline Copper-Indium Chalcogenides and Conductive Polymers. 2003.

25. **Kadri Kriis**. Asymmetric Synthesis of C_2-Symmetric Bimorpholines and Their Application as Chiral Ligands in the Transfer Hydrogenation of Aromatic Ketones. 2004.

26. **Jekaterina Reut**. Polypyrrole Coatings on Conducting and Insulating Substracts. 2004.

27. **Sven Nõmm**. Realization and Identification of Discrete-Time Nonlinear Systems. 2004.

28. **Olga Kijatkina**. Deposition of Copper Indium Disulphide Films by Chemical Spray Pyrolysis. 2004.

29. **Gert Tamberg**. On Sampling Operators Defined by Rogosinski, Hann and Blackman Windows. 2004.

30. **Monika Übner**. Interaction of Humic Substances with Metal Cations. 2004.

31. **Kaarel Adamberg**. Growth Characteristics of Non-Starter Lactic Acid Bacteria from Cheese. 2004.

32. **Imre Vallikivi**. Lipase-Catalysed Reactions of Prostaglandins. 2004.

33. **Merike Peld**. Substituted Apatites as Sorbents for Heavy Metals. 2005.

34. **Vitali Syritski**. Study of Synthesis and Redox Switching of Polypyrrole and Poly(3,4-ethylenedioxythiophene) by Using *in-situ* Techniques. 2004.

35. **Lee Põllumaa**. Evaluation of Ecotoxicological Effects Related to Oil Shale Industry. 2004.

36. **Riina Aav**. Synthesis of 9,11-Secosterols Intermediates. 2005.

37. **Andres Braunbrück**. Wave Interaction in Weakly Inhomogeneous Materials. 2005.

38. **Robert Kitt**. Generalised Scale-Invariance in Financial Time Series. 2005.

39. **Juss Pavelson**. Mesoscale Physical Processes and the Related Impact on the Summer Nutrient Fields and Phytoplankton Blooms in the Western Gulf of Finland. 2005.

40. **Olari Ilison**. Solitons and Solitary Waves in Media with Higher Order Dispersive and Nonlinear Effects. 2005.

41. **Maksim Säkki**. Intermittency and Long-Range Structurization of Heart Rate. 2005.

42. **Enli Kiipli**. Modelling Seawater Chemistry of the East Baltic Basin in the Late Ordovician–Early Silurian. 2005.

43. **Igor Golovtsov**. Modification of Conductive Properties and Processability of Polyparaphenylene, Polypyrrole and polyaniline. 2005.

44. **Katrin Laos**. Interaction Between Furcellaran and the Globular Proteins (Bovine Serum Albumin β-Lactoglobulin). 2005.

45. **Arvo Mere**. Structural and Electrical Properties of Spray Deposited Copper Indium Disulphide Films for Solar Cells. 2006.

46. **Sille Ehala**. Development and Application of Various On- and Off-Line Analytical Methods for the Analysis of Bioactive Compounds. 2006.

47. **Maria Kulp**. Capillary Electrophoretic Monitoring of Biochemical Reaction Kinetics. 2006.

48. **Anu Aaspõllu.** Proteinases from *Vipera lebetina* Snake Venom Affecting Hemostasis. 2006.

49. **Lyudmila Chekulayeva**. Photosensitized Inactivation of Tumor Cells by Porphyrins and Chlorins. 2006.

50. **Merle Uudsemaa**. Quantum-Chemical Modeling of Solvated First Row Transition Metal Ions. 2006.

51. **Tagli Pitsi**. Nutrition Situation of Pre-School Children in Estonia from 1995 to 2004. 2006.

52. **Angela Ivask**. Luminescent Recombinant Sensor Bacteria for the Analysis of Bioavailable Heavy Metals. 2006.

53. **Tiina Lõugas**. Study on Physico-Chemical Properties and Some Bioactive Compounds of Sea Buckthorn (*Hippophae rhamnoides* L.). 2006.

54. **Kaja Kasemets**. Effect of Changing Environmental Conditions on the Fermentative Growth of *Saccharomyces cerevisae* S288C: Auxo-accelerostat Study. 2006.

55. **Ildar Nisamedtinov**. Application of ^{13}C and Fluorescence Labeling in Metabolic Studies of *Saccharomyces* spp. 2006.

