Ann. Rev. Phytopathol. 1986. 24:289–310

MOLECULAR EVOLUTION OF PLANT RNA VIRUSES

R. W. Goldbach

Department of Molecular Biology, Agricultural University, De Dreijen 11, 6703 BC Wageningen, The Netherlands

INTRODUCTION

Viruses with single-stranded-RNA genomes and parasitic on eukaryotic cells form a very diverse group with a wide variation in genome structure and expression (56). This is particularly true for plant RNA viruses, which may have either segmented or nonsegmented genomes that are expressed using a variety of translation strategies (12, 14, 26). Based on these differences, and on other characteristics like host range, vector specificity, and particle morphology, the plant RNA viruses have been classified into a rather large number of groups. The most important of these are summarized in Table 1.

During the last few years the complete nucleotide sequences of the genome of AlMV[1] (6, 10, 11), BMV (2, 3), CMV (50, 51), TMV (18), and CPMV (35, 61) have been elucidated. This has created the possibility of determining possible relationships between these and other eukaryotic RNA viruses in terms of RNA and protein sequence. Such studies have revealed striking sequence homologies among nonstructural proteins of plant and animal RNA viruses, which must reflect (until recently unsuspected) evolutionary relationships. In this review I summarize the main conclusions from these protein-sequence comparisons, which throw a totally new light on the evolution and systematics of RNA viruses. In addition, I discuss the possible evolutionary mechanisms that may underly the interviral protein homologies detected so far.

[1]See Table 1 for abbreviations of virus names.

0066-4286/86/0901-0289$02.00

Table 1 Genome types of plant positive-strand RNA viruses

Group	Type Virus	Abbreviation	Particle structure	RNA structure 5' end	RNA structure 3' end
Monopartite genome					
Tymoviruses	Turnip Yellow Mosaic	TYMV	isometric	cap	tRNA(val)
Sobemoviruses	Southern Bean Mosaic	SBMV	isometric	VPg	X_{OH}
Tombusviruses	Tomato Bushy Stunt	TBSV	isometric	cap	X_{OH}
Tobacco Necrosis Virus	Tobacco Necrosis	TNV	isometric	ppA	?
Luteoviruses	Barley Yellow Dwarf	BYDV	isometric	VPg	X_{OH}
Tobamoviruses	Tobacco Mosaic	TMV	rod	cap	tRNA(his)
Potexviruses	Potato Virus X	PVX	rod	cap	poly A (?)
Potyviruses	Potato Virus Y	PVY	rod	VPg	poly A
Closteroviruses	Sugar Beet Yellows	SBYV	rod	?	?
Bipartite genome					
Comoviruses	Cowpea Mosaic	CPMV	isometric	VPg	poly A
Nepoviruses	Tobacco Ringspot	TRSV	isometric	VPg	poly A
Tobraviruses	Tobacco Rattle	TRV	rod	cap	X_{OH}
Dianthoviruses	Carnation Ringspot	CRSV	isometric	?	?
Pea Enation Mosaic Virus	Pea Enation Mosaic	PEMV	isometric	VPg	X_{OH}
Furoviruses	Beet Necrotic Yellow Vein	BNYVV	rod	cap	poly A
Tripartite genome					
Alfalfa Mosaic Virus	Alfalfa Mosaic	AlMV	bacilliform	cap	X_{OH}
Ilarviruses	Tobacco Streak	TSV	quasi-isometric	cap	X_{OH}
Bromoviruses	Brome Mosaic	BMV	isometric	cap	tRNA(tyr)
Cucumoviruses	Cucumber Mosaic	CMV	isometric	cap	tRNA(tyr)
Hordeiviruses	Barley Stripe Mosaic	BSMV	rod	cap	tRNA(tyr)
Tomato Spotted Wilt Virus	Tomato Spotted Wilt	TSWV	isometric	?	X_{OH}

TRANSLATION STRATEGIES OF PLANT RNA VIRUSES

As mentioned above, the apparent heterogeneity among plant RNA viruses is mostly caused by differences in genome structure and translation strategies. Before discussing possible relationships between viruses that are very different in these respects, I will briefly describe the major types of translation mechanisms found.

