A peer-reviewed version of this preprint was published in PeerJ on 28 June 2016.

Olspert A, Hosmillo M, Chaudhry Y, Peil L, Truve E, Goodfellow I. 2016. Protein-RNA linkage and posttranslational modifications of feline calicivirus and murine norovirus VPg proteins. PeerJ 4:e2134 https://doi.org/10.7717/peerj.2134

Protein-RNA linkage and posttranslational modifications of feline calicivirus and murine norovirus VPg proteins

Authors:

Allan Olspert[1], Myra Hosmillo[2], Yasmin Chaudhry[2], Lauri Peil[3], Erkki Truve[1] and Ian Goodfellow[2*]

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

2) Division of Virology, Department of Virology, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge, UK

3) Faculty of Science and Technology, Institute of Technology, University of Tartu, Nooruse 1, 50411 Tartu, Estonia

***Corresponding author**

Ian Goodfellow

Division of Virology, Department of Virology, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge, UK

Email: IG299@cam.ac.uk

Abstract (max 500 words):302

Members of the *Caliciviridae* family of positive sense RNA viruses cause a wide range of diseases in both humans and animals. The detailed characterization of the calicivirus life cycle had been hampered due to the lack of robust cell culture systems and experimental tools for many of the members of the family. However a number of caliciviruses replicate efficiently in cell culture and have robust reverse genetics systems available, most notable feline calicivirus (FCV) and murine norovirus (MNV). These are therefore widely used as representative members with which to examine the mechanistic details of calicivirus genome translation and replication. The replication of the calicivirus RNA genome occurs via a double stranded RNA intermediate in the cytoplasm of the infected cell which is then used as a template for the production of new positive sense viral RNA, which is covalently linked to the virus-encoded protein VPg. The covalent linkage to VPg occurs during genome replication via the nucleotidylylation activity of the viral RNA-dependent RNA polymerase. Using FCV and MNV, we used mass spectrometry-based approach to identify the specific amino acid linked to the 5' end of the viral nucleic acid. We observed that both VPg proteins are covalently linked to GDP moieties via tyrosine positions 24 and 26 for FCV and MNV respectively. These data fit with previous observations indicating that mutations introduced into these specific amino acids are deleterious for viral replication and fail to produce infectious virus. In addition, we also detected serine phosphorylation sites within the FCV VPg protein with positions 80 and 107 found consistently phosphorylated on VPg-linked viral RNA isolated from infected cells. This work provides the first direct experimental characterisation of the linkage of infectious calicivirus viral RNA to the VPg protein and highlights that post-translational modifications of VPg may also occur during the viral life cycle.

Introduction

The RNA genomes of several positive sense RNA viruses are covalently linked to virus-encoded protein, referred to as VPg. Vertebrate RNA viruses that encode a VPg protein include picornaviruses, astroviruses and caliciviruses (Goodfellow, 2011). In picornaviruses, VPg is a 22 amino acid peptide that is linked to the 5' end of the RNA via a phosphodiester bond between the hydroxyl group of a tyrosine residue and the 5' phosphate group of the viral genomic RNA, which invariably starts with a pUpU sequence (Ambros & Baltimore, 1978; Rothberg et al., 1978). The linkage of VPg to picornavirus RNA occurs during viral genome replication in a process whereby the viral RNA-dependent RNA polymerase uses an RNA structure as a template for the addition of a pUpU moiety to a highly conserved tyrosine residue within the VPg peptide (Hewlett & Florkiewicz, 1980; Goodfellow et al., 2000; Paul et al., 2000). The VPg protein of caliciviruses and astroviruses is typically 13-15 KDa in size and is essential for the infectivity of viral RNA (Goodfellow et al., 2005; Chaudhry et al., 2006; Fuentes et al., 2012; Hosmillo et al., 2014). In the case of FCV and MNV, VPg plays an essential role in viral protein synthesis that has been linked to the ability of VPg to interact with cellular translation initiation factors, eIF4E in the case of FCV and eIF4G in the case of MNV (Goodfellow et al., 2005; Chaudhry et al., 2006; Chung et al., 2014). In plants, members of *Secoviridae, Potyviridae, Luteoviridae* families and *Sobemovirus* genus possess VPg proteins of 2-24 kDa in size, covalently linked to the 5' terminal uridine residue of the viral RNA via tyrosine or serine residue of the VPg (Jiang & Laliberte, 2011). Whilst VPg of plant viruses play multiple roles in the viral life cycle, the best characterized potyvirus VPg interacts with the canonical factor eIF4F also confirming a role in virus translation and the regulation of host gene expression (Jiang & Laliberte, 2011).

The covalent linkage of VPg to the 5' end of calicivirus RNAs has previously been examined

by iodination of purified viral RNA, confirming that VPg is linked to both genomic and subgenomic viral RNAs (Herbert, Brierley & Brown, 1997). Reverse genetics has also been used to identify the key amino acids in VPg required for viral infectivity of both FCV and MNV (Mitra, Sosnovtsev & Green, 2004; Subba-Reddy, Goodfellow & Kao, 2011; Leen et al., 2013); however, the multifunctional role of VPg has complicated the direct identification of amino acids involved in linkage to viral RNA as well as the precise nature of the nucleotide that is linked. Therefore direct experimental confirmation of the amino acid-RNA linkage to infectious calicivirus RNA is still lacking.

Here we have used mass-spectrometry based characterization to identify the amino acids involved in VPg-linkage to viral RNA in both FCV and MNV, demonstrating a direct linkage to pGp moieties. In addition, we identified the possible posttranslational modifications that may contribute to the regulation of VPg function during the calicivirus life cycle.

Materials and Methods

Virus culture and isolation of viral VPg-linked RNA

The F9 strain of FCV and the CW1 isolate of MNV were grown in Crandell-Reese feline kidney cells and RAW264.7 cells respectively. For each preparation at least five 170 cm^2 flasks were infected with a multiplicity of infection of 0.2 TCID50/cell. Infected cells were harvested at ~15 hours post infection. Cells were resuspended directly in lysis buffer and the total RNA was isolated using the GenElute mammalian total RNA miniprep kit as per the manufacturer's instructions. Eluted samples were combined, further concentrated by ethanol precipitation and resuspended in nuclease free water. Where required, preparations of VPg-linked RNA were treated with RNase cocktail (Ambion) at 37°C for 1 hour prior to the addition of SDS-PAGE sample buffer and separation by 15% SDS-PAGE.

