Journal of General Virology (1997), **78**, 237–241. Printed in Great Britain SHORT COMMUNICATION

In vivo expression of an overlapping gene encoded by the cucumoviruses

Bu-Jun Shi, Shou-Wei Ding† and Robert H. Symons

Department of Plant Science, Waite Institute, University of Adelaide, Glen Osmond, SA 5064, Australia

We recently reported the molecular characterization and functional analysis of an overlapping gene 2b encoded by RNA 2 of the Q strain of cucumber mosaic cucumovirus (Q-CMV). We show here that the homologous gene encoded by the V strain of tomato aspermy cucumovirus (V-TAV) and the WAII strain of CMV (WAII-CMV), which is in a different subgroup to Q-CMV, is also expressed *in vivo* by demonstrating the accumulation of the mRNA (RNA 4A) and its protein in infected plants. Interestingly, RNA 4A of V-TAV is encapsidated in virions as found previously for Q-CMV whereas WAII-CMV contains very little RNA 4A in virions. As the 2b gene is conserved in all 10 cucumoviral species or strains sequenced to date and the 2b gene is expressed for three of these viruses, we conclude that the 2b gene is a common feature of the *Cucumovirus* genus.

The genus *Cucumovirus* contains three species, cucumber mosaic virus (CMV), tomato aspermy virus (TAV) and peanut stunt virus (PSV). Strains of CMV are further classified into two major subgroups (I and II) based on serological and nucleic acid hybridization analyses (Piazzolla *et al.*, 1979). Recent data showed that viruses from the same subgroup share 92–98% sequence similarities and those from different subgroups share 71–79% sequence similarities (Palukaitis *et al.*, 1992).

It has been long known that the tripartite single-stranded RNA genome of cucumoviruses encodes four proteins. Proteins 1a and 2a, encoded by RNAs 1 and 2 respectively (Rezaian *et al.*, 1984, 1985), are involved in replication of the virus (Nitta *et al.*, 1988; Hayes & Buck, 1990). RNA 3 encodes two proteins (Davies & Symons, 1988), 3a and coat protein (CP). While 3a is translated from RNA 3, CP is expressed via RNA 4, a subgenomic RNA of RNA 3 (Schwinghamer & Symons, 1977). Both 3a and CP are required for virus cell-to-cell movement (Suzuki *et al.*, 1991).

Recently, we reported that RNA 2 of Q-CMV, which is a subgroup II strain, encodes an additional open reading frame, ORF 2b (Ding *et al.*, 1994). The mRNA for the 2b gene, called RNA 4A (Peden & Symons, 1973), is encapsidated in the viral particles, is 682 nucleotides (nt) in length, and is identical in sequence to the 3′-terminal 682 nt of RNA 2; thus, it is a subgenomic RNA of RNA 2. RNA 4A encodes the complete ORF 2b of 100 codons; the translational product of Q-CMV ORF 2b has been detected in infected plants by Western blot analysis (Ding *et al.*, 1994) and the 2b gene shown to encode a host-specific long-distance virus movement function (Ding *et al.*, 1995 *a*). We show here that the 2b gene encoded by two other cucumoviruses, tomato aspermy virus V strain (V-TAV) and cucumber mosaic virus WAII strain (WAII-CMV, a subgroup I strain; Wahyuni *et al.*, 1992) is also expressed in infected plants.

V-TAV was described by Habili & Francki (1974) and the complete nucleotide sequences of RNAs 1 and 2 were determined by Bernal *et al.* (1991) and Moriones *et al.* (1991). Fig. 1(*a*) shows the structural features of V-TAV RNA 2. Total RNAs were prepared from V-TAV-infected *Nicotiana glutinosa* and the 3′-terminal 628 nt of RNA 2 amplified by RT-PCR using primers BJ1 (5′ AGGATCCTGGGACCCCTAGGGGGAACCTACGGA 3′, complementary to nucleotides 3049–3074 of RNA 2) and BJ2 (5′ TGGATCC***ATG***GCAAGCATCGAGATCCCTCTACA 3′, corresponding to nucleotides 2447–2470 of RNA 2) (Ding *et al.*, 1995 *b*). A *Bam*HI site (underlined) was added to the 5′ end of both primers and the initiation codon of ORF 2b is shown as bold italics. The amplified PCR fragment was cloned at the *Sma*I site of pBluescript SK(+) to yield pBSK2NVT.