The various translation strategies developed by eukaryotic RNA (ribonucleic acid) viruses are all directed to enable monocistronic translation. This is necessary since eukaryotic ribosomes usually start translation only at the first AUG[2] initiation codon from the 5'-terminus of a messenger RNA (mRNA) (31–33). Hence, most plant RNA viruses employ one (or a combination) of the following translation strategies to express downstream cistrons:

1. *Segmentation of the genome*. First of all the genome of RNA viruses may be divided into more than a single RNA segment. By this segmentation the number of downstream cistrons is often minimized, and most of the genomic RNA segments contain only a single cistron (e.g. AlMV and BMV; see Figure 1).
2. *Translation from subgenomic mRNAs*. The virus may produce subgenomic mRNAs to express downstream cistrons [e.g. AlMV and BMV (RNA3), and TMV; see Figures 1 and 2].
3. *Read-through translation*. The first cistron in the viral RNA may be followed by a "leaky" termination codon that can be suppressed by a host-encoded transfer RNA (tRNA), thereby permitting some of the ribosomes to read-through into a second cistron (e.g. TMV; see Figure 2).
4. *Polyprotein processing*. The genomic RNA may be a monocistronic mRNA coding for a so-called "polyprotein" from which functional, smaller proteins are generated by proteolytic cleavages (e.g. CPMV; see Figure 3).

Despite differences in genome structure and translation strategy, TMV is similar to AlMV, BMV, and CMV in containing four cistrons, its two most downstream cistrons encoding, moreover, similar-sized proteins of ~ 35 Kilodalton (K) and ~ 20K (i.e. the coat protein), respectively (cf Figures 1 and 2). CPMV, on the other hand, is very dissimilar, resembling in genome structure and expression the animal picornaviruses, like poliovirus and foot-and-mouth-disease virus (FMDV).

[2]Many terms are defined in the glossary at the end of the text.

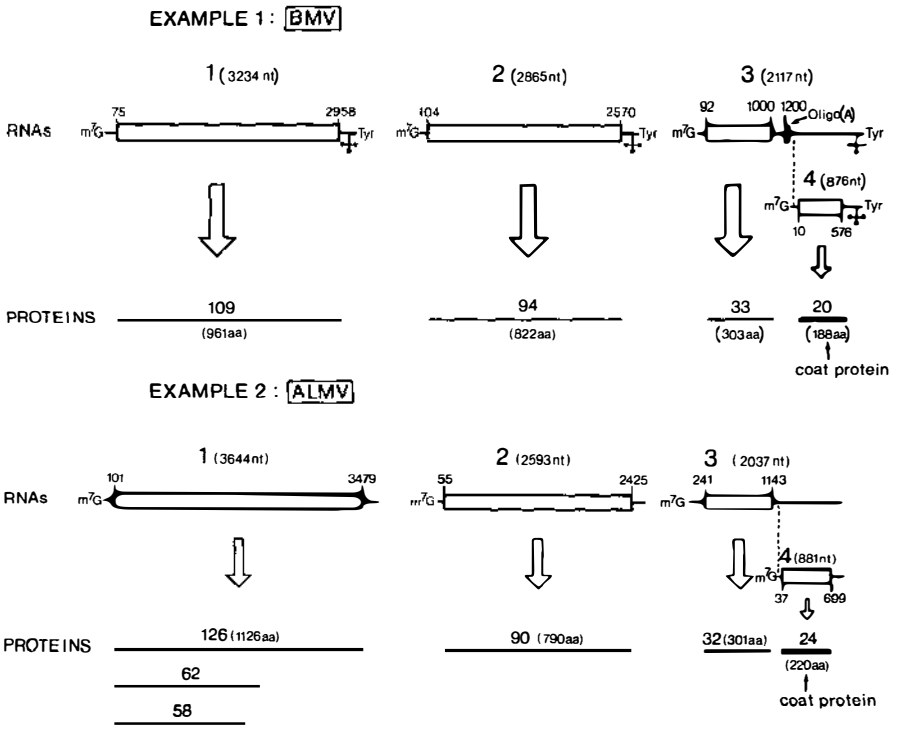

Figure 1 Genomic organization and translation strategy of plant viruses with a tripartite RNA genome (examples: AlMV and BMV). Open-reading frames in the RNAs are represented as open bars, with the nucleotide positions of start and stop codons indicated, and proteins are represented as single lines. Numbers on top of each protein indicate the molecular weight ($\times 10^{-3}$); numbers below each protein indicate the number of amino acid residues. RNA1 and RNA2 both function as monocistronic mRNA. RNA3 contains two cistrons, one encoding a ~ 30K nonstructural protein and the other encoding the ~ 20K coat protein. The coat-protein cistron is only translated from a subgenomic mRNA, RNA4.

