

Similarities in the genomic sequence and coat protein structure of plant viruses

M. R. N. MURTHY and H. S. SAVITHRI*

Molecular Biophysics Unit and * Department of Biochemistry, Indian Institute of Science, Bangalore 560012, India

Abstract. The genomic sequences of several RNA plant viruses including cucumber mosaic virus, brome mosaic virus, alfalfa mosaic virus and tobacco mosaic virus have become available recently. The former two viruses are icosahedral while the latter two are bullet and rod shaped, respectively in particle morphology. The non-structural 3a proteins of cucumber mosaic virus and brome mosaic virus have an amino acid sequence homology of 35% and hence are evolutionarily related. In contrast, the coat proteins exhibit little homology, although the circular dichroism spectrum of these viruses are similar. The non-coding regions of the genome also exhibit variable but extensive homology. Comparison of the brome mosaic virus and alfalfa mosaic virus sequences reveals that they are probably related although with a much larger evolutionary distance. The polypeptide folds of the coat protein of three biologically distinct isometric plant viruses, tomato bushy stunt virus, southern bean mosaic virus and satellite tobacco necrosis virus have been shown to display a striking resemblance. All of them consist of a topologically similar 8-standard β -barrel. The implications of these studies to the understanding of the evolution of plant viruses will be discussed.

Keywords. Plant viruses; genomic sequence; coat protein structure; comparison; evolution.

Introduction

Many different types of viruses cause disease in plants. They are classified on the basis of their host and vector specificity, particle morphology, serological relationships and nature of nucleic acid (Kurstak, 1981). Plant viruses of a variety of shapes and sizes have been extensively studied. Many of them are isometric with a diameter of approximately 300 Å. They consist of a protein shell of identical subunits arranged with icosahedral symmetry. The shell serves to protect the genome.

Tobacco mosaic virus (TMV) particles are rod shaped with flat ends. The virion contains 2130 protein subunits which encapsidate a single stranded RNA of molecular mass 2.06×10^6 daltons. An intriguing class of plant viruses is that of alfalfa mosaic virus (AMV). Electron microscopic observation of purified AMV reveals bacillus shaped particles of different lengths. A mixture of 3 large particles is necessary for infection. Different length particles encapsidate different RNA molecules of mass 1.1, 1.0, 0.7, and 0.3 million daltons. Hence, AMV is said to have a divided genome. Bromo and cucumoviruses also possess a similar genomic organisation, although they are isometric with a $T = 3$ lattice (Caspar and Klug, 1962). The divided nature of the

Abbreviations used: TMV, Tobacco mosaic virus; TYMV, turnip yellow mosaic virus; CPMV, cowpea mosaic virus; AMV, alfalfa mosaic virus; SBMV, souther bean mosaic virus; TBSV, tomato bushy stunt virus; SDS, sodium dodecyl sulphate; CMV, cucumber mosaic virus; BMV, bean mosaic virus; STNV, satellite tobacco necrosis virus.

genomes of these viruses is revealed by gel electrophoresis of purified RNA preparations. In these viruses, the larger two RNAs of approximately 1 million daltons are encapsidated individually while the smaller two RNAs are encapsidated together resulting in particles of nearly identical sedimentation coefficients. Hence these viruses exhibit a single symmetrical peak in sucrose density gradients. This divided nature of the genome is characteristic of plant viruses and is not found in animal or bacterial viruses.

In the sixties and early seventies, a large number of biochemical experiments were carried out on plant viruses which led to a classification scheme based on the nature of forces that stabilize the virion rather than the similarities of biological expression (Kaper, 1975). In this scheme biologically distinct viruses such as turnip yellow mosaic virus (TYMV), cowpea mosaic virus (CPMV) etc are grouped together based on the strong hydrophobic association of protein subunits and relatively weak protein-nucleic acid interactions. Viruses of this group are insensitive to sodium dodecyl sulphate (SDS), RNase and neutral chloride salts. They also form empty capsids *in vivo* indicating that the strong association of protein subunits governs the assembly of these viruses. In contrast, cucumber mosaic virus (CMV) and AMV are grouped together based on the relatively strong protein-nucleic acid interactions in these viruses. They are degraded by high concentration of neutral chloride salts and low concentrations of SDS and are sensitive to ribonuclease. Southern bean mosaic virus (SBMV) and tomato bushy stunt virus (TBSV) on the other hand, are stabilized by protein-protein as well as protein-nucleic acid interactions.