Firstly, we determined whether RNA 4A (mRNA of the 2b gene) was produced *in vivo* from RNA 2 of V-TAV. Encapsidated viral RNAs and total RNAs were extracted from

Author for correspondence: Bob Symons.
Fax +61 8 8303 7102. e-mail bsymons@waite.adelaide.edu.au

† **Present address:** Institute of Molecular Agrobiology, National University of Singapore, 59A, The Fleming, Science Park Drive, Singapore 118240.

The partial RNA 2 nucleotide sequence of WAII-CMV reported in this paper has been deposited in GenBank under accession no. U64435.

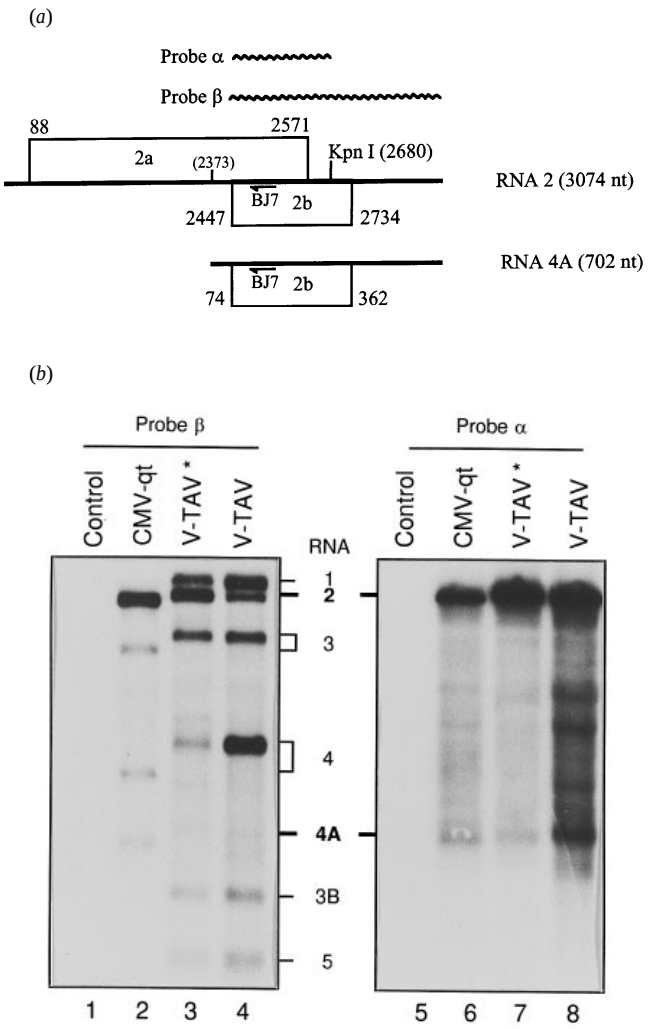

Fig. 1. Structure and expression of V-TAV RNA 2. (*a*). Structure of V-TAV RNA 2 and RNA 4A. ORFs for both 2a and 2b genes are shown as open boxes. The wavy lines indicate the positions of strand-specific RNA probes, used in the Northern blot hybridization, which are complementary to 627 nt (probe α, between 2b initiation site and 3′ end) and 204 nt (probe *β*, between 2b initiation site and *Kpn*I site) of RNA 2. The arrowhead lines within RNA 2 and RNA 4A ORFs indicate the positions of primers used in RNA dideoxynucleotide sequencing. (*b*) Northern blot analysis of viral RNAs. V-TAV virion RNAs (0·2 μg, lanes 3 and 7) and total RNAs (5 μg) extracted from *N. glutinosa* inoculated with sterile water (lanes 1 and 5), CMV-qt (lanes 2 and 6) or V-TAV (lanes 4 and 8) were electrophoresed on a 1·2% agarose gel containing 1·1% formaldehyde, transferred onto a Hybond-N+ membrane (Amersham) and hybridized with probe *β* (lanes 1–4) or probe α (lanes 5–8). The positions of viral RNAs 1, 2, 3, 4, 4A, 3B and 5 are indicated, of which the RNA 3B is a newly discovered subgenomic RNA derived from RNA 3 (unpublished).

virus-infected *N. glutinosa* and analysed by Northern blot hybridization using *in vitro* transcribed ^{32}P-labelled RNA probes as described in Ding *et al.