HOMOLOGY BETWEEN CPMV AND THE PICORNAVIRUSES

Similarities in Protein Capsid, Genome Structure, and Genome Expression

Although comoviruses have a divided genome, consisting of a larger (B) and a smaller (M) RNA segment, and picornaviruses do not, members of both groups are very similar in genomic organization and capsid architecture (19, 20):

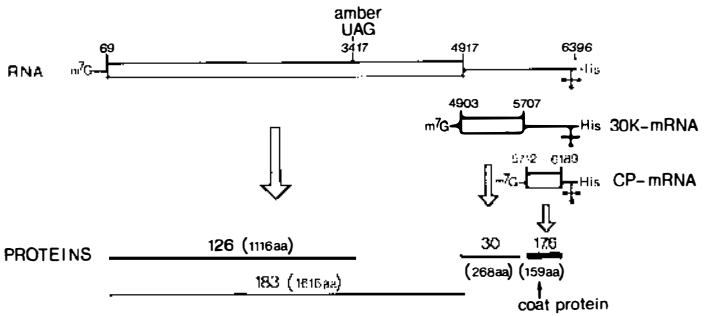

Figure 2 Genomic organization and translation strategy of TMV. For explanation of symbols see legend of Figure 1. The 183K protein is an extension of the 126K protein and is produced by read-through at the amber stop codon of position 3417. This stop codon can be partially suppressed by a host-encoded tRNA. From the 3'-terminal part of the genomic RNA, two subgenomic mRNAs are derived that code for a 30K protein and the coat protein, respectively.

1. *Genome structure*. Members of both groups have positive-stranded-RNA genomes supplied with a 5'-terminal protein (denoted VPg: *V*iral *P*rotein *g*enome-linked) and a 3'-poly(A) tail.
2. *Genome expression*. Their RNAs are expressed by the production of polyproteins from which the mature, functional proteins are released by proteolytic cleavages.
3. *Capsid morphology*. Both picornaviruses and comoviruses have icosahedral capsids composed of more than a single species of coat protein that are arranged in a pseudo T = 3 lattice. The full three-dimensional structure of two different picornaviruses, human rhinovirus HRV14 (52) and poliovirus (24), has recently been established by computerized X-ray crystallography. This shows that the three large coat proteins VP1, VP2, and VP3 of these viruses are folded in very similar edge-shaped β-barrel structures. Additionally, each of these proteins has a long amino-terminal extension—VP4 representing in fact the amino-terminal extension of VP2. Recent X-ray diffraction studies performed by J. E. Johnson and colleagues (personal communication) have revealed that the two coat proteins VP37 and VP23 of CPMV (type member of the comoviruses) together form three β-barrel domains that clearly correspond to the β-barrel cores of the picornaviral coat proteins.

Homology Among Nonstructural Proteins

While the comparison made in the preceding paragraph indicates that comoviruses and picornaviruses are similar in capsid architecture and genome structure and expression, the finding that these viruses encode proteins with similar functions and exhibiting extensive amino acid sequence homology

