

**Supplementary Data**

**VPg impact on ryegrass mottle virus serine-like 3C protease proteolysis and structure**

**Gints Kalnins<sup>1†</sup>, Rebeka Ludviga<sup>2†</sup>, Ieva Kalnciema<sup>3</sup>, Gunta Resevica<sup>3</sup>, Vilija Zeltina<sup>3</sup>, Janis Bogans<sup>4</sup>, Kaspars Tars<sup>1</sup>, Andris Zeltins<sup>3</sup>, Ina Balke<sup>2\*</sup>**

<sup>1</sup> Structural biology group, Latvian Biomedical Research and Study Centre, Ratsupites street 1, k-1, Riga LV-1067, Latvia

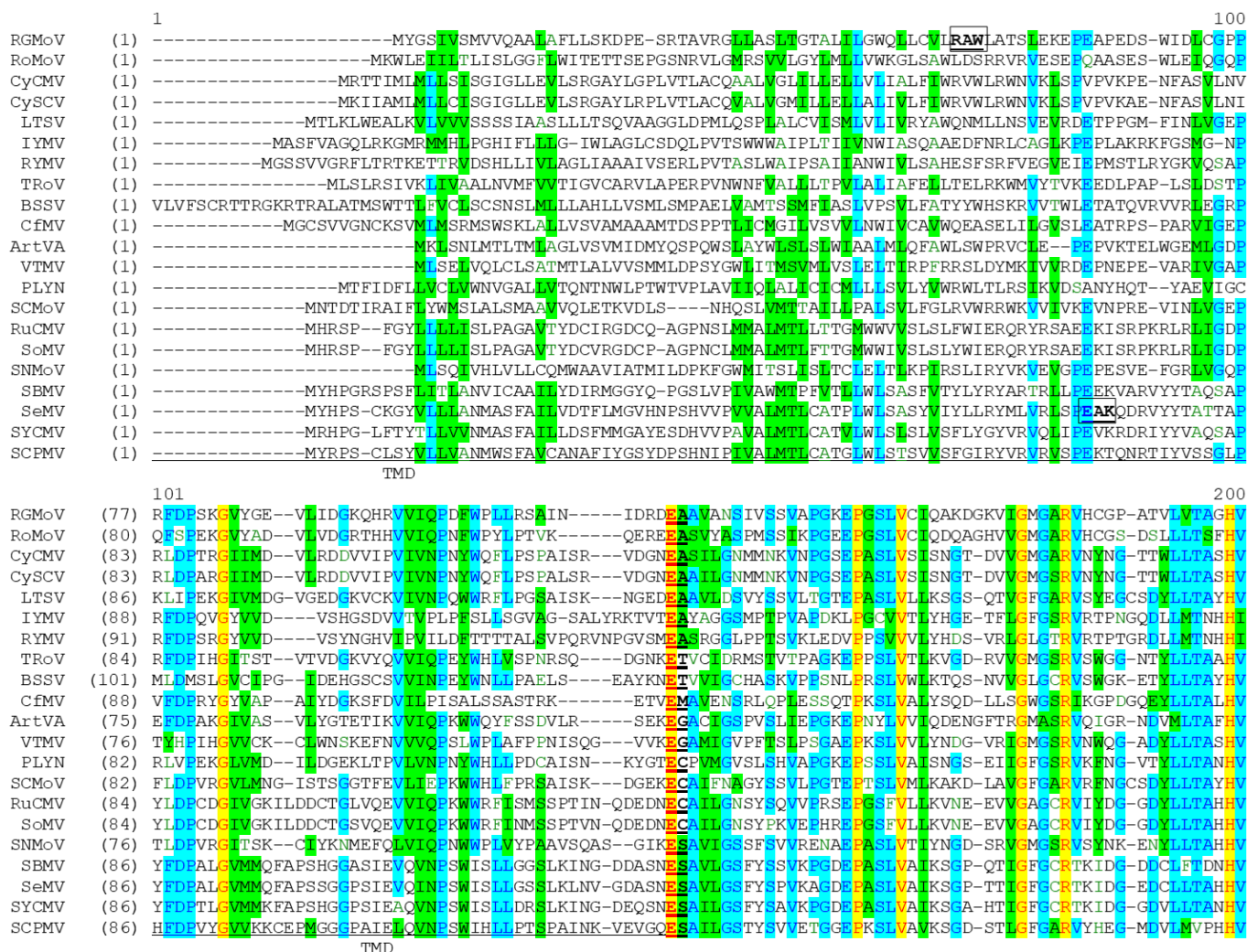
<sup>2</sup> Plant virus protein research group, Latvian Biomedical Research and Study Centre, Ratsupites street 1, k-1, Riga LV-1067, Latvia