* (1995*a*). Probe *β* (Fig. 1*a*), transcribed from pBSK2NVT, is complementary in sequence to the 3′-terminal 628 nt of V-TAV RNA 2 and was expected to hybridize to all V-TAV RNAs that contain the 3′ conserved region and/or the ORF 2b coding region. This probe detected seven major RNA species (Fig. 1*b*, RNAs 1–5) both in total V-TAV-infected plant RNAs (lane 4) and in virion RNAs (lane 3). The four larger RNAs corresponded to the known RNAs 1–4 whereas the two smaller RNAs at the bottom of the gel represent two novel subgenomic RNAs of RNA 3, RNAs 3B and 5 (unpublished). The RNA band above 3B most likely corresponds to RNA 4A as it is similar in size to the known RNA 4A (667 nt) of a Q-CMV chimera, CMV-qt (lane 2). The genome structure of CMV-qt is identical to that of Q-CMV except for the ORF 2b coding sequence, which is from V-TAV (Ding *et al.*, 1996).

Probe α (Fig. 1*a*), which is complementary in sequence to the 5′ 234 nt of the V-TAV ORF 2b coding sequence (nt 2447–2680 of RNA 2), strongly hybridized only to two (RNAs 2 and 4A) of the seven major RNA species detected above from either CMV-qt or V-TAV (Fig. 1*b*, lanes 6, 7 and 8). The nature of the bands between RNAs 2 and 4A (lanes 6–8) detected by probe α has not been determined. Taken together, the above results indicate that the V-TAV RNA 4A predicted (Ding *et al.*, 1994) was produced in infected plants and was encapsidated in virions.

RNA dideoxy sequencing (Fichot & Girard, 1990) was used to map the 5′-terminal nucleotide of the V-TAV RNA 4A using total V-TAV-infected plant RNAs and primer BJ7 (5′ TCTCGTGTAGAGGGATCT 3′, complementary to nt 2460–2477 of V-TAV RNA 2 as shown in Fig. 1*a*). The 5′-terminal nucleotide of RNA 4A was mapped to nt 2373 of RNA 2 and the size of V-TAV RNA 4A was determined as 702 nt. The size of V-TAV RNA 4A correlates well with its relative rate of migration in denaturing agarose gel electrophoresis as compared to the known size of 667 nt for CMV-qt RNA 4A (compare lanes 3 and 2, Fig. 1*b*). Thus, RNA 4A of V-TAV encodes the complete ORF 2b with a 5′ UTR of 73 nt and a 3′ UTR of 340 nt (Fig. 1*a*). Significantly, the initiation site of V-TAV RNA 4A determined in this work agrees exactly with that predicted from sequence similarities of RNA 2 among five cucumoviruses around the known initiation site of Q-CMV RNA 4A (Ding *et al.*, 1994).