Recently the three dimensional structure of TBSV (Harrison *et al.*, 1978) SBMV, (Abad-Zapatero *et al.*, 1980), Satellite tobacco necrosis virus (STNV)(Liljas *et al.*, 1982), TMV (Bloomer *et al.*, 1978) and the genomic sequences of AMV (Barker *et al.*, 1983). BMV (Ahliquist *et al.*, 1981), CMV (Gould and Symons, 1982) and TMV (Goelet *et al.*, 1982) have become available. A comparison of these results to understand the importance of structural and non-structural proteins in the life cycle of these viruses and to detect any possible evolutionary relationships is presented in this paper.

Results

Comparison of the structure of spherical plant viruses

Three-dimensional structures of TBSV, SBMV, and STNV have been determined to near atomic resolution (Harrison *et al.*, 1978; Abad-Zapatero *et al.*, 1980; Liljas *et al.*, 1982). The coat proteins of all these viruses have a disordered amino terminal arm that seems to interact with RNA. Following the amino terminal arm, the polypeptide chain folds into an 8-stranded anti-parallel β -barrel (S-domain). In addition to the arm and S-domain, TBSV coat protein has also a carboxy terminal domain (P-domain) where the polypeptide fold is similar to that of the S-domain. Quantitative comparison of the polypeptide folds of these viruses by the method of Rossmann and Argos (1975) indicated a striking similarity of S-domains (Rossmann *et al.*, 1983). The extent of similarity was found to be better when compared to the similarity in the NAD binding domains of dehydrogenases, although, smaller than the similarity between the α and β chains of haemoglobin. Based on these observations, a divergent evolutionary tree of

these viruses was suggested. Also, the similarity between the P and S domains in TBSV was taken to suggest a possible gene duplication (Rossmann *et al.*, 1983).

Comparisons with TYMV

Comparison of the secondary structure of TYMV predicted from its amino acid sequence (Peter *et al.*, 1972) with those of TBSV and SBMV reveals significant similarities (Argos, 1981). Recently, the coat protein sequences of SBMV (Hermodson *et al.*, 1982) and TBSV (Hopper *et al.*, 1984) have become available. We have compared these sequences with that of TYMV using Needleman and Wunch (1970) algorithm as modified by Dayhoff (1972). The mutation data matrix was taken from Dayhoff (1972). The table corresponds to 256 accepted point mutations per 100 residues. 10 randomized sequences were used to obtain the mean and standard deviation for the scores corresponding to unrelated sequences with an identical amino acid composition. Alignment score was expressed in units of standard deviation above the mean. The results are shown in table 1. As observed earlier (Harrison *et al.*, 1984), it is evident that SBMV and TBSV sequences are significantly homologous. The alignment score of 7.8 is comparable to the score of 8.1 between haemoglobin and myoglobin (Dayhoff, 1972). In contrast, the score between haemoglobin α and β chains is 19.3 (Dayhoff, 1972). TYMV also exhibits a weak relationship with TBSV, while no similarity is found between SBMV and STNV despite the striking similarity in the polypeptide folds of their coat proteins. Although this might suggest convergence to a structure suitable for self assembly of spherical viruses, it might be pointed out that the memory of ancestry is retained in the three-dimensional structures for considerably longer time scales compared to the similarity in sequences.

Table 1. Comparison of the coat protein sequences of SBMV, TBSV, TYMV and STNV. The alignment scores are expressed in units of standard deviation above the mean of scores for unrelated sequences with an identical amino acid composition.