<sup>3</sup> Plant Virology group, Latvian Biomedical Research and Study Centre, Ratsupites street 1, k-1, Riga LV-1067, Latvia

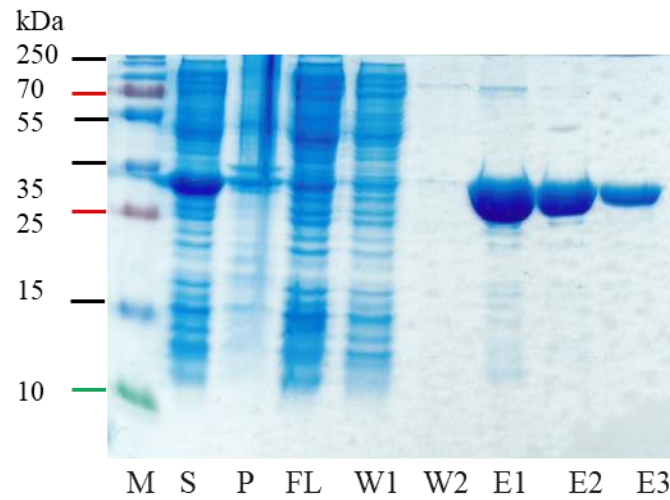
<sup>4</sup> Biotechnology core facility, Latvian Biomedical Research and Study Centre, Ratsupites street 1, k-1, Riga LV-1067, Latvia

\* Correspondence: inab@biomed.lu.lv

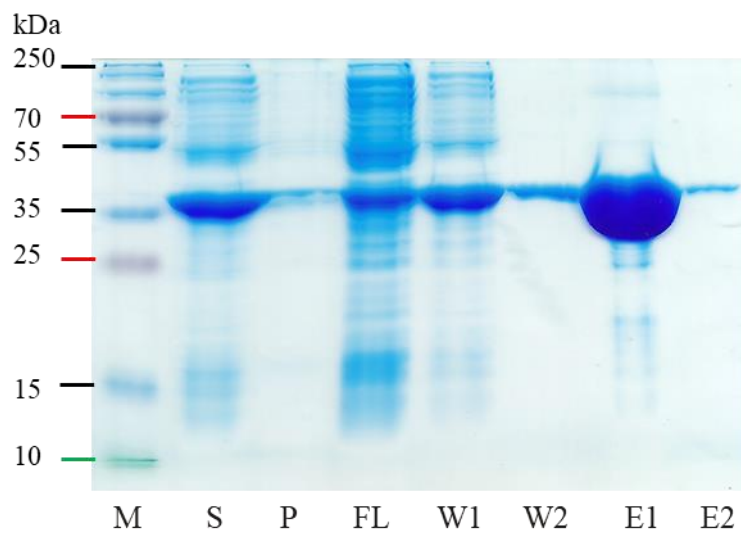
† These authors contributed equally to this work.