95

96 **Recombinant protein expression and purification**

97 Untagged derivatives of FCV and MNV VPg proteins were expressed and purified in *E. coli*
98 as previously described (Goodfellow et al., 2005; Chaudhry et al., 2006).

99

100 **Mass-spectrometric analysis of FCV and MNV VPg-linked RNA**

101 VPg, covalently bound to the RNA, was trypsin digested and the RNA subsequently
102 hydrolyzed in 10% trifluoroacetic acid for 48 h at room temperature. 2-10 µg of total RNA
103 was used per analysis. The samples were then dried under vacuum, purified with StageTips
104 (Rappsilber, Mann & Ishihama Y, 2007) and analyzed by LC–MS/MS using an Agilent 1200
105 series nanoflow system (Agilent Technologies) connected to a LTQ Orbitrap mass-
106 spectrometer (Thermo Electron) equipped with a nanoelectrospray ion source (Proxeon), as
107 described previously (Olspert et al., 2011a). LTQ Orbitrap was operated in the data dependent
108 mode with a full scan in the Orbitrap (mass range m/z 300–1900, resolution 60 000 at m/z
109 400, target value 1 Â 106 ions) followed by up to five MS/MS scans in the LTQ part of the
110 instrument (normalized collision energy 35%, wideband activation enabled, target value 5000
111 ions). Fragment MS/MS spectra from raw files were extracted as MSM files and then merged
112 to peak lists using Raw2MSM version 1.11, selecting top eight peaks for each 100 Da (Olsen
113 et al., 2005). MSM files were searched with the Mascot 2.3 search engine (Matrix Science)
114 against the protein sequence database composed of VPg sequences and common contaminant
115 proteins such as trypsin, keratins etc. Search parameters were as follows: 5 ppm precursor
116 mass tolerance and 0.6 Da MS/MS mass tolerance, three missed trypsin cleavages plus a
117 number of variable modifications such as oxidation (M), oxidation (HW), ethyl (DE),
118 phospho (ST), phospho (Y), pAp (STY), pGp (STY), pCp (STY) and pUp (STY). For both
119 viruses at least two independent biological samples were analyzed. For publication the spectra

were auto-annotated with xiSPEC (http://spectrumviewer.org) and images were prepared using Inkscape (http://www.inkscape.org).

RESULTS AND DISCUSSION

Conservation of the calicivirus 5' end and VPg sequences

We initially compared the VPg sequences from a number of representative caliciviruses to identify amino acids that were highly conserved across the genera (Figure 1A). Calicivirus VPg sequences vary in length from 65 amino acids for bovine nebovirus to 138 amino acids for Norwalk virus. The recent solution of the structures of the VPg proteins from MNV, FCV, and porcine sapovirus (PSaV) highlight the presence of conserved helical bundles at the core of VPg, tightly bound in hydrophobic interactions (Leen et al., 2013; Hwang et al., 2015). A limited number of amino acids were highly conserved across all calicivirus genera, most notably a lysine rich N-terminal region and a central motif containing the EYDEΦ sequence, with Φ representing any aromatic amino acid (Figure 1A). The tyrosine within this motif, position 24 and 26 of FCV and MNV respectively, have previously been proposed as a possible site for VPg nucleotidylylation based on data using either *in vitro* biochemical assays (Machín, Martín Alonso & Parra, 2001; Belliot et al., 2008; Han et al., 2010) or a novel cell-based assay (Subba-Reddy, Goodfellow & Kao, 2011). However discordant data was obtained for MNV where *in vitro* biochemical assays identified tyrosine at position 117 as the site for nucleotidylylation (Han et al., 2010). Importantly, mutational analysis of this amino acid in the context of the MNV infectious clone confirmed that Y117 was not required for viral infectivity whereas Y24 was essential (Subba-Reddy, Goodfellow & Kao, 2011). This highlights that, at least for MNV, *in vitro* biochemical approaches using recombinant purified proteins, are confounded by an apparent lack of specificity of the viral RdRp.

Alignment of the 5' ends sequences of representative calicivirus genomes has demonstrated that almost all genomic and subgenomic RNAs start with a GU dinucleotide (Figure 1B and 1C). Tulane virus is an exception to this, as published data would indicate that the genome begins with a GGG sequence (Farkas et al., 2008). It is worth noting that there is only a single full-length genome sequence available for Tulane virus, therefore confirmation of the 5' end may require additional viral sequences. Therefore, based entirely on sequence conservation we would expect calicivirus VPg proteins to be guanylylated on the conserved tyrosine within the EYDEΦ sequence equivalent to positions 24 and 26 for FCV and MNV respectively. These amino acids are predicted to lie within the structured region of the FCV and MNV proteins as highlighted in Figure 1D and E respectively.

Purification of viral VPg-linked RNA

In order to identify the amino acid and the nucleotide involved in the covalent linkage of calicivirus RNA to VPg, a source of viral VPg-linked RNA that would yield sufficient nucleotide-linked VPg was required. Attempts were initially made to purify sufficient quantities of VPg-linked viral RNA from infectious virions, however the yields were insufficient to allow the robust detection of VPg peptides by mass spectrometry (data not shown). As an alternative approach, we isolated total RNA from infected cells as we have previously demonstrated that RNA isolated in this way is covalently linked to the mature form of VPg only, that the RNA is infectious when transfected into permissive cells, that the VPg-linked RNA is translationally competent and that the linkage to VPg is essential for the infectivity of the viral RNA (Goodfellow et al., 2005; Chaudhry et al., 2006; Hosmillo et al., 2014). Taken together, these observations confirm that RNA isolated from infected cells

provides a robust source of authentic viral RNA. Cells permissive for either FCV or MNV infection were infected with a high multiplicity of infection and total RNA purified using a silica column based purification method. To determine if sufficient quantities of VPg-linked RNA were present within these preparations, the RNA was digested with a cocktail of ribonuclease A and T1, separated by SDS-PAGE and proteins visualised by staining with colloidal Coomassie blue (Figure 2). Recombinant VPg proteins expressed and purified from *E. coli* without any exogenous non-viral amino acids were used to confirm the expected mass of the VPg proteins. FCV VPg was readily visible following Coomassie staining of RNase digested RNA, however MNV VPg, because of its larger mass, was obscured by the RNase present within the sample. Western blot analysis (data not shown), and subsequent mass spectrometry (see below) was used to confirm the isolation of the MNV VPg protein.