56. **Alar Leibak**. On Additive Generalisation of Voronoï's Theory of Perfect Forms over Algebraic Number Fields. 2006.

57. **Andri Jagomägi**. Photoluminescence of Chalcopyrite Tellurides. 2006.

58. **Tõnu Martma**. Application of Carbon Isotopes to the Study of the Ordovician and Silurian of the Baltic. 2006.

59. **Marit Kauk**. Chemical Composition of $CuInSe_2$ Monograin Powders for Solar Cell Application. 2006.

60. **Julia Kois**. Electrochemical Deposition of $CuInSe_2$ Thin Films for Photovoltaic Applications. 2006.

61. **Ilona Oja Açik**. Sol-Gel Deposition of Titanium Dioxide Films. 2007.

62. **Tiia Anmann**. Integrated and Organized Cellular Bioenergetic Systems in Heart and Brain. 2007.

63. **Katrin Trummal**. Purification, Characterization and Specificity Studies of Metalloproteinases from *Vipera lebetina* Snake Venom. 2007.

64. **Gennadi Lessin**. Biochemical Definition of Coastal Zone Using Numerical Modeling and Measurement Data. 2007.

65. **Enno Pais**. Inverse problems to determine non-homogeneous degenerate memory kernels in heat flow. 2007.

66. **Maria Borissova**. Capillary Electrophoresis on Alkylimidazolium Salts. 2007.

67. **Karin Valmsen**. Prostaglandin Synthesis in the Coral *Plexaura homomalla*: Control of Prostaglandin Stereochemistry at Carbon 15 by Cyclooxygenases. 2007.

68. **Kristjan Piirimäe**. Long-Term Changes of Nutrient Fluxes in the Drainage Basin of the Gulf of Finland – Application of the PolFlow Model. 2007.

69. **Tatjana Dedova**. Chemical Spray Pyrolysis Deposition of Zinc Sulfide Thin Films and Zinc Oxide Nanostructured Layers. 2007.

70. **Katrin Tomson**. Production of Labelled Recombinant Proteins in Fed-Batch Systems in *Escherichia coli*. 2007.

71. **Cecilia Sarmiento**. Suppressors of RNA Silencing in Plants. 2008.

72. **Vilja Mardla**. Inhibition of Platelet Aggregation with Combination of Antiplatelet Agents. 2008.

73. **Maie Bachmann**. Effect of Modulated Microwave Radiation on Human Resting Electroencephalographic Signal. 2008.

74. **Dan Hüvonen**. Terahertz Spectroscopy of Low-Dimensional Spin Systems. 2008.

75. **Ly Villo**. Stereoselective Chemoenzymatic Synthesis of Deoxy Sugar Esters Involving *Candida antarctica* Lipase B. 2008.

76. **Johan Anton**. Technology of Integrated Photoelasticity for Residual Stress Measurement in Glass Articles of Axisymmetric Shape. 2008.

77. **Olga Volobujeva**. SEM Study of Selenization of Different Thin Metallic Films. 2008.

78. **Artur Jõgi**. Synthesis of 4'-Substituted 2,3'-dideoxynucleoside Analogues. 2008.

79. **Mario Kadastik**. Doubly Charged Higgs Boson Decays and Implications on Neutrino Physics. 2008.

80. **Fernando Pérez-Caballero**. Carbon Aerogels from 5-Methylresorcinol-Formaldehyde Gels. 2008.