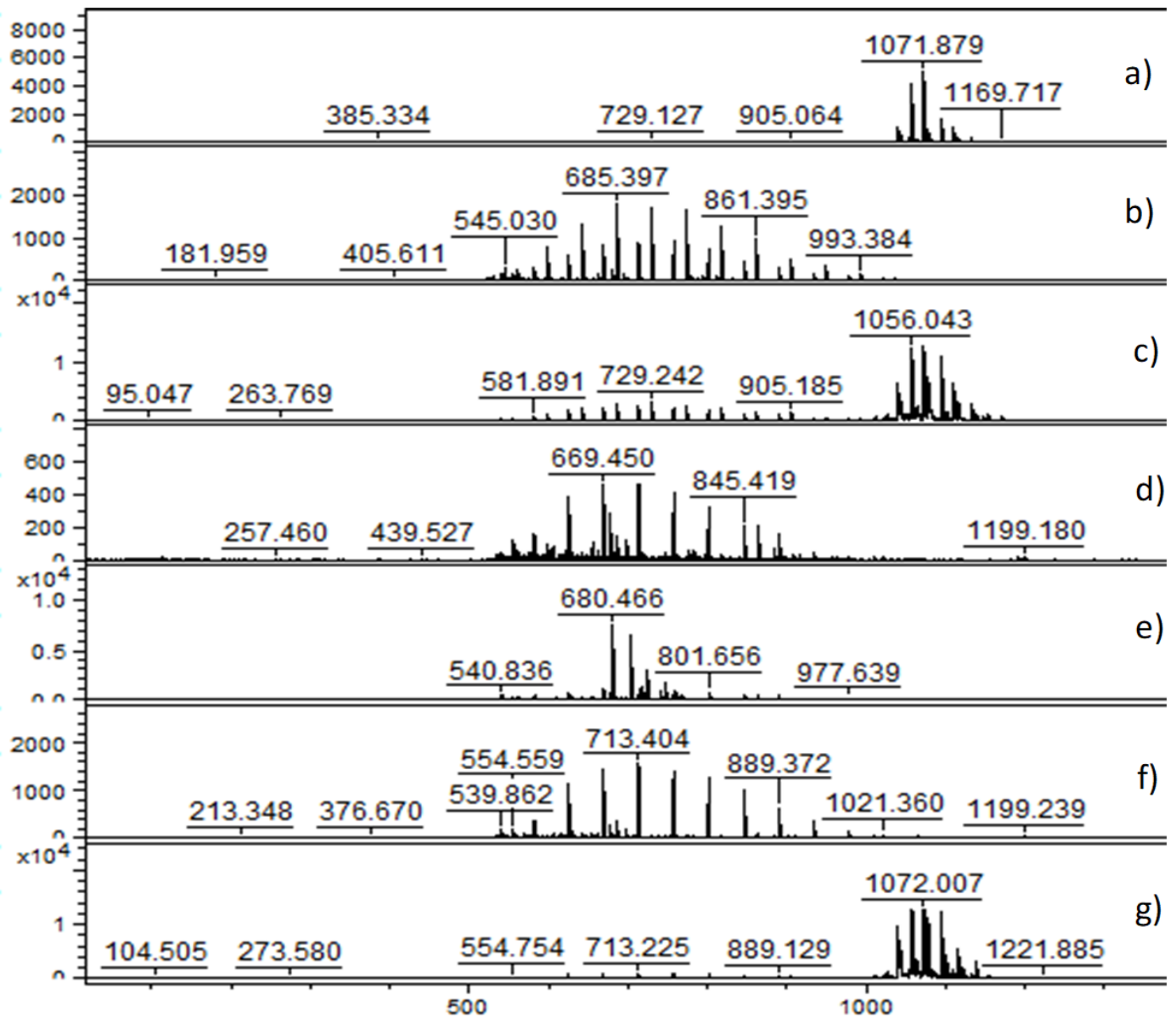
**Figure S1.** Sobemovirus Pro N-terminal part aa alignment. Artemisia virus A (ArtVA\_NC\_017914), blueberry shoestring virus (BSSV\_NC\_029578), cocksfoot mottle virus (CfMV\_NC\_002618), Cymbidium chlorotic mosaic virus (CyCMV\_NC\_027123), Cymbidium chlorotic spot virus (CyCSV\_KR996515), Imperata yellow mottle virus (IYMV\_NC\_011536), lucerne transient streak virus (LTSV\_NC\_001696), papaya lethal yellowing virus (PLYV\_NC\_018449), ryegrass mottle virus (RGMoV\_EF091714), Rottboellia yellow mottle virus (RoMoV\_NC\_027198), Rubus chlorotic mottle virus (RuCMV\_NC\_011187), rice yellow mottle virus (RYMV\_NC\_001575), southern bean mosaic virus (SBMV\_NC\_004060), subterranean clover mottle virus (SCMoV\_NC\_004346), southern cowpea mosaic virus (SCPMV\_NC\_001625), Sesbania mosaic virus (SeMV\_NC\_002568), Solanum nodiflorum mottle virus (SNMoV\_KC577470), sowbane mosaic virus (SoMV\_GQ845002), soybean yellow common mosaic virus (SYCMV\_KX096577), turnip rosette virus (TRoV\_AY177608), velvet tobacco mottle virus (VTMoV\_NC\_014509). **RAW** & **EAK**: start position of *E. coli* expressed Pro; underlined: transmembrane domain (TMD); **EA**: protease cleavage site for TMD.



