Supporting Information

AbouHaidar et al. 10.1073/pnas.1402814111

SI Materials and Methods

Isolation of RYMV and Cloning into Plasmid Expression Tag 29(c). A full-size cDNA copy of the RNA of small circular satellite of the rice yellow mottle virus (scRYMV) was cloned in pBlueScript (pBS) (+) (1). A head-to-tail dimer and trimer were generated in pBS (+) using the unique XmaI site of the scRYMV cDNA. For protein expression analysis, the EcoRI-HindIII fragments of the recombinant pBS-dimer and pBS-trimer clones were inserted into the plasmid expression tag (pET)29(c) expression vector (Novagen), fusing it in-frame with 39 aa of the vector at the N terminus to give the clones pETdimer and pETtrimer (pET52), respectively.

Overexpression of the scRYMV Generated Polypeptide in *Escherichia coli* and Generation of Antisera. *E. coli* BL21 (DE3) cells transformed with pET52 were induced with 1 mM IPTG to express the scRYMV-encoded protein. Cells carrying pET29 were used as a negative control. Upon SDS/PAGE analysis of total proteins, a distinct and highly intense band was observed at 19 kDa only in cells carrying pET52 and not in cells carrying the pET29(c). The size of this protein corresponded to the estimated molecular weight of the fusion protein, which, when optimally induced, could form as much as 60% of the total cellular protein. This 19-kDa protein band was excised from the gel, and the protein eluted (5 mg) by electroelution was used to raise rabbit antiserum (2). An antibody titer of the order of 1:1,000 was obtained.

In Vitro Translation and Immunoprecipitation Reactions. To isolate the scRYMV circular RNA free from any other forms of scRYMV RNA or RYMV genomic RNA, total nucleic acids were prepared from purified RYMV virus preparations and electrophoresed on a 7 M urea/4% (wt/vol) polyacrylamide gel. Circular RNA band was excised from the gel, electrophoresed again through urea-PAGE gel and the RNA eluted using the QIAEX II gel elution kit (Qiagen). The pET52 and pETdimer DNA templates for the translation reactions were prepared as described earlier (2). In vitro translation was performed by using the T7/T3 (TNT) coupled in vitro transcription/translation wheat germ system (Promega).

Typically, the translations were performed using 1 μg of DNA templates (pETdimer or trimer) or 3 μg of total RYMV viral RNA template containing 150 ng of the scRYMV circular RNA and 25 μCi of L-[^{35}S]cysteine for each reaction as per the manufacturer's instructions, with internal controls Luciferase DNA and Tobacco Mosaic Virus RNA. For the scRYMV circular RNA alone, 150 ng of the purified, eluted RNA was used for the translations.

Immunoprecipitation of translation products was accomplished as described previously (3).

While the entire protein(s) from a single immunoprecipitation reaction was loaded per well of the SDS/PAGE gel, 1/10th of the reaction volume was directly loaded on the gel for detecting the total translation products. Gels were dried, and detection of radioactive signals was carried out using Bio-Rad Personal FX Molecular Imager. Molecular weight size markers (Fermentas) were run on the same SDS/PAGE gel, and their positions were marked using radioactive ink. For all immunoprecipitation experiments, minus antibody and minus antigen negative controls were used.

Northern Blot Analysis for Detection of scRYMV Concatemers. Different forms of scRYMV RNA were detected by Northern hybridization using scRYMV (anti-sense) riboprobe. Probe was prepared using plasmid pBS (+) carrying a single copy of the scRYMV cDNA, which was first linearized with EcoRI or KpnI followed by in vitro transcription using T3 RNA polymerase (Fermentas) and UTP[α-^{35}S] (1,000 Ci/mMol; Amersham Biosciences). Probes with a specific activity of 1–3 × 10^8 dpm/μg were obtained. For detection of the linear, circular, and concatameric forms of scRYMV RNA, total RNA (3 μg) from RYMV virus particles and from RYMV-infected or uninfected rice leaves were electrophoresed on a denaturing 7 M urea, 8% polyacrylamide gel until the xylene cyanol-FF dye ran out of the gel. RNA bands from denatured gels were transferred onto nylon membranes (Hybond-N; Amersham Biosciences) and subjected to Northern analysis as described previously (4). Molecular weight size markers (RiboRuler Low Range RNA ladder; Fermentas) were loaded on the same gel, stained with ethidium bromide, and their positions were marked with radioactive ink before exposure to PI screen. The identities of the concatemers were established to be derived from the scRYMV, by hybridization to specific scRYMV labeled probes and electroelution of the individual dimer and trimer bands, followed by RT-PCR and sequencing.