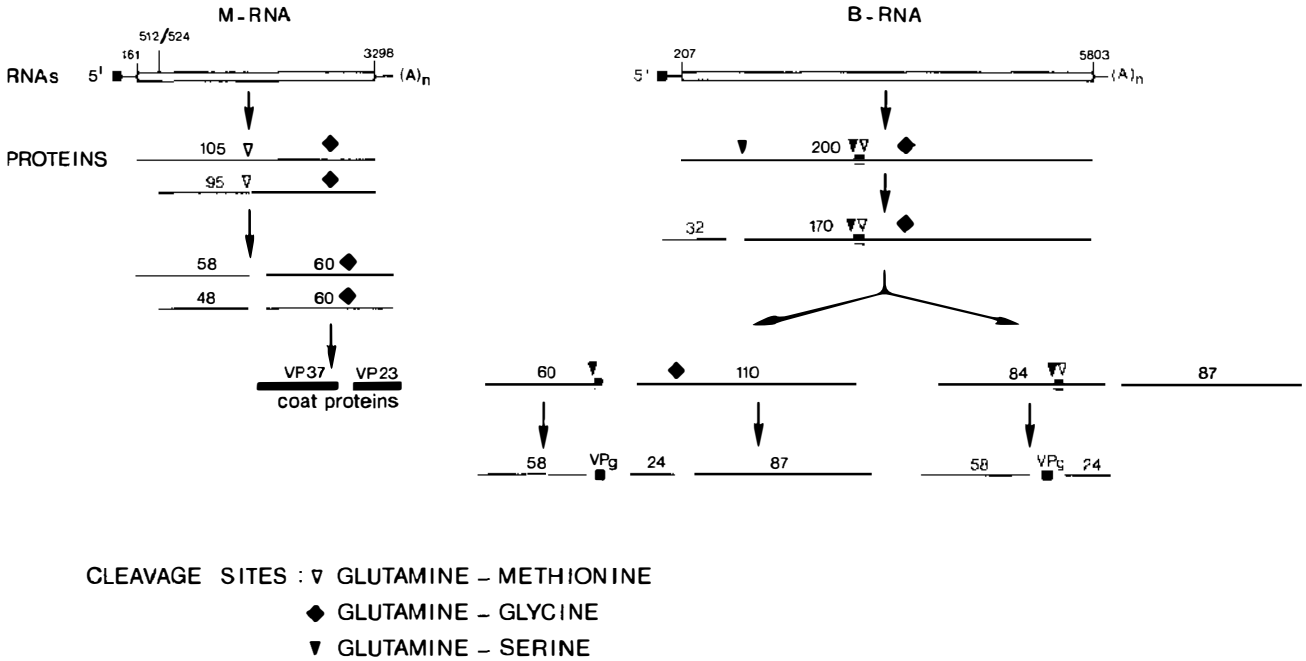

Figure 3 Genomic organization and translation strategy of CPMV. The black square at the 5'-termini of the RNAs indicates VPg; for other symbols see legend of Figure 1. M-RNA encodes two overlapping polyproteins of 105K and 95K; B-RNA encodes a polyprotein of 200K. Proteolytic cleavages occur at glutamine-methionine, glutamine-glycine, and glutamine-serine dipeptide sequences. Functions of cleavage products are indicated in Figure 4.

(16) indicates that the similarities between the two virus groups probably are not accidental but reflect an evolutionary relationship. In Figure 4 the polyproteins of CPMV and poliovirus are compared, with the regions of amino acid sequence homology (as found by computerized alignment) indicated by cross-hatching. Strikingly, when placing M-RNA to the left of B-RNA, the general gene order in the (split) CPMV genome is similar to that of the (unsplit) poliovirus genome. In particular, the genetic organization of the B-RNA, encoding two proteases, VPg and the core polypeptide of the viral replicase (i.e. the polypeptide in which the RNA-synthesizing activity resides), is amazingly colinear with that of the P2 and P3 regions of poliovirus. Whereas the coat proteins of CPMV do not show any sequence homology to the picornaviral proteins (but do show structural homology), four of the five proteins generated from the B-RNA-coded 200K polyprotein exhibit significant (20–30%) sequence homology to picornaviral counterparts (see Figure 4). These proteins are the 58K protein (homologous to picornaviral protein 2C, formerly called X), VPg, the 24K protease (homologous to protease 3C), and the 87K protein (homologous to the picornaviral polymerase 3D). In addition, the 32K protease of CPMV (17) and protease 2A of poliovirus (58) are encoded in similar regions of the respective viral genomes (Figure 4). All these striking analogies together provide strong evidence that como- and picornaviruses are somehow evolutionarily related. For further discussion on this point see below.

HOMOLOGY BETWEEN AlMV, BMV, CMV, TMV, AND SINDBIS VIRUS

Conserved Sequences in Nonstructural Proteins of AlMV, BMV, CMV, and TMV

Another striking example of interviral relationship has been found for TMV, AlMV, BMV, and CMV on one hand and the animal Sindbis virus (alphaviruses) on the other (4, 9, 23, 50, 51).