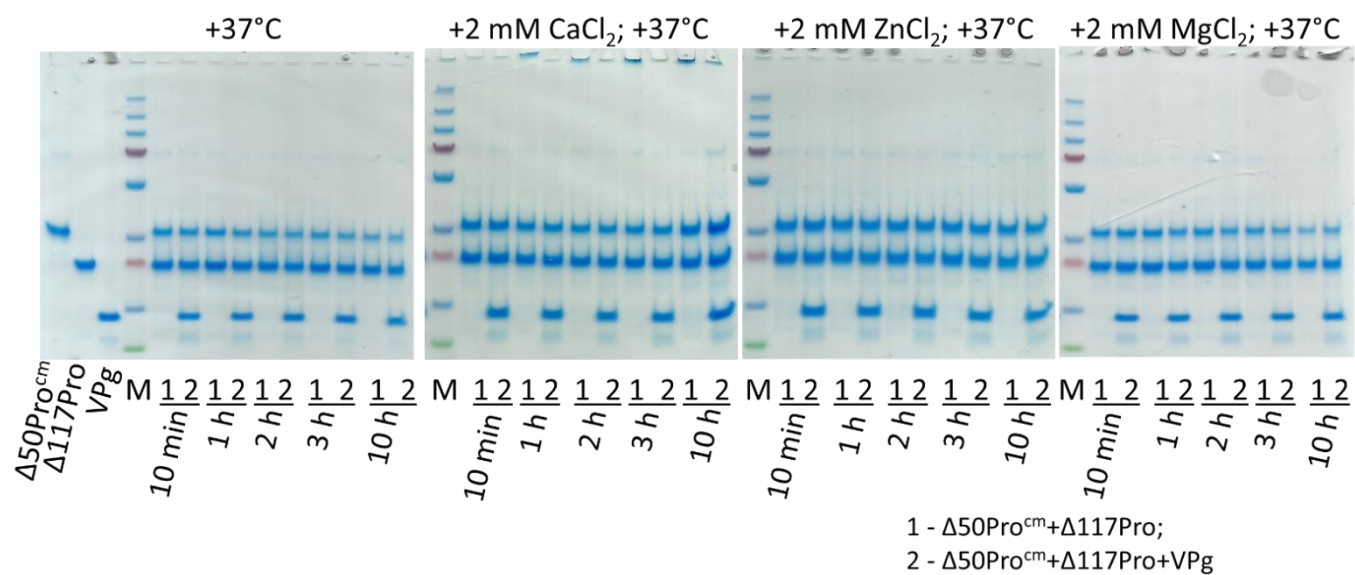
**Figure S2.**  $\Delta 50\text{Pro}^{\text{cm}}$  cleavage test *in cis*. M: protein marker (Page Ruler Plus, Thermo Fisher Scientific); S: soluble protein fraction; P: pellet; FL: flow fraction after IMAC column; W1: first wash; W2: second wash; E1: first elution fraction; E2: second elution fraction; E3: third elution fraction.