Detection of FCV and MNV VPg peptides and post-translational modifications

Initial attempts were made to analyse the RNase treated and trypsin-digested VPg-linked RNA preparations as a method to identify the nucleotide and amino acids involved in the covalent linkage, however this approach failed to produce spectra that allowed for the detection of the nucleotide-linked VPg peptides. As an alternative approach, we used tryptic digestion of VPg-linked RNA preparations followed by acid hydrolysis of the RNA-linked peptides as described previously (Olspert et al., 2011a,b). The RNA-linked amino-acid residue modification after RNA hydrolysis using this method is known to be a 5',3'-diphosphate nucleotide, pNp (N denoting adenosine, cytidine, guanosine or uridine), and the possible phosphodiester bond acceptor residues are serine, tyrosine and threonine. FCV and MNV VPg-linked RNA preparations were prepared and analysed by Orbitrap MS. The sequence coverage obtained for the FCV and MNV VPg proteins were 70% and 63%, respectively. The identified peptides are shown in Table 1, Figure 3 and Figure 4. The

regions not detected were most likely absent due to trypsin digestion producing peptides too short for detection.

Using this approach we determined that the FCV VPg is linked to RNA through the tyrosine residue at position 24 (Y24) and the corresponding modification was pGp, as assigned by the modification delta mass and the corresponding fragmentation spectrum (Figure 3B). This is in agreement with the high degree of conservation of a 5' G nucleotide in all calicivirus genomes (Figure 1B and 1C). The spectra were searched against all possible nucleotides (pGp, pUp, pCp and pAp) but no other matches were detected indicating that all the detected VPg peptides were derived from linkage to the positive strand of viral RNA. The corresponding FCV VPg peptide was never detected without pGp modification. In FCV VPg we also identified two potential phosphorylation sites; threonine at position 80 (Figure 3C) and serine at position 107 (Figure 3D) were consistently detected as phosphorylated but the same peptides were also detected without the modification. This might indicate loss of modification during sample handling and/or transient nature of the modification. Threonine at position 11 was also detected in a phosphorylated form (Table 1), however this was identified in only one FCV sample. Unfortunately, due to peptide's short length and low amount of matched fragmentation ions, this is a low confidence observation. Threonine at position 11 is not conserved between FCV isolates, with all other isolated possessing a proline at this position (data not shown), while T80 and S107 are 100% conserved across all FCV isolates. Amino acids T80 and S107 are in disordered region of the FCV VPg protein (Figure 1D).

The MNV VPg protein was identified as also linked to RNA through a tyrosine residue at position 26 and the modification was also pGp (Figure. 4B). Surprisingly for MNV the corresponding peptide was also detected without the RNA modification suggesting that RNA

modification was lost during sample preparation. In addition we detected random aspartate and glutamate ethylation, methionine and tryptophan oxidations (Table 1 and data not shown), which are known to be generated in vitro during sample preparation (Stadtman & Levine, 2003; Xing et al., 2008; Olspert et al., 2011a) and were therefore not considered of biological relevance.

Previous studies on caliciviruses VPg nucleotidylylation have examined the ability of the viral RdRp to uridylylate VPg by leading to the formation of a VPg-pUpU(OH) moiety (Rohayem et al., 2006; Belliot et al., 2008). As all calicivirus genomic and subgenomic RNAs begin with a single G, uridylylated VPg, if formed, would be expected to prime only negative sense RNA synthesis on the 3' polyA tail present on the viral genomic RNA. The end result would be VPg-pUpU at the 5' end of negative sense viral RNA. Whether VPg is uridylylated during calicivirus replication remains to be determined, however our data would indicate that only GDP was found linked to either the FCV or the MNV VPg proteins isolated from infected cells. Given the asymmetry of calicivirus genome replication, the number of negative sense RNA molecules present within an infected cell can be >1000 fold lower than the corresponding positive sense RNA molecule (Vashist, Urena & Goodfellow, 2012), which even if it were linked to uridylylated VPg, may be present in such low quantities to make detection challenging. We have previously demonstrated that the norovirus RdRp possesses the ability to initiate RNA synthesis *de novo* and that this activity is regulated by binding to the viral capsid protein VP1 (Subba-Reddy, Goodfellow & Kao, 2011). This has allowed us to propose a model whereby negative sense RNA synthesis occurs via a primer independent *de novo* mechanism of RNA synthesis but positive sense RNA synthesis occurs via a VPg-primed mechanism [Reviewed in (Thorne & Goodfellow, 2014)]. In this model, only guanylylated VPg would be produced, fitting with our experimental observations.

In picornaviruses, the VPg protein is removed immediately upon viral RNA release into the cytoplasm by the host cell enzyme TDP2 (Virgen-Slane et al., 2012), but the removal of the VPg protein is not essential for the replication of the incoming viral RNA (Langereis et al., 2014). In contrast, ongoing picornavirus replication may require the activity of TDP2 (Maciejewski et al., 2016) Given the absolute requirement of VPg for calicivirus genome translation we would expect that the calicivirus VPg protein remains attached to the viral RNA, at least during the initial stages of the viral life cycle. Fitting with this hypothesis, preliminary *in vitro* analysis would indicate that purified TDP2 is unable to cleave the MNV VPg protein from viral RNA under conditions that readily removes the poliovirus VPg (Bert Semler and Sonia Maciejewski, pers. comm., 2016).