Protein Extractions and Western and North-Western Analyses. For protein extractions from *E. coli* carrying pET52 or pET29(c) (negative control), 0.5 mL of cells (OD_{600} of 0.8) were pelleted and the pellets boiled in 2× Laemmli SDS gel-loading buffer (5). Proteins from RYMV virus particles were prepared by boiling 25 μg purified virus for 5 min in equal volume of 2× SDS/PAGE loading buffer. For preparation of total proteins from RYMV-infected rice plants, 0.2 g of plant tissue was homogenized in liquid nitrogen in a buffer containing 2% (vol/vol) Triton X-100, 0.5% Nonidet P-40, and 7 M urea. After a brief spin at 10,000 × *g*, 50 μL of the supernatant (500 μg) of dissolved proteins was used for SDS/PAGE and Western blotting.

For North-Western analysis, 500 μg of all protein preparations from RYMV-infected or uninfected rice, from *E. coli* cells carrying pET52 or pET29 and 25 μg of RYMV virus particles were electrophoresed on a 6–20% gradient SDS/polyacrylamide gel, and Western-transferred (2). North-Western analysis was performed as described previously (4). The ^{35}S-labeled scRYMV and potato virus X (PVX) RNA (+) sense riboprobes were generated by T7 RNA polymerase in vitro transcription of a HindIII-linearized pBS DNA carrying a single (monomer) copy of scRYMV or a clone of 1.8 Kbp at 3′ end of PVX RNA. All protein samples for Western and North-Western analyses were prepared under the same conditions. Equivalent protein amounts (500 μg, calculated by Bradford assay and by OD at 280 nm) were loaded on SDS/PAGE. Equal loading for all SDS/PAGE gels was also confirmed by Coomassie blue staining.

Mutation of Initiation AUG to AAU and Tandem Termination Codons (UGAUGA to CUCGAG). For mutation of AUG into AAU, an AseI-tailed (ATTAAT) forward primer specific for 5′ end of the scRYMV trimer and an XhoI-tailed (CUCGAG) reverse primer were used for PCR and cloning. For testing the function of the tandem termination codons, NdeI-tailed (CATATG) forward primer specific for the 5′ end of the scRYMV pETdimer was used in a PCR, which also contains an XhoI-tailed (CUCGAG) reverse primer specific for the 3′ end of the scRYMV sequence (just before the last UGAUGA sequence). This resulted in the extension of the 19-kDa recombinant scRYMV

protein to a His-tagged 20-kDa protein that terminated at the singular UGA codon at the end of the C-terminal His-tag coding sequence. This protein is then easily detectable by the polyclonal anti–His-tag antibody (Genscript).

LC-MS/MS Analysis and Database Search of scRYMV-Specific Proteins from pET52-Expressing *E. coli* and RYMV-Infected Rice. Equal amounts (500 μg) of proteins extracted from plants or *E. coli* cells were first loaded onto SDS/PAGE gels followed by silver staining. Equivalent bands from lanes containing healthy control or infected rice samples or from *E. coli* cells containing pET29 or pET52 plasmids were cut out of the gel based on the size of the molecular weight ladder bands. Silver stained gel bands were destained and subjected to in-gel digestion with trypsin. Proteins were reduced with 10 mM DTT for 30 min at 56 °C, alkylated with 100 mM iodoacetamide for 30 min at room temperature. Gel bands were shrunk with 50% (vol/vol) acetonitrile/25 mM ammonium bicarbonate for 15 min and then rehydrated with 50 mM ammonium bicarbonate (pH 7.8). Proteins were subjected to overnight digestion with trypsin at 37 °C. Supernatants were collected and treated with two alternative 50 mM ammonium bicarbonate/acetonitrile washes. Extracted peptides were lyophilized, zip tipped, and subjected to LC-MS/MS analyses using an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific) coupled to a nano-LC system (Eksigent; AB SCIEX). Zip-tipped peptide digests (20 μL) were sampled with an autosampler (Eksigent; AB SCIEX) injected onto a trap column (ZORBAX 300SB-C18, 5 μm, 5 × 0.3 mm; Agilent), followed by an analytical column (ZORBAX 300SB-C18, 3.5 μm, 150 × 0.75 mm; Agilent) in which the peptides were resolved under gradient elution for 90 min at a flow rate of 400 nL/min. Raw data files were converted to mzXML files and searched against target databases with the following parameters: (*i*) mass tolerance in MS mode of 20 ppm; (*ii*) mass tolerance in MS/MS mode of 0.6 Da; (*iii*) fixed modification, carbamidomethylation; (*iv*) variable modifications, oxidation; and (*v*) missed cleavages, 2. The output from the searches was validated by using Scaffold 3.6 (Proteome Software).