**Figure S3.**  $\Delta 50\text{Pro}^{\text{cm}}$ -VPg cleavage test *in trans*. M: protein marker (Page Ruler Plus, Thermo Fisher Scientific); S: soluble protein fraction; P: pellet; FL: flow fraction after IMAC column; W1: first wash; W2: second wash; E1: first elution fraction; E2: second elution fraction.

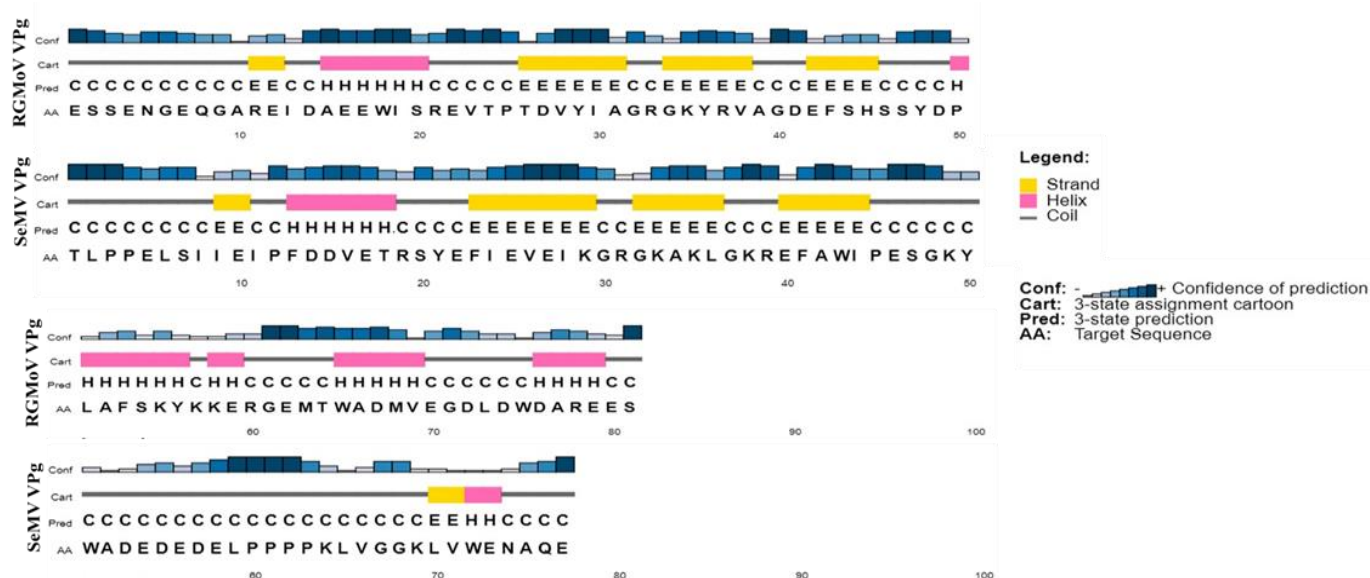


**Figure S4.** Mass spectrometry analysis of peptide cleavage. A: patterns of the synthetic peptide (Abz-RDE/AAVK(Dnp)-NH<sub>2</sub>, MW 1072.9 Da, Pepscan) in the peptide range; c: patterns of  $\Delta$ 117Pro<sup>cm</sup> (S<sup>159</sup> was mutated to A) in the peptide range; d: patterns of the synthetic peptide and  $\Delta$ 117Pro<sup>cm</sup> in the peptide range; e: patterns of  $\Delta$ 117Pro in peptide range; f: patterns of the synthetic peptide and  $\Delta$ 117Pro in the peptide range; g: patterns of the synthetic peptide and  $\Delta$ 117Pro in the peptide range with 20 mM DTT.