In conclusion, this work has provided experimental confirmation that at least for FCV and MNV, the covalent linkage of the VPg proteins to the 5' end of the viral genome occurs specifically via a highly conserved tyrosine residue to the 5' G nucleotide. The identification of potential phosphorylated sites in the FCV VPg protein may provide a mechanism by which the function of VPg is temporally regulated during the viral life cycle.

Reference:

Ambros V., Baltimore D. 1978. Protein is linked to the 5' end of Poliovirus RNA by a phosphodiester linkage to tyrosine. *Journal of Biological Chemistry* 253:5263–5266.

Belliot G., Sosnovtsev S V., Chang KO., McPhie P., Green KY. 2008. Nucleotidylylation of the VPg protein of a human norovirus by its proteinase-polymerase precursor protein. *Virology* 374:33–49. DOI: 10.1016/j.virol.2007.12.028.

Chaudhry Y., Nayak A., Bordeleau M-E., Tanaka J., Pelletier J., Belsham GJ., Roberts LO.,

Goodfellow IG. 2006. Caliciviruses differ in their functional requirements for eIF4F components. *Journal of Biological Chemistry* 281:25315–25325. DOI: 10.1074/jbc.M602230200.

Chung L., Bailey D., Leen EN., Emmott EP., Chaudhry Y., Roberts LO., Curry S., Locker N., Goodfellow IG. 2014. Norovirus translation requires an interaction between the C terminus of the genome-linked viral protein VPg and eukaryotic translation initiation factor 4G. *Journal of Biological Chemistry* 289:21738–21750. DOI: 10.1074/jbc.M114.550657.

Farkas T., Sestak K., Wei C., Jiang X. 2008. Characterization of a rhesus monkey calicivirus representing a new genus of Caliciviridae. *Journal of Virology* 82:5408–5416. DOI: 10.1128/JVI.00070-08.

Fuentes C., Bosch A., Pinto RM., Guix S. 2012. Identification of human astrovirus genome-linked protein (VPg) essential for virus infectivity. *Journal of Virology* 86:10070–10078. DOI: 10.1128/JVI.00797-12.

Goodfellow I., Chaudhry Y., Richardson A., Meredith J., Almond JW., Barclay W., Evans DJ. 2000. Identification of a cis-acting replication element within the poliovirus coding region. *Journal of Virology* 74:4590–600. DOI: 10.1128/JVI.74.10.4590-4600.2000.

Goodfellow I., Chaudhry Y., Gioldasi I., Gerondopoulos A., Natoni A., Labrie L., Laliberte JF., Roberts L. 2005. Calicivirus translation initiation requires an interaction between VPg and eIF4E. *EMBO Reports* 6:968–972. DOI: 10.1038/sj.embor.7400510.

Goodfellow I. 2011. The genome-linked protein VPg of vertebrate viruses - a multifaceted protein. *Current Opinion in Virology* 1:355–362. DOI: 10.1016/j.coviro.2011.09.003.

Han KR., Choi Y., Min BS., Jeong H., Cheon D., Kim J., Jee Y., Shin S., Yang JM. 2010. Murine norovirus-1 3Dpol exhibits RNA-dependent RNA polymerase activity and nucleotidylylates on Tyr of the VPg. *Journal of General Virology* 91:1713–1722. DOI:

10.1099/vir.0.020461-0.

Herbert TP., Brierley I., Brown TD. 1997. Identification of a protein linked to the genomic and subgenomic mRNAs of feline calicivirus and its role in translation. *Journal of General Virology* 78:1033–40.

Hewlett MJ., Florkiewicz RZ. 1980. Sequence of picornavirus RNAs containing a radioiodinated 5 ' -linked peptide reveals a conserved 5 ' sequence. *Proceedings of the National Academy of Sciences* 77:303–307.

Hosmillo M., Chaudhry Y., Kim DS., Goodfellow I., Cho KO. 2014. Sapovirus translation requires an interaction between VPg and the cap binding protein eIF4E. *Journal of Virology* 88:12213–12221. DOI: 10.1128/JVI.01650-14.

Hwang H-J., Min HJ., Yun H., Pelton JG., Wemmer DE., Cho K-O., Kim J-S., Lee CW. 2015. Solution structure of the porcine sapovirus VPg core reveals a stable three-helical bundle with a conserved surface patch. *Biochemical and Biophysical Research Communications* 459:610–616. DOI: 10.1016/j.bbrc.2015.02.156.

Jiang J., Laliberte J-F. 2011. The genome-linked protein VPg of plant viruses - A protein with many partners. *Current Opinion in Virology* 1:347–354. DOI: 10.1016/j.coviro.2011.09.010.

Langereis MA., Feng Q., Nelissen FHT., Virgen-Slane R., Van der Heden Van Noort GJ., Maciejewski S., Filippov D V., Semler BL., Van Delft FL., Van Kuppeveld FJM. 2014. Modification of picornavirus genomic RNA using “click” chemistry shows that unlinking of the VPg peptide is dispensable for translation and replication of the incoming viral RNA. *Nucleic Acids Research* 42:2473–2482. DOI: 10.1093/nar/gkt1162.

Leen EN., Kwok KYR., Birtley JR., Simpson PJ., Subba-Reddy C V., Chaudhry Y., Sosnovtsev S V., Green KY., Prater SN., Tong M., Young JC., Chung LMW., Marchant J., Roberts LO., Kao CC., Matthews S., Goodfellow IG., Curry S. 2013. Structures of the

compact helical core domains of feline calicivirus and murine norovirus VPg proteins. *Journal of Virology* 87:5318–30. DOI: 10.1128/JVI.03151-12.

Machín a., Martín Alonso JM., Parra F. 2001. Identification of the amino acid residue involved in rabbit hemorrhagic disease virus VPg uridylylation. *Journal of Biological Chemistry* 276:27787–92. DOI: 10.1074/jbc.M100707200.