The 2b protein encoded by V-TAV was expressed in *Escherichia coli* as a fusion protein with glutathione *S*-transferase (GST) and a 2b-specific antiserum raised in rabbits as described (Ding *et al.*, 1994). The plasmid expressing the GST-2b fusion protein in *E. coli* was constructed by cloning the *Bam*HI fragment containing the V-TAV ORF 2b coding sequence and the 3′ UTR from pBSK2NVT into the *Bam*HI site of pGEX-2 (Smith & Johnson, 1988). To detect the 2b protein in virus-infected plants by Western blot analyses, leaf tissues from infected *N. glutinosa* plants were harvested 2 weeks after inoculation and successive protein extracts were prepared using three different extraction buffers according to von Arnim *et al.* (1993).

The results (Fig. 2) showed that the 2b protein (marked by an arrowhead) accumulated in *N. glutinosa* plants infected with V-TAV (lane 2). This protein was not detectable by the GST-

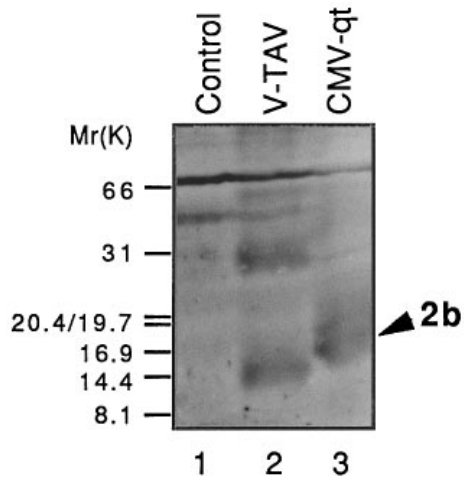

Fig. 2. Western immunoblotting detection of the gene 2b product in V-TAV-infected *N. glutinosa*. Protein extracts were prepared from healthy plants (lane 1), or plants infected with V-TAV (lane 2) or CMV-qt (lane 3) 2 weeks after inoculation. The position of the 2b protein is indicated by an arrowhead. Protein M_r markers were obtained from Promega.

2b antiserum in uninfected plants (lane 1) or in plants infected with Q-CMV (not shown), but it was detected in plants infected with CMV-qt (lane 3). Further, this protein was not detected when the preimmune serum or an antiserum against an unrelated GST fusion protein (GST fused with ORF 4 of the barley yellow dwarf virus-PAV, a gift of Masoud Shamsbakhsh) was used as the first antibody (not shown). Interestingly, the V-TAV 2b protein became soluble only after boiling in the ESB buffer containing SDS, urea and β-mercaptoethanol (von Arnim *et al*. 1993), suggesting an association of the 2b protein with cell walls/membranes. The size of the 2b protein was estimated as M_r 14000 under the conditions used which is larger than M_r 11000 predicted from the sequence; a similar difference was found previously for the Q-CMV 2b protein (Ding *et al*., 1994). In addition, it is likely that the other larger protein species of about M_r 28000 present in plants infected with V-TAV or CMV-qt is related to the V-TAV 2b protein. Most of this larger protein in extracts from CMV-qt-infected plants was not solubilized in the ESB buffer (not shown). It is not clear whether this larger species represents an aggregate of the 2b protein with other proteins or is simply a dimeric form of the 2b protein.

Using a strategy similar to that described for V-TAV, we next investigated whether the 2b gene was also expressed *in vivo* by WAII-CMV. The ORF 2b coding sequence plus the 3′ UTR of WAII-CMV RNA 2 was obtained by RT-PCR using primers derived from the known RNA 2 sequence of another subgroup I strain, Fny-CMV (Rizzo & Palukaitis, 1988), cloned (named as pBSK2NWAC) and sequenced. A region of 148 nt of WAII-CMV RNA 2 that is 5′ to the initiation codon of ORF 2b was directly sequenced by AMV reverse transcriptase (Fichot & Girard, 1990) using primer BJ8 that binds to a region 37 nt 3′ to the ORF 2b start codon (Fig. 3*a*). Thus, the 3′-terminal sequence of 779 nt of WAII-CMV RNA 2 was determined (GenBank accession no. U64435).