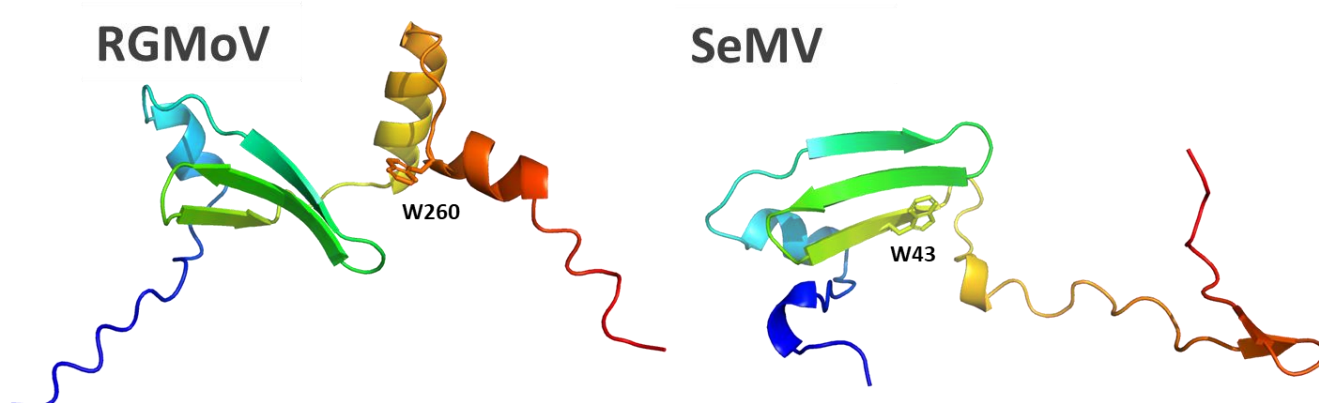


**Figure S5.** *In vitro in cis* cleavage test of Δ50Pro<sup>cm</sup>, Δ117Pro and VPg. M: protein marker (Page Ruler Plus, Thermo Fisher Scientific).

A



B



**Figure S6. *In silico* analysis of RGMoV and SeMV VPg sequences.** A: RGMoV and SeMV VPg secondary structure element determination by PsiPred [1] (C: coil; E: strand; H: helix); B: RGMoV and SeMV VPg 3D structure determination by AlphaFold [2], key interacting residues are selected as a stick model.

## References

1. Jones, D.T. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **1999**, 292, 195-202.
2. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A., *et al.* Highly accurate protein structure prediction with alphafold. *Nature* **2021**, 596, 583-589.

