

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

Allan Olsper¹, Lauri Peil², Eugénie Hébrard³, Denis Fargette³ and Erkki Truve¹

Correspondence
Erkki Truve
erkki.truve@ttu.ee

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

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

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

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.

Received 16 August 2010
Accepted 8 November 2010

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 NSELY^{RNA}PDQSSGPARELDAET^{PHOS}YTERLEQQIAGFTEYNISGITV
 KTSREWTTAEALRVARYKPLGGGKAWGDS^{PHOS}DDEDTQE^{396/78}

327/1 S^{RNA}PFEIYGKRFREANS^{PHOS}EEYDESLRHGVLVYAEYDFSGDITIRASS^{PHOS}
 NTWWRERTKYHAEERRKSGQLSWADRFGDDS^{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.

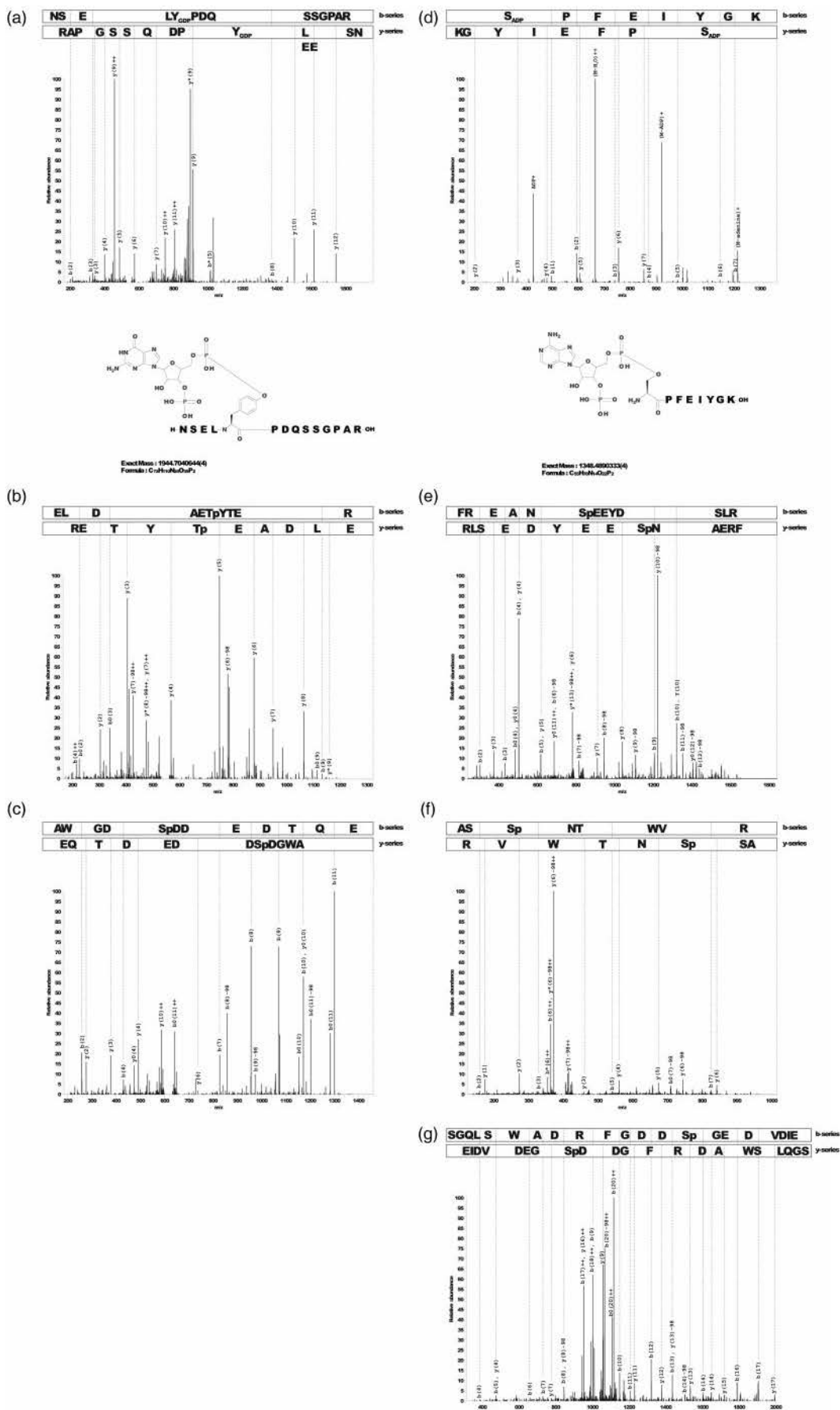


Fig. 2. Identification of PTMs of CfMV and RYMV VPg by MS/MS analysis. Co-purified VPg linked to RNA was trypsin-digested and RNA was degraded with acidic hydrolysis. The peptides were analysed by nano-LC/MS/MS and resulting data were searched against corresponding sequence databases by MASCOT. The b 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)DDEDTOE]. (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 [SGQLSWADRFQDD(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, $P=4 \times 10^{-3}$; site 14, $P=2 \times 10^{-5}$; site 33, $P=3 \times 10^{-2}$ and site 41, $P=9 \times 10^{-7}$. 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 E³⁹⁶/T³⁹⁷ 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-Bréchet *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 × 10⁶ 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., Ndjondjop, 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). Nucleotidylation 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 *rymv1-2*-mediated resistance. *Arch Virol* **153**, 1161–1164.
- Hébrard, E., Bessin, Y., Michon, T., Longhi, S., Uversky, V. N., Delalande, F., Van Dorsseleer, 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-Bréchet, 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., Dulleman, 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.