Let us first consider the homology between the four plant viruses. AlMV, BMV, and CMV, though biologically distinct and members of different groups (see Table 1), have a similar genomic structure and translation strategy. The genome of these viruses is tripartite; each of the two largest RNAs, RNA1 and RNA2, encodes a single protein of ~ 120K and ~ 95K, respectively, while RNA3 is dicistronic. Only the 5'-proximal cistron is translated directly from RNA3, giving a protein of ~ 35K (for brevity denoted protein 3a here). The second cistron in RNA3, encoding the coat protein of ~ 20–25K, is only translated from a subgenomic mRNA, RNA4 (see Figure 1). While the amino acid sequences of the coat proteins of AlMV, BMV, and CMV show little or no homology (4, 43), the 3a proteins of these viruses, in

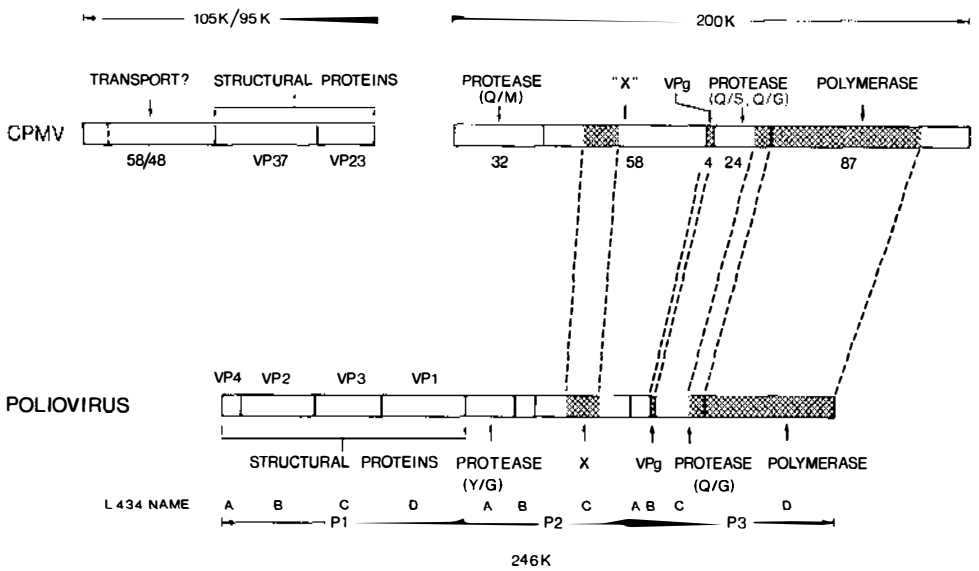

Figure 4 Comparison of the polyproteins encoded by CPMV (above) and poliovirus (below). The (proposed) functions of the cleavage products derived from these polyproteins are indicated. Regions of significant (> 20%) sequence homology among CPMV and polio proteins are cross-hatched. Reprinted with permission from Franssen et al (16), with modifications.

contrast, exhibit significant sequence homology. The most extensive homology is found between the BMV and CMV proteins (~ 34%) and the least homology between the AlMV and CMV protein (~ 14%) (43, 53). In addition, comparison of the RNA1 and RNA2 translation products of these three viruses and comparison with the 126K (p126) and 183K (p183) proteins of TMV (cf Figure 2) revealed additional sequence homology (Figure 5; see also 9, 23, 50, 51). The RNA1 products of AlMV, BMV, and CMV, and p126 of TMV, all contain two distinct regions ("domains") of conserved sequences (23–28% homology) separated by a less conserved sequence near the middle of each protein, while the RNA2 products of the tripartite genome viruses contain a third conserved block with sequence homology to the read-through portion of p183 of TMV. Homology between the TMV 30K protein and any of the 3a proteins of the tripartite-genome viruses was not detected (4).

Homology to Sindbis Virus Proteins

Strikingly, further (computerized) comparisons with proteins of other viruses revealed that the three conserved domains present in the two largest nonstructural proteins of the tripartite-genome viruses and of TMV can be aligned with sequences in three proteins specified by Sindbis virus, an animal alpha-