Database Searching. Charge state deconvolution and deisotoping were not performed. All MS/MS samples were analyzed using X! Tandem (version X! Tandem Sledgehammer 2013.09.01.1; GPM). X! Tandem was set up to search the RYMV_Extension database assuming the digestion enzyme trypsin. X! Tandem was searched with a fragment ion mass tolerance of 0.80 Da and a parent ion tolerance of 20 ppm. Carbamidomethyl of cysteine was specified in X! Tandem as a fixed modification. Glu→pyro-Glu of the N terminus, ammonia loss of the n-terminus, Gln→pyro-Glu of the N terminus, deamidation of asparagine and glutamine, oxidation of methionine and tryptophan, and dioxidation of methionine and tryptophan were specified in X! Tandem as variable modifications.

Criteria for Protein Identification. Scaffold (version Scaffold_4.2.1; Proteome Software) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95% probability by the Peptide Prophet algorithm (6) with Scaffold delta-mass correction. Protein identifications were accepted if they could be established at greater than 99.9% probability and contained at least two identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (7). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

1. Collins RF, Gellatly DL, Sehgal OP, Abouhaidar MG (1998) Self-cleaving circular RNA associated with rice yellow mottle virus is the smallest viroid-like RNA. *Virology* 241(2):269–275.
2. Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY), 2nd Ed.
3. Philipson L, et al. (1978) Translation of MuLV and MSV RNAs in nuclease-treated reticulocyte extracts: enhancement of the gag-pol polypeptide with yeast suppressor tRNA. *Cell* 13(1):189–199.
4. Tsai M-S, Hsu Y-H, Lin N-S (1999) Bamboo mosaic potexvirus satellite RNA (satBaMV RNA)-encoded P20 protein preferentially binds to satBaMV RNA. *J Virol* 73(4):3032–3039.
5. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227(5259):680–685.
6. Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* 74(20):5383–5392.
7. Nesvizhskii AI, Keller A, Kolker E, Aebersold R (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* 75(17):4646–4658.

Table S1. LC-MS/MS analysis of the scRYMV-specific proteins expressed in *E. coli*

Peptide	Modification(s)	Charge	Observed *m/z*	Delta, Da	X! Tandem, −log e	No. of peptides	Actual mass (Da)
(–)MSQEELGGTQEFHPGRPGRLGA(V)	Acetyl (+42)	4	600.31	2.1	5.24	5	2,397.19
(R)ASLWSEPGLQGAWR(R)	—	2	779.39	−0.0006	14.82	97	1,556.77
(P)GISPGSTWAAR(S)	—	2	551.78	−0.0064	11.77	36	1,101.55
RLGAVHRASLWSEPGLQGAWRRN	Dioxidation (+32), Dioxidation (+32)	4	604.06	2	3.21	1	2,412.20
(R)NFTRVDLGGEPCTGR(R)	Deamidated (+1),	4	420.71	0.027	3.55	3	1,678.80
(R)AQGVAVER(A)	Carbamidomethyl (+57) Deamidated (+1)	2	415.23	−0.0026	7.14	13	828.44
(R)APAASSCAGGGDFVSSLTDTDEPR(G)	Carbamidomethyl (+57)	3	790.01	−0.014	15	96	2,367.02
(R)GLEAKPVCWDHSDHQSSCSGSFQLR(R)	Carbamidomethyl (+57)	3	963.43	−0.02	14.92	7	2,887.26
(R)RFCFEPYRH(A)	Carbamidomethyl (+57)	3	437.87	−0.0089	7.96	77	1,310.59
(G)EPCTGRRCGASLASK(G)	Carbamidomethyl (+57), Carbamidomethyl (+57)	4	413.45	1	3.68	1	1,649.78
(R)RCGASLASKGPGGETGLLGPLGPS(V)	Carbamidomethyl (+57)	3	745.05	−0.025	4.96	1	2,238.12
(T)AAAKFERQHMDSPDLGTLVPR(G)	—	4	586.05	2.0	3.66	6	2,340.17
(R)QHMDSPDLGTLVPR(G)	Ammonia-loss (−17), Oxidation (+16)	2	782.87	−0.0058	15.00	21	1563.73
(R)GSMGYLWIR(I)	Oxidation (+16)	2	549.77	−0.0043	10.34	58	1,097.53
(G)ETGLLGPLGPSVIVLRQL(–)	Dehydrated (−18)	4	462.29	2.0	4.48	1	1,845.12
(R)FCFEPYRHAKRNLEAPRNFTRVDLGG(–)	Carbamidomethyl (+57)	4	788.66	1	7.35	1	3,150.62
(–)RALPTMSQEELGGTQEFHPGR(P)	—	4	586.05	0.045	4.33	2	2,340.17