**Table S1.** List of used oligonucleotides (primers).

Construct	Primer	Primer sequence
N6H-Δ50Pro-E/A-VPg	RG-SerP-RAW-NdeI-F	5' CCA TAT GCG CGC TTG GCT AGC AAC CTC TCT GGA 3'
	RG-SP-VPg-E/A-R	5' TCC CTG CTC TCC GTT TGC ACT GGA TGC ACA GTT TGC ATG GAA AAT 3'
	RG-SP-VPg-E/A-F	5' ATT TTC CAT GCA AAC TGT GCA TCC AGT GCA AAC GGA GAG CAG GGA 3'
	RG2-VPg-HindIII-R	5' AAG CTT TCA TTC TTC ACG GGC GTC CCA ATC A 3'
	RG-VPg-P16-EcoRI-F	5' TGA ATT CTC CCA CTC AAG TTA TGA TCC TCT GGCT 3'
N6H-Δ50Pro-TMD-E/A	RGSP-E/A-R	5' AGC CAC AGC AGC AGC ATC ACG ATC T 3'
	RGSP-E/A-F	5' AGA TCG TGA TGC TGC TGC TGT GGC T 3'
	RGSP-XhoI-R	5' AGT GGA CTC GAG CCC CCA TCC CA 3'
VPg-P16-C6H	RG-P16-C6H-HindIII-R	5' CAA GCT TTC AGT GAT GAT GAT GAT GAT GAG CTG AGG GGG ACC CCT GGA 3'
	RG2-VPg-NcoI-F	5' CCA TGG GA AAC GGA GAG CAG GGA GCG CGC GA 3'
VPg-E/A-P16-C6H	RG-VPg-E/A-P16-F	5' TGA AGC GTC CAC AGG AAA TGA TAT TCC TTT AAA CT 3'
	RG-VPg-E/A-P16-R	5' AGT TTA AAG GAA TAT CAT TTC CTG TGG ACG CTT CA 3'
	RG2-VPg-NcoI-F	5' CCA TGG GA AAC GGA GAG CAG GGA GCG CGC GA 3'
	RG-VPg-P16-EcoRI-F	5' TGA ATT CTC CCA CTC AAG TTA TGA TCC TCT GGC T 3'
Δ50Pro-C6H	Pro-C-6-His-HindIII-R	5' AAG CTT TCA GTG ATG ATG ATG ATG ATG TTC ACT GGA TTC ACA GTT TGC ATG GA 3'
	RG-SerP-RAW-NcoI-F	5' CCA TGG CTC GCG CTT GGC TAG CAA CCT CTCTGGA 3'
Δ50Pro-VPg-wt-C6H	VPg-C-6-His-HindIII-R	5' AAG CTT TCA GTG ATG ATG ATG ATG ATG TTC TTC ACG GGC GTC CCA ATCA 3'
	RG-SerP-RAW-NcoI-F	5' CCA TGG CTC GCG CTT GGC TAG CAA CCT CTC TGG A 3'
	RG2-SerP-HindIII-R	5' AAG CTT TCA TTC ACT GGA TTC ACA GTT TGC ATG GAA AAT GGA GA 3'
Δ50Pro-N6H	RG-SerP-RAW-NdeI-F	5' CCA TAT GCG CGC TTG GCT AGC AAC CTC TCT GGA 3'
	SP-AAV-NcoI-N6H-F	5' ACC ATG GGC CAT CAT CAT CAT CAT CAC GCT GCT GTG GCT AAT TCC ATC GTCT 3'
Δ117SP <sup>cm</sup>	SP-S/A-F	5' ACC CGT CCA GGG TGG GCA GGG ACA CCC CT 3'
	SP-S/A-R	5' AGG GGT GTC CCT GCC CAC CCT GGA CGG GT 3'
	SP-NCE-HindIII-R	5' GAA GCT TTC ATT CAC AGT TTG CAT GGA AAA TGG AGA GGT TCGT 3'
Δ23Pro-C6H	RG-SerP-E/S-Nco-F	5' CCA TGG TCC AGG ACT GCG GTG AGA GGG CTG CTA 3'

**Table S2.** Data collection and refinement statistics.

	<b>Δ117 Pro</b> (PDB ID 6FEZ)	<b>Δ117 Pro-E/A-VPg</b> (PDB ID 6FFO)	<b>Δ117 Pro<sup>em</sup></b> (PDB ID 7YZV)
<b>Data collection</b>			
Space group	P 1 21 1	P 21 21 21	C 1 2 1
Wavelength, Å	0.9184	0.96411	0.9184
Unit cell dimensions, Å			
a, b, c (Å)	37.11, 115.99, 57.92	38.94, 67.52, 74.08	112.22 36.95 69.50
α, β, γ (°)	90.00° 105.27° 90.00°	90.00° 90.00° 90.00°	90.00° 126.29° 90.00°
Resolution, Å	2.3 (2.30-2.42)	2.1 (2.1-2.21)	1.6 (1.60-1.68)
R <sub>merge</sub> , %	12.3 (54.8)	7.7 (38)	14.8 (66)
I/σ	5.3 (2.1)	13.9 (4.2)	6.7 (2.2)
Completeness, %	95.98 (98.69)	99.8 (99.8)	95.9 (97.0)
Redundancy	3.0 (3.1)	5.8 (5.9)	5.7 (5.8)
<b>Refinement</b>			
Chains in asymmetric unit	2	1	1
Resolution (Å)	57.99-2.3	49.90-2.10	56.02 – 1.60
R <sub>work/free</sub> (%)	22.946/29.018	17.623/23.481	22.21/26.27
No. of Rfree set (%)	5.1	5.0	4.75
Average B-factor (Å <sup>2</sup> )	45.366	31.683	18.13
R.m.s deviations			
Bond lengths, Å	0.016	0.018	0.0053
Bond angles, (°)	1.976	1.897	0.7979
Ramachandran outliers, %			
Favored	92	95	96
Allowed	6	5	4
Outliers	2	0	0

The numbers in parentheses are for the highest-resolution shell.