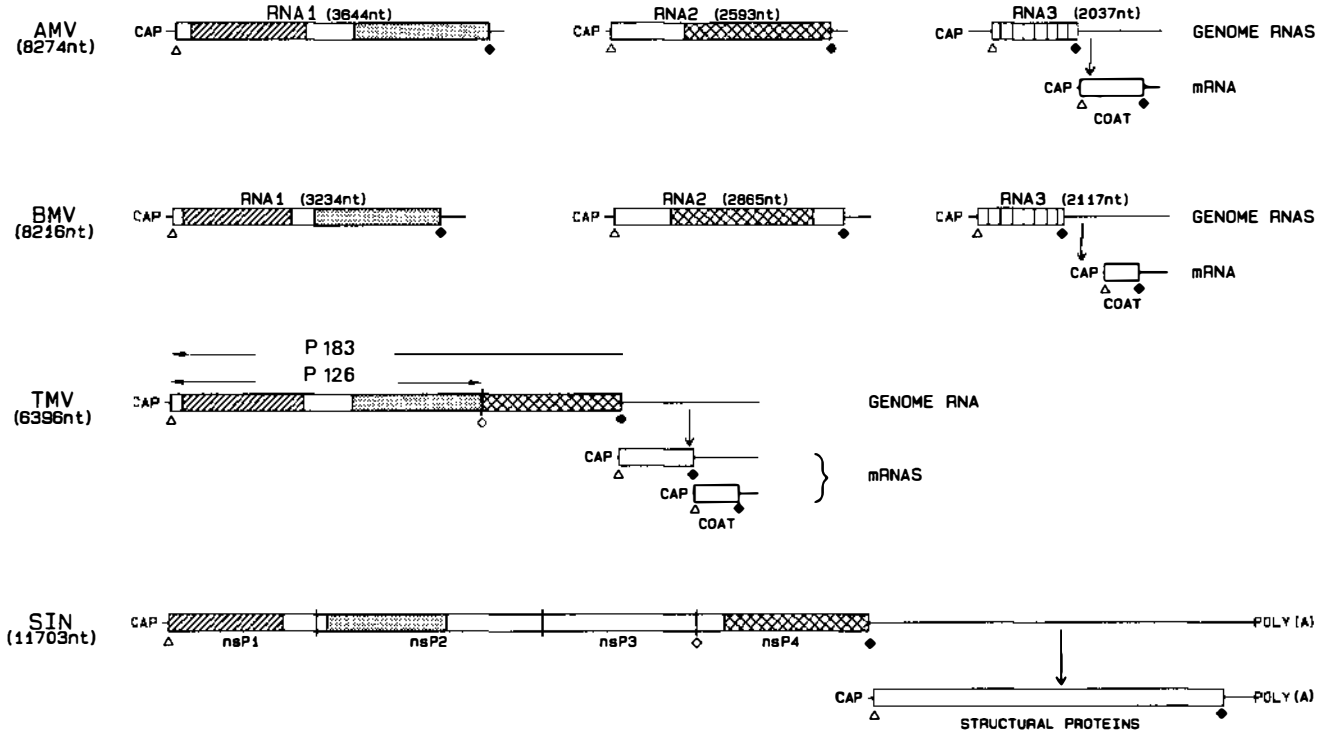

Figure 5 Comparison of the genomic RNAs of AlMV, BMV, TMV, and Sindbis virus (SIN). Open-reading frames are represented as open bars and regions of amino acid sequence homology are indicated by similar shading.

Symbols: Triangle, start codons; solid diamond, stop codons; and open diamond, suppressable stop codons (in TMV and Sindbis virus). Reprinted with permission from Ahlquist et al (4).

virus (Figure 5; 4, 23, 27). Sindbis virus has a positive-stranded-RNA genome, the complete sequence of which (11.7 kilobases) has been determined (54, 55). The genomic RNA encodes two large polyproteins, a nonstructural polyprotein (p270, translated directly from the genomic RNA) and a structural polyprotein (p130, translated from a subgenomic RNA). Both polyproteins are processed proteolytically; p270 is cleaved into four proteins (denoted by nsP1–nsP4), one of which, nsP4, is produced by translational read-through at an opal termination codon similar to that found for TMV (see Figure 5). Haseloff et al (23) found that the read-through product nsP4 is homologous to the RNA2 products of AlMV and BMV and to the read-through part of TMV p183, with sequence homology ranging between 18 and 20% (Figure 5). Additionally, Ahlquist et al (4) established that nsP1 and nsP2 are homologous to the RNA1 products of AlMV and BMV and to p126 of TMV, with sequence homology ranging between 9 and 19%. No homologous counterpart for nsP3 could be detected among the various plant viral protein sequences, nor could homology among the structural proteins of the viruses considered be detected (4).

For CMV the situation is similar as described above for AlMV and BMV, with the translation products of RNA1 and RNA2 containing the same three conserved domains (50, 51).