Maciejewski S., Nguyen JHC., Gómez-Herreros F., Cortés-Ledesma F., Caldecott KW., Semler BL. 2016. Divergent Requirement for a DNA Repair Enzyme during Enterovirus Infections. *mBio* 7:e01931–15. DOI: doi:10.1128/mBio.01931-15.

Mitra T., Sosnovtsev S V., Green KY. 2004. Mutagenesis of tyrosine 24 in the VPg protein is lethal for feline calicivirus. *Journal of Virology* 78:4931–4935. DOI: 10.1128/JVI.78.9.4931-4935.2004.

Olspert A., Arike L., Peil L., Truve E. 2011a. Sobemovirus RNA linked to VPg over a threonine residue. *FEBS Letters* 585:2979–2985. DOI: 10.1016/j.febslet.2011.08.009.

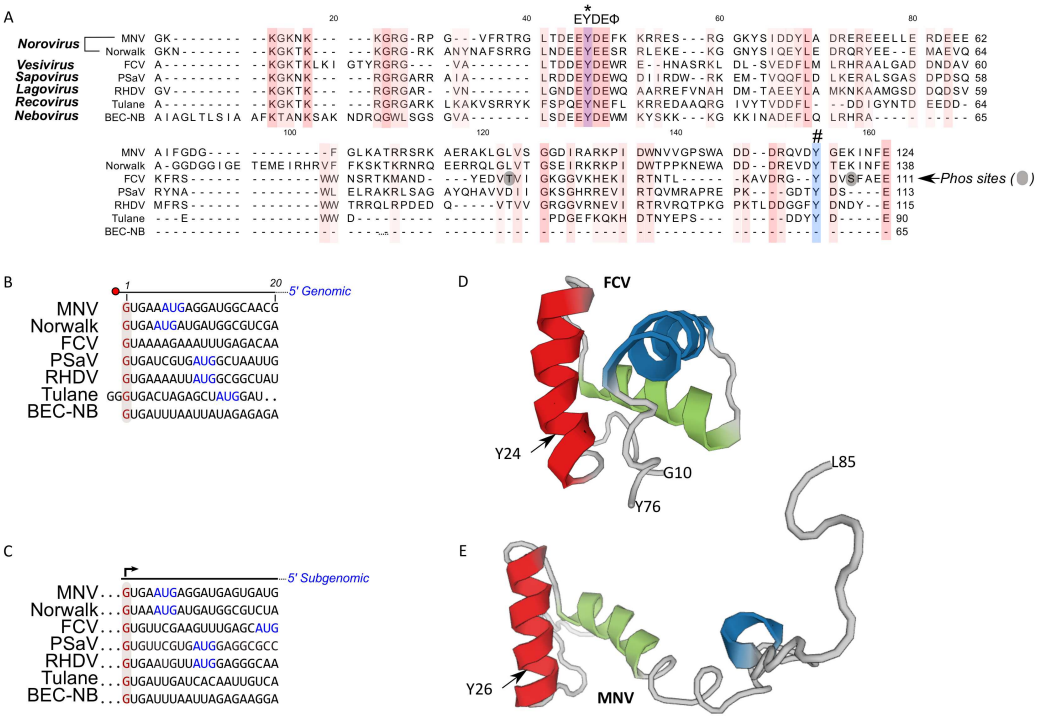
Olspert A., Peil L., Hébrard E., Fargette D., Truve E. 2011b. Protein-RNA linkage and post-translational modifications of two sobemovirus VPgs. *Journal of General Virology* 92:445–452. DOI: 10.1099/vir.0.026476-0.

Paul A V., Rieder E., Kim DW., van Boom JH., Wimmer E. 2000. Identification of an RNA hairpin in poliovirus RNA that serves as the primary template in the in vitro uridylylation of VPg. *Journal of Virology* 74:10359–70. DOI: 10.1128/JVI.74.22.10359-10370.2000.

Rohayem J., Robel I., Jäger K., Scheffler U., Rudolph W. 2006. Protein-primed and de novo initiation of RNA synthesis by norovirus 3Dpol. *Journal of Virology* 80:7060–7069. DOI: 10.1128/JVI.02195-05.

Rothberg PG., Harris TJ., Nomoto A., Wimmer E. 1978. O4-(5'-uridylyl)tyrosine is the bond between the genome-linked protein and the RNA of poliovirus. *Proceedings of the National Academy of Sciences* 75:4868–72.

345 Stadtman ER., Levine RL. 2003. Free radical-mediated oxidation of free amino acids and
346 amino acid residues in proteins. *Amino Acids* 25:207–218. DOI: 10.1007/s00726-003-
347 0011-2.
348 Subba-Reddy C V., Goodfellow I., Kao CC. 2011. VPg-primed RNA synthesis of norovirus
349 RNA-dependent RNA polymerases by using a novel cell-based assay. *Journal of*
350 *Virology* 85:13027–37. DOI: 10.1128/JVI.06191-11.
351 Thorne LG., Goodfellow IG. 2014. Norovirus gene expression and replication. *Journal of*
352 *General Virology* 95:278–91. DOI: 10.1099/vir.0.059634-0.
353 Vashist S., Urena L., Goodfellow I. 2012. Development of a strand specific real-time RT-
354 qPCR assay for the detection and quantitation of murine norovirus RNA. *Journal of*
355 *Virological Methods* 184:69–76. DOI: 10.1016/j.jviromet.2012.05.012.
356 Virgen-Slane R., Rozovics JM., Fitzgerald KD., Ngo T., Chou W., van der Heden van Noort
357 GJ., Filippov DV., Gershon PD., Semler BL. 2012. An RNA virus hijacks an incognito
358 function of a DNA repair enzyme. *Proceedings of the National Academy of Sciences*
359 109:14634–14639.
360 Xing G., Zhang J., Chen Y., Zhao Y. 2008. Identification of four novel types of in vitro
361 protein modifications. *Journal of Proteome Research* 7:4603–4608. DOI:
362 10.1021/pr800456q.
363