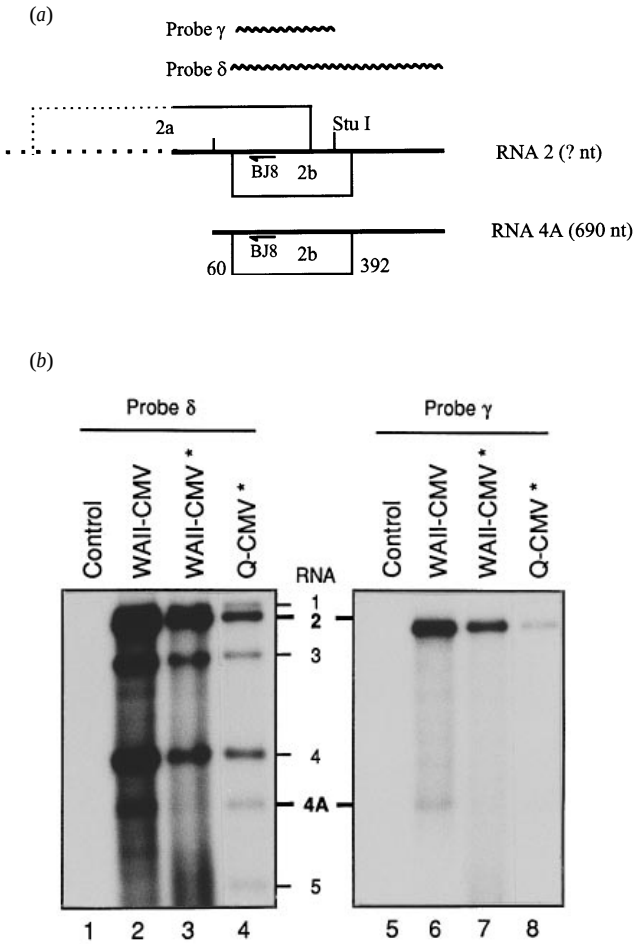

Fig. 3. Structure and expression of WAII-CMV RNA 2. (*a*). Structure of WAII-CMV RNA 2 and RNA 4A. Only 3′-terminal sequence of 779 nt of RNA 2 was determined. The remaining undetermined sequence is indicated by the dotted line. ORFs for both 2a and 2b genes are shown as open boxes. The wavy lines indicate the positions of strand-specific RNA probes, used in the Northern blot hybridization, which are complementary to 630 nt (probe δ, between 2b initiation site and 3′ end) and 251 nt (probe γ, between 2b initiation site and *Stu*I site) of RNA 2. The arrowhead lines within RNA 2 and RNA 4A ORFs indicate the positions of primers used in RNA dideoxynucleotide sequencing. (*b*) Northern blot analysis of viral RNAs. Virion RNAs (0·2 μg) of WAII-CMV (lanes 3 and 7) or Q-CMV (lanes 4 and 8) and total RNAs (5 μg) extracted from *N. glutinosa* inoculated with sterile water (lanes 1 and 5) or WAII-CMV (lanes 2 and 6) were electrophoresed on a 1·2% agarose gel containing 1·1% formaldehyde, transferred onto a Hybond-N^+ membrane (Amersham) and hybridized with probe δ (lanes 1–4) or probe γ (lanes 5–8). The positions of viral RNAs 1, 2, 3, 4, 4A and 5 are indicated.

Northern blot analysis (Fig. 3*b*) using probe δ (transcribed from pBSK2NWAC and complementary in sequence to the 3′-terminal 630 nt of WAII-CMV) identified in WAII-CMV-infected plants an RNA species similar in size to RNA 4A of Q-CMV (compare lanes 2 and 4). As the RNA species also hybridized (lane 6) to probe γ whose sequence is complementary only to the 5′ 240 nt of the WAII-CMV ORF 2b coding sequence (Fig. 3*a*), we conclude that this RNA

corresponds to the predicted RNA 4A. In contrast to RNA 4As of Q-CMV (Ding *et al.*, 1994) and V-TAV (see above), only very low levels of WAII-CMV RNA 4A which accumulated in infected plants (lanes 2 and 6) were encapsidated in the viral particles (lanes 3 and 7), indicating varying specificity of *in vivo* encapsidation of WAII-CMV RNAs.