	TBSV	SBMV	TYMV	STNV
TBSV	—	7.8	3.2	2.8
SBMV		—	1.5	-2.0
TYMV			—	-1.7

Comparison of sequences of viruses with tripartite genomes

The smallest of the RNAs (RNA4) of CMV, BMV and AMV code for their respective coat proteins (Schwinghamer and Symons, 1977; Shih and Kaesberg, 1973). The larger two RNAs (RNA1 and RNA2) code for proteins of unknown function. RNA3 is dicistronic and codes for a protein called 3a protein and the coat protein. The function of 3a protein is not known. In addition to RNAs 1-4, CMV supports a fifth RNA (RNA5), which causes severe necrosis in tomato plants (Kaper and Waterworth, 1981).

We have compared the sequences of RNA3 of BMV, CMV and AMV using a modified method of Jukes and Cantor (Murthy, 1983; Savithri and Murthy, 1983).

Non-structural proteins of BMV, CMV and AMV

The 3a proteins of BMV and CMV were found to possess a homology of 35% revealing their evolutionary relatedness (Murthy, 1983). In contrast, the 3a protein of AMV did not reveal significant homology with the corresponding proteins of BMV and CMV (Savithri and Murthy, 1983). However, six segments with good homology (30%) between AMV and BMV or CMV 3a proteins could be identified. In pairwise comparison, AMV 3a protein could be aligned with 18% homology with that of BMV and 14% with that of CMV. The minimum base change per codon required to convert AMV 3a protein sequence to BMV or CMV for the 274 common positions was 1.31. Although these similarities are on the borderline of significance (Doolittle, 1981) taken together with the similarity of their genomic organization and similarity in biophysical properties, we suggested that the viruses are evolutionarily related. More recently, Haseloff *et al.*, (1984) have demonstrated striking homology among the non-structural proteins encoded by AMV and BMV RNAs 1 and 2. These results probably reflect the common evolutionary origin of plant viruses with a tripartite genome. In addition, they point out the importance of non-structural proteins in the life cycle of these viruses.

Coat protein sequences of BMV and CMV

In contrast to the non-structural proteins, the coat protein sequences did not reveal significant similarities. In order to examine if similarities exist at a secondary structural level, the probable helical and sheet regions in these proteins were predicted by the method of Ptitsyn and Finklestein (Finklestein, personal communication). These predictions reveal significant β -structure in the coat protein of CMV while much less β -structure for BMV protein (figure 1). However, the CD spectrum of CCMV, a bromovirus serologically related to BMV (Verduin, 1978) is similar to that of CMV (Savithri, H. S. and Murthy, M. R. N., unpublished results) and both appear to be rich in β -structure (Verduin, 1978). Hence it is likely that despite the differences in the sequence and predicted secondary structures, the coat proteins of these viruses might have a similar three-dimensional structure.

Comparison of CMV, RNA and CARNA5

Gel electrophoresis of purified RNA from CMV has revealed that CARNA5 is probably co-encapsidated with RNA3 (Kaper and Waterworth, 1981). This might result from a common coat protein recognition site on CARNA5 and RNA3. Comparison of the CMV RNA3 and CARNA5 sequences did not reveal significant homology. It was possible to detect a 12 consecutive residue segment in CARNA5 (153–164) identical to residues 1904–1915 of CMV RNA3. The probability that these common residues occur by chance is 4%. This might indeed be the coat protein recognition site although 12 residues seem to be rather short for excluding cellular RNAs. Comparison of CARNA5 with RNA3 of BMV revealed that the 13 residues 1730–1742 of BMV RNA3 are identical to residues 77–89 of CARNA5. In this context,