364

365 **Figure 1. Comparison of calicivirus VPg and 5' genomic and subgenomic sequences**. (A)
366 Amino acid alignment of VPg sequences among representatives of calicivirus genera: MNV
367 (DQ285629), Norwalk (AF093797), FCV (M86379), PSaV (AF182760), RHDV (Z49271),
368 Tulane (EU391643) and BEC-NB (AY082891). The conserved amino acids are coloured
369 including the highly conserved central motif of VPg, EYDEΦ (Φ is any aromatic acid). An
370 asterisk (*) indicates the conserved tyrosine (Y) residue essential for calicivirus replication. A
371 hash (#) indicates the Y residue identified necessary for MNV nucleotidylylation using an *in*
372 *vitro* biochemical approach (Han et al., 2010). The identified phosphorylation (Phos) sites in
373 the FCV VPg protein are shaded. Alignment of the first 20 nucleotides of the genomic (B) and
374 (C) subgenomic RNAs of representative caliciviruses. The putative VPg-linked 5' G
375 nucleotides are highlighted and shown in red. AUG are shown in blue. The solution structure
376 of the FCV (D, PDB:2M4H) and MNV (E, PDB: 2MG4) VPg proteins are also shown. The
377 FCV structure represents amino acid G10 to Y27 whereas the MNV VPg structure
378 encompasses amino acids G11 to L85.

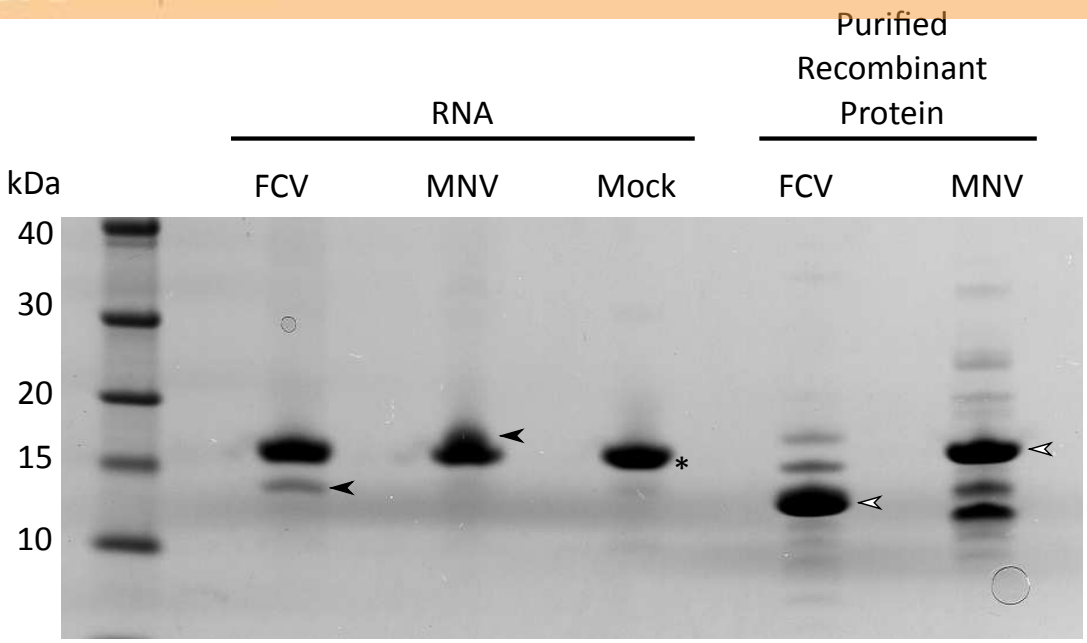
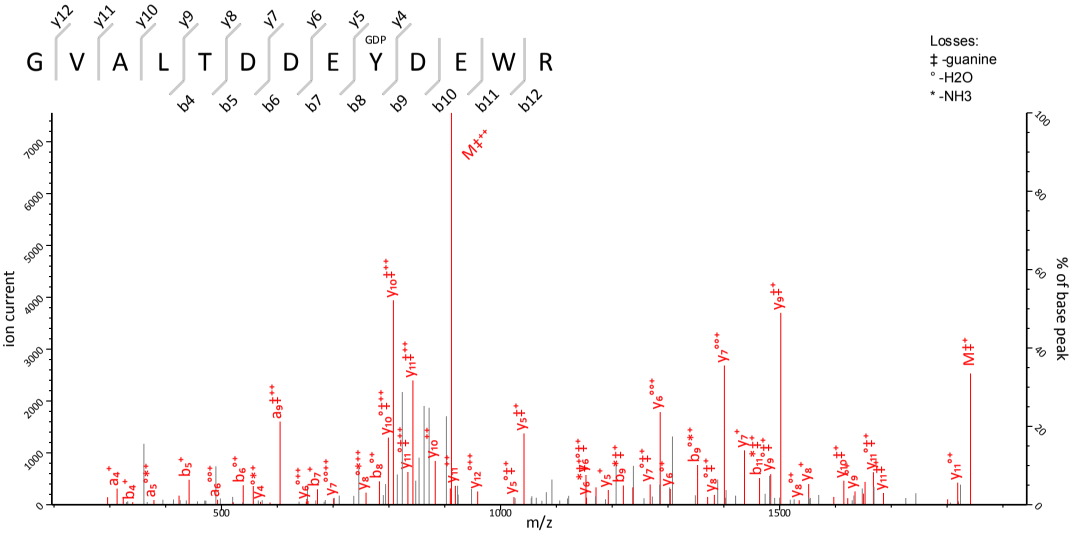


Figure 2. Isolation and characterization of calicivirus VPg-linked RNA. Total RNA was isolated from FCV, MNV or mock-infected cells then ~10µg was subjected to RNase treatment. The calicivirus VPg linked to the RNA were subsequently analysed in SDS-PAGE, alongside their corresponding recombinant proteins. White arrowheads indicate the recombinant VPg used as a marker with black arrowhead indicating the position of VPg linked to the RNA. An asterisk (*) is used to highlight the position of the RNase A in the treated samples.