The 5′-terminal nucleotide of WAII-CMV RNA 4A was further determined by RNA dideoxy sequencing as described above. This mapped the initiation site of RNA 4A (data not shown) and predicted a size of 690 nt for RNA 4A of WAII-CMV. The initiation site of WAII-CMV RNA 4A determined here is also in complete agreement with the site predicted for Fny-CMV RNA 4A (Ding *et al.*, 1994). The conserved sequence (V-TAV RNA 2 ^{2362}UUGAUCAAUUC**G**UUUU 3′) surrounding the initiation nucleotide (G in bold) of RNA 4A found among the cucumoviruses (Ding *et al.*, 1994) may be functionally important to the production of RNA 4A.

Repeated attempts failed to express the WAII-CMV 2b protein in *E. coli* using several protein expression systems (pGEX, pQE30 and pET), indicating a possible toxic effect of the protein to the bacterium. In addition, the WAII-CMV 2b protein was not detectable by the antiserum against the Q-CMV 2b-GST fusion protein, despite the fact that the 2b proteins of the two strains are 53% identical in sequence. Since the 2b protein was detected for both Q-CMV and V-TAV and RNA 4A accumulated in WAII-CMV-infected plants, we did not continue to try to demonstrate the accumulation of the WAII-CMV 2b protein.

We have shown in this work that the 2b gene encoded by two cucumoviruses (V-TAV and WAII-CMV) is expressed *in vivo* by detecting the accumulation of the mRNA (V-TAV and WAII-CMV) and protein product (V-TAV) in infected plants while the expression of the Q-CMV 2b gene in infected plants has been reported previously (Ding *et al.*, 1994). Furthermore, the 2b gene is conserved in RNA 2 of the 10 cucumoviruses or strains sequenced to date (Ding *et al.*, 1994; our unpublished observations). It is therefore concluded that the occurrence and expression of the 2b gene are common features of the *Cucumovirus* genus, and that the cucumoviruses contain three genomic RNAs (RNAs 1–3) and two subgenomic RNAs (RNAs 4 and 4A) that function as messenger RNAs for the *in vivo* expression of the five cucumoviral proteins, 1a, 2a, 3a, CP and 2b, respectively.

The authors wish to thank Neil Shirley for oligonucleotide synthesis, Wanxiang Li for technical help, Masoud Shams-bakhsh for antiserum against GST-ORF4 of BYDV-PAV and laboratory colleagues for useful discussions. This work was supported by the Australian Research Council Special Research Centre for Basic and Applied Plant Molecular Biology. B.J.S. is the recipient of a University of Adelaide Overseas Postgraduate Research Scholarship.

References

Bernal, J. J., Moriones, E. & Garcia-Arenal, F. (1991). Evolutionary relationships in the cucumoviruses: nucleotide sequence of tomato aspermy virus RNA 1. *Journal of General Virology* **72**, 2191–2195.

Davies, C. & Symons, R. H. (1988). Further implications for the evolutionary relationships between tripartite plant viruses based on cucumber mosaic virus RNA 3. *Virology* **165**, 216–224.

Ding, S. W., Anderson, B. J., Haase, H. R. & Symons, R. H. (1994). New overlapping gene encoded by the cucumber mosaic virus genome. *Virology* **198**, 593–601.

Ding, S. W., Li, W. X. & Symons, R. H. (1995*a*). A novel naturally occurring hybrid gene encoded by a plant RNA virus facilitates long distance virus movement. *EMBO Journal* **14**, 5762–5772.