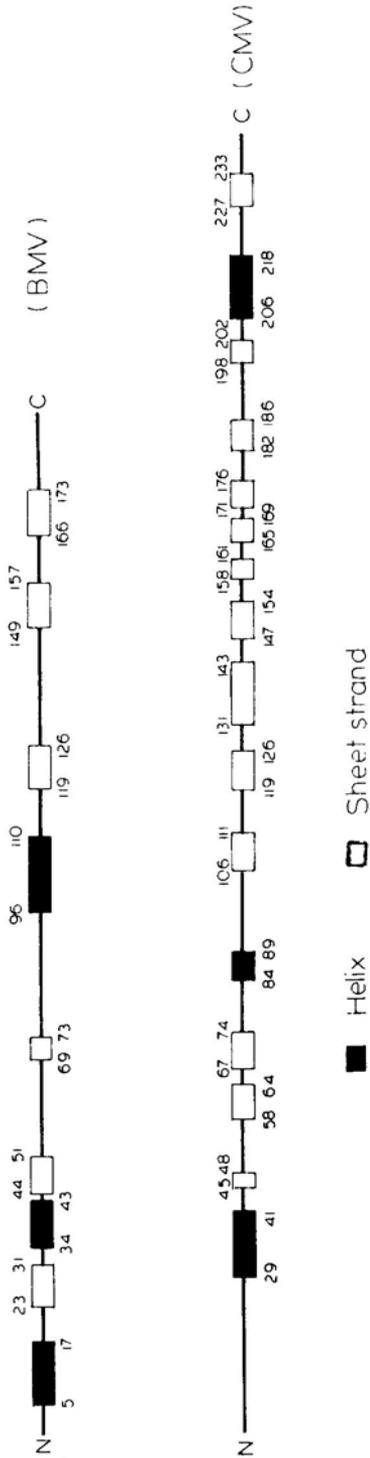


Figure 1. Predicted secondary structure of the coat proteins of CMV and BMV.

it would be of interest to examine if BMV supports the satellite RNA associated with CMV. However, the 12 and 13 residues segments of CARNA5 which are identical to the segments 1904–1915 and 1730–1742 of CMV and BMV, respectively, do not have any homology. Hence, the observed similarity is likely to be a random event.

Non-coding regions

The 5'-terminal non-coding regions of these viruses did not reveal significant homology. However, the 3'-terminal non-coding regions were highly homologous. This region might be folded into a tRNA like structure in many plant viruses. The 66% homology for the 3'-terminal 150 residues between the RNA3 of BMV and CMV point out the functional importance of the tRNA like structure in the life cycle of these viruses.

Discussion

SBMV and TBSV belong to the T = 3 group of plant viruses with an unsegmented, single stranded RNA genome. The particle stability in these viruses results from both protein-protein and protein-nucleic acid interactions. Although biologically distinct, these viruses are derived from a common ancestor. TBSV has probably undergone a gene duplication. These viruses are also likely to be evolutionarily related to TYMV as suggested by the analysis of their coat protein sequences and secondary structure. STNV on the other hand, is a T = 1 virus. The surprising similarity in the three dimensional structure of STNV is probably a result of divergent evolution of their protein coats.

CMV and BMV are icosahedral viruses with T = 3 particle morphology. AMV is bacillus shaped and consists of particles of different lengths. Despite these apparent differences in the particle morphology of these viruses, they have a similar genomic organization and share many common biophysical properties. Analysis of their genomic sequences indicates that they are evolutionarily related. It appears likely that all plant viruses with a tripartite genome might have resulted from a common ancestor. This ancestral virus might have been a tripartite virus or might have been a virus with a single stranded genome which has subsequently evolved into a tripartite virus. Recently, Haseloff *et al.* (1984) have demonstrated striking similarity between AMV RNA1 and 2 and TMV RNA. These results probably reflect the common evolutionary ancestry of viruses with unsegmented and divided genomes. It is also possible that similar genes from host organisms might have been incorporated into different viral genes resulting in the observed similarities (Haseloff *et al.*, 1984). The mechanism of RNA encapsidation within the protein shells of these viruses of different shapes and structures are strikingly dissimilar. The observed homology in the sequences of non-structural proteins and lack of homology in the structural proteins suggest that the mechanism of self-assembly of simple viruses might evolve in a variety of different modes.

Acknowledgements

We thank Dr. Finklestein for predicting the secondary structure of CMV and BMV coat proteins. This work was supported by Department of Science and Technology, New Delhi.