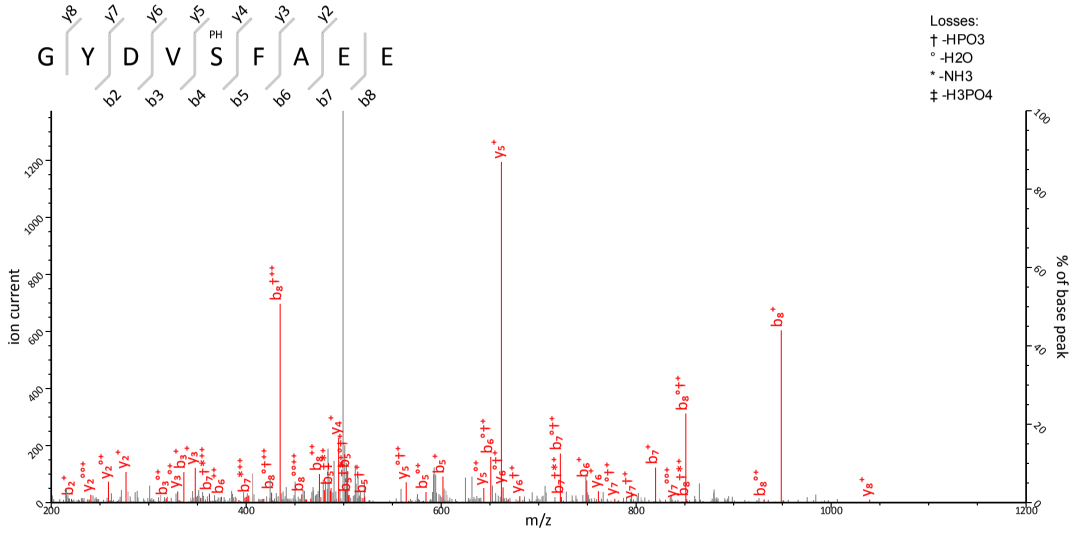
A

```
 1 akgktklkIG TYRgrGVALT DDEYDEWReh nasrKLDLSV EDFLMLRhrA ALGADDNDAV
61 KFRSWWNSRT KMANDYEDVT VIGKGGVKHE KIRTNTLKAV DRGYDVSFAE E
```

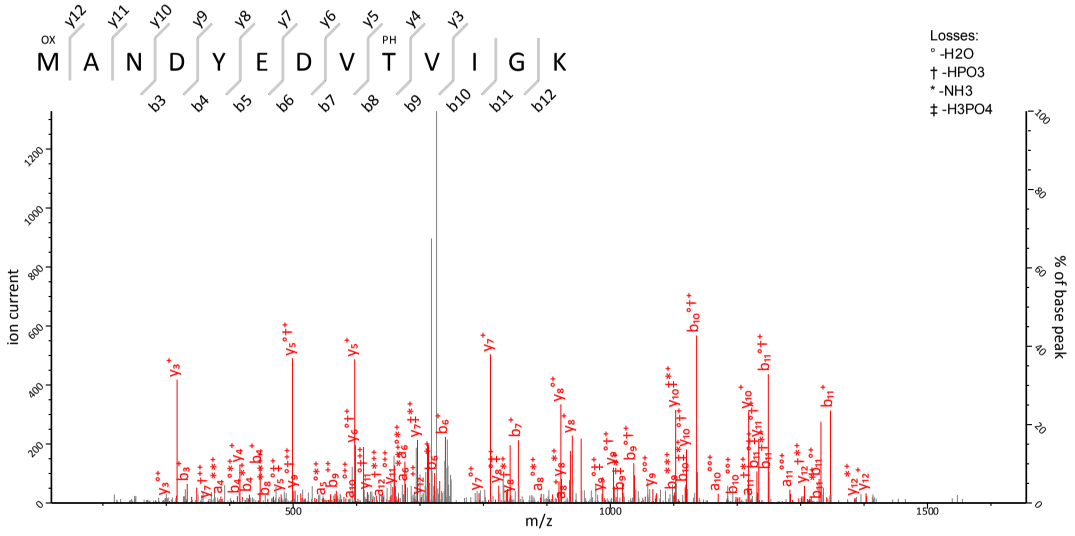
B



C



D



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.1960v1 | CC-BY 4.0 Open Access | rec: 14 Apr 2016, publ: 14 Apr 2016

Figure 3. Mass-spectrometric characterization of FCV VPg. (A) FCV VPg, regions for which peptides were detected are shown in red, the amino acid residue linked to RNA is in blue and underlined, phosphorylated residues are in bold green, the regions not detected are in lowercase black. (B–D) Identification of post-translational modifications of FCV 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 a/b and y ions represent peptide N- and C-terminal fragment ions produced by collision-induced dissociation in the mass spectrometer. Matched ions are indicated in red and corresponding losses indicated at the top left corner of each plot, M denotes the precursor peptide (with corresponding losses). (B) Identification of the residue covalently linked to RNA. The FCV VPg peptide, GVALTDDEYDEWR, was identified to contain a Y9 linked pGp modification (indicated with GDP), a corresponding degradation product of viral RNA. (C-D) VPg peptide GYDVSFAEE was detected to contain a phosphorylation at S5 (indicated with PH) and peptide MANDYEDVTVIGK was detected to be phosphorylated at T9 (in addition to occasional M1 oxydation [OX] occurring during sample handling).