81. **Sirje Vaask**. The Comparability, Reproducibility and Validity of Estonian Food consumption surveys. 2008.
82. **Anna Menaker**. Electrosynthesized Conducting Polymers, Polypyrrole And Poly(3,4-ethylenedioxythiophene), For Molecular Imprinting. 2009.
83. **Lauri Ilison.** Solitons And Solitary Waves In Hierarchical Korteweg-de Vries Type Systems. 2009.
84. **Kaia Ernits**. Study Of In_2S_3 And Zns Thin Films Deposited By Ultrasonic Spray Pyrolysis And Chemical Deposition. 2009.
85. **Veljo Sinivee.** Portable Spectrometer For Ionizing Radiation "gammamapper". 2009.
86. **Jüri Virkepu**. On Lagrange Formalism For Lie Theory And Operadic Harmonic Oscillator In Low Dimensions. 2009.
87. **Marko Piirsoo**. Deciphering Molecular Basis Of Schwann Cell Development. 2009.
88. **Kati Helmja**. Determination Of Phenolic Compounds And Their Antioxidative Capability In Plant Extracts. 2010.
89. **Merike Sõmera.** Sobemoviruses: Genomic Organization, Potential For Recombination And Necessity Of P1 In Systemic Infection. 2010.
90. **Kristjan Laes**. Preparation And Impedance Spectroscopy Of Hybrid Structures Based On $CuIn_3Se_5$ Photoabsorber. 2010.
91. **Kristin Lippur.** Asymmetric Synthesis Of 2,2'-bimorpholine And Its 5,5'-substituted Derivatives. 2010.
92. **Merike Luman.** Dialysis Dose And Nutrition Assessment By An Optical Method. 2010.
93. **Mihhail Berezovski**. Numerical Simulation Of Wave Propagation In Heterogeneous And Microstructured Materials. 2010.
94. **Tamara Aid-pavlidis.** Structure And Regulation Of Bdnf Gene. 2010.
95. **Olga Bragina.** The Role Of Sonic Hedgehog Pathway In Neuro- And Tumorigenesis. 2010.
96. **Merle Randrüüt.** Wave Propagation In Microstructured Solids: Solitary And Periodic Waves. 2010.
97. **Marju Laars.** Asymmetric Organocatalytic Michael And Aldol Reactions Mediated By Cyclic Amines. 2010.
98. **Maarja Grossberg.** Optical Properties Of Multinary Semiconductor Compounds For Photovoltaic Applications. 2010.
99. **Alla Maloverjan.** Vertebrate Homologues Of Drosophila Fused Kinase And Their Role In Sonic Hedgehog Signalling Pathway. 2010.
100. **Priit Pruunsild.** Neuronal Activity-dependent Transcription Factors And Regulation Of Human *Bdnf* Gene. 2010.

101. **Tatjana Knjazeva.** New Approaches In Capillary Electrophoresis For Separation And Study Of Proteins. 2011.
102. **Atanas Katerski.** Chemical Composition Of Sprayed Copper Indium Disulfide Films For Nanostructured Solar Cells. 2011.
103. **Kristi Timmo.** Formation of Properties of $CuInSe_2$ and $Cu_2ZnSn(S,Se)_4$ Monograin Powders Synthesized in Molten KI. 2011.
104. **Kert Tamm**. Wave Propagation and Interaction in Mindlin-Type Microstructured Solids: Numerical Simulation. 2011.
105. **Adrian Popp**. Ordovician Proetid Trilobites in Baltoscandia and Germany. 2011.
106. **Ove Pärn**. Sea Ice Deformation Events in the Gulf of Finland and This Impact on Shipping. 2011.
107. **Germo Väli**. Numerical Experiments on Matter Transport in the Baltic Sea. 2011.
108. **Andrus Seiman**. Point-of-Care Analyser Based on Capillary Electrophoresis. 2011.
109. **Olga Katargina**. Tick-Borne Pathogens Circulating in Estonia (Tick-Borne Encephalitis Virus, *Anaplasma phagocytophilum*, *Babesia* Species): Their Prevalence and Genetic Characterization. 2011.
110. **Ingrid Sumeri**. The Study of Probiotic Bacteria in Human Gastrointestinal Tract Simulator. 2011.
111. **Kairit Zovo**. Functional Characterization of Cellular Copper Proteome. 2011.
112. **Natalja Makarytsheva**. Analysis of Organic Species in Sediments and Soil by High Performance Separation Methods. 2011.
113. **Monika Mortimer**. Evaluation of the Biological Effects of Engineered Nanoparticles on Unicellular Pro- and Eukaryotic Organisms. 2011.
114. **Kersti Tepp**. Molecular System Bioenergetics of Cardiac Cells: Quantitative Analysis of Structure-Function Relationship. 2011.
115. **Anna-Liisa Peikolainen**. Organic Aerogels Based on 5-Methylresorcinol. 2011.
116. **Leeli Amon**. Palaeoecological Reconstruction of Late-Glacial Vegetation Dynamics in Eastern Baltic Area: A View Based on Plant Macrofossil Analysis. 2011.
117. **Tanel Peets**. Dispersion Analysis of Wave Motion in Microstructured Solids. 2011.
118. **Liina Kaupmees**. Selenization of Molybdenum as Contact Material in Solar Cells. 2011.