Peptide (T)AAAKFERQHMDSPDLGTLVPR(G) occurs close to the start of the 19-kDa fusion protein. Peptide MSQEELGGTQEFHPGRPGRLGA(V) represents the start of the 16 kDa generated as a result of internal initiation at the AUG codon within the tandem UGAUGA sequence. Peptides marked in red represent the readthrough peptides and those marked in blue are the peptides connecting terminal peptide of the 16 kDa with the readthrough sequence. Compare the frequency (n = 77) of the terminal peptide (R)RFCFEPYRH(A) of the 16 kDa with that (n = 1) of the (R)FCFEPYRHAKRNLEAPRNFTRVDLGG(–) readthrough peptide that ends at the downstream UAG codon. The last peptide (shown in green) RALPTMSQEELGGTQEFHPGR(P) represents the crossing over of the scRYMV RNA translation from the end of the third (last) reading frame to the beginning of the first reading frame, showing continued protein synthesis around the circular scRYMV RNA.

Table S2. LC-MS/MS analysis of the scRYMV-encoded polypeptides from RYMV-infected rice

Peptide	Modification (s)	Charge	Observed *m/z*	Delta, Da	X! Tandem, −log e	No. of peptides	Actual mass (Da)
MSQEELGGTQEFHPGRPGR(L)	Deamidated (+1),	3	706.34	2	2.85	9	2,116.00
(R)LGAVHRASLWSEPGLQGAWR(R)	Deamidated (+1)	3	731.07	0.032	2.43	119	2,190.18
(R)RFCFEPYRHAKRNLEAPR(N)	Carbamidomethyl (+57),	3	783.72	0.97	1.4	2	2,348.15
(R)NFTRVDLGGEPCTGR(R)	Deamidated (+1)	2	841.39	2	1.47	3	1,680.78
(R)RFCFEPYRH(A)	Deamidated (+1), Carbamidomethyl (+57)	—	—	—	—	—	—
	Carbamidomethyl (+57)	3	437.87	−0.0064	6.89	63	1,310.59
(R)RCGASLASKGPGGETGLLGPLGPSV(I)	Carbamidomethyl (+57)	4	585.8	2	3.17	7	2,339.19
(R)SVGTTRTISH(R)	—	2	529.79	0.0067	3.96	6	1,057.56
(R)SVGTTRTISHRAPAASSCAGGDFV(S)	Carbamidomethyl (+57)	3	821.41	0.028	1.07	2	2,461.21
(R)AQGVAVERAWPPRGLEAKPVC(W)	Carbamidomethyl (+57)	3	764.41	0.0025	1.12	2	2,290.20
(R)GLEAKPVCWDHSDHQSSCSGSFQLR(R)	Carbamidomethyl (+57), Carbamidomethyl (+57)	3	963.43	−0.020	15.00	7	2,887.26
(D)HSDHQSSCSGSFQLR(R)	Carbamidomethyl (+57)	3	578.59	1	3.33	2	1,732.76
(D)FVSSLTDTDEPR(G)	—	2	684.34	1	2.18	100	1,366.66
(R)HPGISPGSTWAARSRAQGVAVE(R)	—	3	745.4	0.043	1.3	40	2,233.18
(A)VERAWPPR(G)	—	2	506.29	1	1.96	4	1,010.56
(L)EAPRNFTR(V)	—	2	495.77	0.017	1.34	1	989.52
(A)AQGAEILFR(A)	—	2	502.78	0.032	1.85	1	1,003.55
(R)FCFEPYRHAKRNLEAPRNFTRDLGGEPCTGR(R)	Carbamidomethyl (+57), Deamidated (+1), Deamidated (+1), Carbamidomethyl (+57)	4	964.48	2.0	3.55	1	3,853.87
(-)RALPTMSQEELGGTQEFHPGRPGRLGA(V)	Oxidation (+6)	4	727.87	−0.0086	13.17	5	2,907.43

Peptide MSQEELGGTQEFHPGRPGR(L) represents the start of all of the scRYMV polypeptides. Peptide RRFCFEPYRH (occurring 63 times) signals the end of the major 16-kDa polypeptide obtained as a result of translation in the first two completely overlapping reading frames. The minor readthrough peptides (shown in red) are generated by translation of the scRYMV in the third reading frame. Marked in blue are the connecting peptides that bridge the 16-kDa sequence with the readthrough sequence. The last peptide in the table (–)RALPTMSQEELGGTQEFHPGRPGRLGA(V) (marked in green) represents the crossing over of the scRYMV RNA translation from the end of the third (last) reading frame to the beginning of the first reading frame. Its low frequency implies that this could be a rare, yet probable occurrence.