A

```
  1 gkkgknkkgr GRPGVFRtrG LTDEEYDEFK KrresrggkY SIDDYLADRE REEELLERDE
 61 EEAIFGDGFG LKatrrsrka erakLGLVSG GDIRarkPID WNVVGPSWAD DDRqvdygek
121 infe
```

B

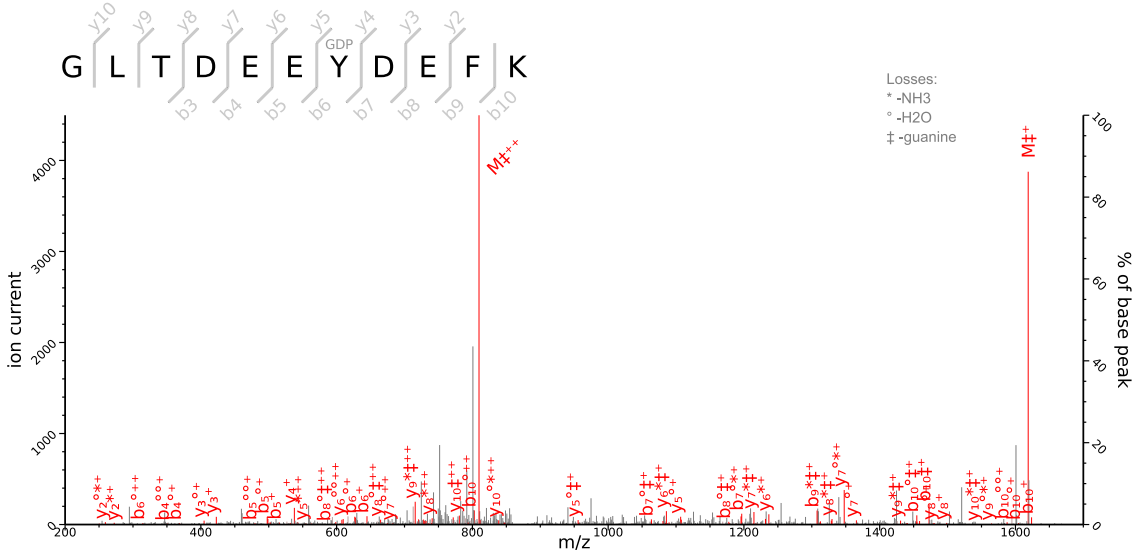


Figure 4. Mass-spectrometric characterization of MNV VPg. (A) MNV VPg, regions for which peptides were detected are shown in red, the amino acid residue linked to RNA is in blue and underlined, the regions not detected are in lowercase black. (B) Determination of the residue covalently linked to MNV RNA 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. Matched ions are indicated in red and corresponding losses indicated at the top left corner of each plot, M denotes the precursor peptide (with corresponding losses). The MNV VPg peptide, GLTDEEYDEFK, was identified to contain Y7 linked pGp modification (indicated with GDP), a corresponding degradation product of viral RNA.

435 Table 1. Examples of detected peptides identified by fragmentation spectra. The post-translational modifications are described (phos –
436 phosphorylation, ox – oxidation, eth – ethylation, GDP - pGp) and the modified position is in bold in the peptide.

Virus	Position	Peptide	Modification	Experimental mass, Da	Calculated mass, Da	Mass error, ppm	Mascot score
FCV	9-13	IG**T**YR	phos	688.2952	688.2945	1.05	20
FCV	9-13	IGTYR		608.329	608.3282	1.39	32
FCV	16-28	GVALTDDE**Y**DEWR	GDP	1992.6959	1992.6928	1.54	61
FCV	16-28	GVALTDDE**Y**DE**W**R	GDP, ox	2008.6909	2008.6877	1.56	52
FCV	16-28	GVALTDDE**Y**D**E**WR	GDP, eth	2020.7274	2020.7241	1.62	50
FCV	16-28	GVALTDDE**Y**DEWREHNASR	GDP	2687.0071	2687.0075	-0.16	10
FCV	35-47	KLDLSVEDFL**M**LR	ox	1593.8448	1593.8436	0.77	86
FCV	36-47	LDLSVEDFL**M**LR	ox	1465.7502	1465.7487	1.09	68
FCV	50-61	AALGAD**D**NDAVK	eth	1186.5837	1186.583	0.62	76
FCV	50-61	AALGADDNDAVK		1158.5542	1158.5517	2.23	63

FCV	64-69	SW**W**NSR	ox	850.3722	850.3722	0.066	29
FCV	64-69	S**WW**NSR	ox, ox	866.3674	866.3671	0.39	16
FCV	64-69	SWWNSR		834.3776	834.3773	0.45	15
FCV	72-84	**M**ANDYEDVTVIGK	ox	1469.672	1469.6708	0.84	96
FCV	72-84	**M**ANDYEDV**T**VIGK	ox, phos	1549.6372	1549.6371	0.076	65
FCV	72-84	MANDYEDVTVIGK		1453.6756	1453.6759	-0.17	111
FCV	103-111	GYDV**S**FAEE	phos	1095.3796	1095.3798	-0.12	31
FCV	103-111	GYDVSFAEE		1015.4134	1015.4135	-0.0039	42
MNV	9-17	GRPGVFR		787.4444	787.4453	-1.1	32
MNV	20-30	GLTDEE**Y**DEFK	GDP	1769.5878	1769.5859	1.05	36
MNV	20-30	GLTDEEYDEFK		1344.572	1344.5721	-0.13	64
MNV	20-31	GLTDEE**Y**DEFKK	GDP	1897.6834	1897.6808	1.36	29
MNV	20-31	GLTDEEYDEFKK		1472.6673	1472.6671	0.13	25

MNV	40-49	YSIDDYLA**D**R	eth	1257.5883	1257.5877	0.48	70
MNV	40-49	YSIDDYLADR		1229.5588	1229.5564	1.92	52
MNV	40-51	YSIDDYLADRER		1514.6998	1514.7001	-0.19	23
MNV	50-58	EEELLER		916.4508	916.4501	0.72	19
MNV	50-58	EREEELLER		1201.5946	1201.5938	0.64	63
MNV	50-72	ERE**E**ELLERDEEEAIFGDGFGLK	eth	2737.3105	2737.3082	0.82	11
MNV	50-72	EREEELLERDEEEAIFGDGFGLK		2709.2807	2709.2769	1.39	52
MNV	52-72	EEELLERDEEEAIFGDGFGLK		2424.1361	2424.1332	1.2	77
MNV	59-72	DEEEAIFGDGFGLK		1525.6944	1525.6936	0.48	103
MNV	85-94	LGLVSGG**D**IR	eth	1013.5865	1013.5869	-0.42	42
MNV	85-94	LGLVSGGDIR		985.5566	985.5556	0.92	55
MNV	97-113	KPIDWNVVGPSWADDDR		1968.9345	1968.933	0.75	33