Ding, S. W., Rathjen, J. P., Li, W. X., Swanson, R., Healy, H. & Symons, R. H. (1995*b*). Efficient infection from cDNA clones of cucumber mosaic cucumovirus RNAs in a new plasmid vector. *Journal of General Virology* **76**, 459–464.

Ding, S. W., Shi, B. J., Li, W. X. & Symons, R. H. (1996). An interspecies hybrid RNA virus is significantly more virulent than either parental virus. *Proceedings of the National Academy of Sciences, USA* **93**, 7470–7474.

Fichot, O. & Girard, M. (1990). An improved method for sequencing of RNA templates. *Nucleic Acids Research* **18**, 6162.

Habili, N. & Francki, R. I. B. (1974). Comparative studies on tomato aspermy and cucumber mosaic viruses. *Virology* **57**, 392–401.

Hayes, R. J. & Buck, K. W. (1990). Complete replication of a eukaryotic virus RNA in vitro by a purified RNA-dependent RNA polymerase. *Cell* **63**, 363–368.

Moriones, E., Roossinck, M. J. & Garcia-Arenal, F. (1991). Nucleotide sequence of tomato aspermy virus RNA 2. *Journal of General Virology* **72**, 779–783.

Nitta, N., Takanami, Y., Kuwata, S. & Kubo, S. (1988). Comparative studies on the nucleotide sequence of cucumber mosaic virus RNA 3 between Y strain and Q-strain. *Annals of the Phytopathological Society of Japan* **54**, 516–522.

Palukaitis, P., Roossinck, M. J., Dietzgen, R. G. & Francki, R. I. B. (1992). Cucumber mosaic virus. *Advances in Virus Research* **41**, 281–348.

Peden, K. W. C. & Symons, R. H. (1973). Cucumber mosaic virus contains a functionally divided genome. *Virology* **53**, 487–492.

Piazzolla, P., Diaz-Ruiz, J. R. & Kaper, J. M. (1979). Nucleic acid homologies of eighteen cucumber mosaic virus isolates determined by competition hybridization. *Journal of General Virology* **45**, 361–369.

Rezaian, M. A., Williams, R. H. V., Gordon, K. H. J. & Symons, R. H. (1984). Nucleotide sequence of cucumber mosaic virus RNA 2 reveals a translation product significantly homologous to corresponding proteins of other viruses. *European Journal of Biochemistry* **143**, 277–284.

Rezaian, M. A., Williams, R. H. V. & Symons, R. H. (1985). Nucleotide sequence of cucumber mosaic virus RNA 1: presence of a sequence complementary to part of the viral satellite RNA and homologies with other viral RNAs. *European Journal of Biochemistry* **150**, 331–339.

Rizzo, T. M. & Palukaitis, P. (1988). Nucleotide sequence and evolutionary relationships of cucumber mosaic virus (CMV) strains: CMV RNA 2. *Journal of General Virology* **69**, 1777–1787.

Schwinghamer, M. W. & Symons, R. H. (1977). Translation of the four major RNA species of cucumber mosaic virus in plant and animal cell-free systems and in toad oocytes. *Virology* **79**, 88–108.

Smith, D. B. & Johnson, K. S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase. *Gene* **67**, 31–40.

Suzuki, M., Kuwata, S., Kataoka, J., Masuta, C., Nitta, N. & Takanami, T. (1991). Functional analysis of deletion mutants of cucumber mosaic

virus RNA 3 using an in vitro transcription system. *Virology* **183**, 106–113.

von Arnim, A., Frischmuth, T. & Stanley, J. (1993). Detection and possible functions of African cassava mosaic virus DNA B gene products. *Virology* **192**, 264–272.

Wahyuni, W. S., Dietzgen, R. G., Hanada, K. & Francki, R. I. B. (1992). Serological and biological variation between and within subgroup I and II strains of cucumber mosaic virus. *Plant Pathology* **41**, 282–297.

Received 19 August 1996; Accepted 30 September 1996