References

- Abad-Zapatero, C, Abdel-Meguid, S. S., Johnson, J. E., Leslie, A. G. W., Rayment, I., Rossmann, M. G., Suck, D. and Tsukihara, T. (1980) *Nature (London)*, **286**, 33.
- Ahlquist, P., Luckow, V. and Kaesberg, P. (1981) *J. Mol. Biol.* **153**, 23.
- Argos, P. (1981) *Virology*, **110**, 55.
- Barker, R. F., Javris, N. P., Thompson, D. V., Loesch-Fries, L. S. and Hall, T. C. (1983) *Nucleic Acids Res.*, **11**, 2881.
- Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. and Klug, A. (1978) *Nature (London)*, **276**, 362.
- Casper, D. L. D. and Klug, A. (1962) *Cold spring Harbor Symposia Quant. Biol.*, **27**, 1.
- Dayhoff (1972) *Atlas of protein sequence and structure*, Natl. Biom. Res. Foundation, Washington D.C.
- Doolittle, R. F. (1981) *Science* **214**, 149.
- Golet, P., Lomonossoff, G. P., Butler, P. J. G., Akam, M. E., Gait, M. J. and Karn, J. (1982) *Proc. Natl. Acad. Sci., USA*, **79**, 5818.
- Gould, A. R. and Symons, R. H. (1982) *Eur. J. Biochem.*, **126**, 211.
- Hopper, P., Harrison, S. C., Saucer, R. T. (1984) *J. Mol. Biol.*, **177**, 701.
- Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. and Bricogne, G. (1978) *Nature (London)*, **276**, 368.
- Haseloff, J., Golet, P., Zimmern, D., Ahlquist, P., Dasgupta, R. and Kaesberg, P. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4358.
- Hermodson, M. A., Abad-Zapatero, C, Abdel-Meguid, S. S., Pundak, S., Rossmann, M. G. and Tremaine, J. H. (1982) *Virology*, **119**, 133.
- Kaper, J. M. (1975) The chemical basis of virus structure, Dissociation and Reassembly in *Frontiers Biology* (eds A. Neuberger and E. L. Tutum) (North Holland/American Elsevier, Amsterdam) Vol. 39.
- Kaper, J. M. and Waterworth, H. E. (1981) in *Handbook of Plant Virus Infections and Comparative Diagnosis* (ed. E. Kurstak) (Amsterdam: Elsevier/North Holland Biomedical Press) p. 258.
- Kurstak, E. (1981) in *Handbook of Plant Virus Infections and Comparative Diagnosis* (North-Holland, Amsterdam, New York: Elsevier)
- Liljas, L., Unge, T., Jones, T. A., Fridborg, K., Lovgren, S., Scogland, V. and Strandberg, B. (1982) *J. Mol. Biol.*, **159**, 93.
- Murthy, M. R. N. (1983) *J. Mol. Biol.*, **68**, 469.
- Needleman, S. B. and Wunch, C. D. (1970) *J. Mol. Biol.*, **48**, 443.
- Peter, R., Stehelin, D., Reinbott, J., Collot, D. and Duranton, H. (1972) *Virology*, **49**, 615.
- Rossmann, M. G., Abad-Zapatero, C, Murthy, M. R. N., Liljas, L., Jones, T. A. and Strandberg, B. (1983) *J. Mol. Biol.*, **165**, 711.
- Rossmann, M. G. and Argos, P. (1975) *J. Biol. Chem.*, **250**, 7525.
- Savithri, H. S. and Murthy, M. R. N. (1983) *J. Biosci.*, **5**, 183.
- Schwinghamer M. W. and Symons, R. H. (1977) *Virology*, **79**, 88.
- Shih, D. S. and Kaesberg, P. (1973) *Proc. Natl. Acad. Sci. USA*, **70**, 1799.
- Verduin, B. J. M. (1978) Ph. D. Thesis, Agricultural University of Wageningen, The Netherlands.