HOMOLOGY AMONG PROTEINS OF OTHER PLANT AND ANIMAL VIRUSES

The nucleotide sequence of the RNA genome of carnation mottle virus (CarMV) a member of the tombusvirus group (Table 1), reveals that this virus is also related to TMV, the tripartite viruses, and Sindbis virus (22). The gene organization of CarMV is most similar to that of TMV. Its nondivided genomic RNA encodes a nonstructural protein of 27K (p27) that can be elongated by read-through at an amber termination codon to a protein of 86K (p86), and possibly by double read-through to a protein of 98K (p98). Two subgenomic mRNAs, derived from the 3'-terminal part of the genomic RNA, contain cistrons for a 7K protein and the 38K coat protein. Protein-sequence comparisons show that p27 is homologous to the conserved domain in the carboxy-terminus of both TMV p126 and the RNA1 product of the tripartite viruses, and to Sindbis nsP2. In addition, there is some homology between the read-through part of CarMV p86 and the conserved domain in the read-through part of TMV p183, the RNA2 products of the tripartite viruses, and nsP4 of Sindbis (22). Strikingly, the third domain conserved among TMV (N-terminus p126), the tripartite viruses (N-terminus RNA1 product), and Sindbis virus (nsP1) is lacking in the CarMV proteins.

Likewise, determination of part of the nucleotide sequence of TRV RNA1

(SYM strain) (7) shows that this virus, though bipartite, is also related to TMV. Genomic RNA1 of TRV encodes a 140K protein that can be elongated by read-through, similar to that found for TMV, to a 170K protein. Two additional downstream cistrons in RNA1, encoding proteins of 29K and 16K, are translated only from subgenomic mRNAs. The read-through part of the 170K protein and the 29K protein clearly shows sequence homology to the read-through part of TMVp183 and to p30, respectively, while the 16K protein appears to be unique for TRV.

In addition to the comparisons presented above, there is another example of homology among proteins of plant and animal viruses. This concerns the reverse transcriptases of cauliflower mosaic virus (CaMV), hepatitis B virus, and retroviruses, which also show significant sequence homology (57, 60). Because these viruses utilize reverse transcription for their genome replication, the evolutionary mechanism underlying their relationship may be very different from that followed by the true RNA viruses. Therefore, these viruses are not further discussed here.

FUNCTIONS OF THE CONSERVED VIRAL PROTEINS

From the above it is obvious that homology may exist between nonstructural proteins of plant and animal RNA viruses previously considered as not, or only distantly, related. As far as we know, this homology primarily concerns proteins that are somehow involved in replication of viral RNA.

This holds completely for the (four) proteins conserved among CPMV and the picornaviruses, with the CPMV 87K protein representing the viral core polymerase, VPg most likely functioning as a primer during initiation of replication, and the 24K protease releasing this primer from the 60K protein (see Figure 3). The latter protein, moreover, having a size of 58K after release of VPg, anchors the replication complex (i.e. the complex of replication proteins and template RNA) to membranes (see references 19 and 21 for detailed discussion). Furthermore, recent analysis of guanidine-resistant and -dependent isolates of poliovirus revealed that all these viruses contain a mutation in the central domain (of $\sim$ 140 amino acid residues) of protein 2C (47), which is strongly conserved among picornaviruses (15) and homologous to the central region of the CPMV 58K protein (Figure 4). Since the major effect of guanidine appears to be blockage of viral RNA synthesis, this observation provides additional evidence that the 58K protein of CPMV and protein 2C of the picornaviruses play a role in this process.

AlMV and BMV protoplast studies indicate that the conserved RNA1 and RNA2 translation products also play an essential part in RNA replication. In isolated protoplasts, RNA1 and RNA2 have been shown to replicate efficiently in the absence of RNA3 (28, 45). Proteins p126 and/or p183 have been

suggested as participants in viral RNA synthesis for TMV, but direct indications are still missing. If TMV encodes its own polymerase, like CPMV does for instance, then p126 and p183 are the only two candidates for this activity, since p30 is certainly too small for this function.

Detailed computer-assisted sequence alignment has revealed that the homologous RNA2 products of both AlMV and BMV, the read-through portion of TMV p183, and nsP4 of Sindbis virus all share two small, conserved blocks that are also found in the RNA-dependent RNA polymerases of CPMV and the picornaviruses (27). This finding suggests that all these viral products represent core polymerase molecules. The consensus sequence of the first conserved block is $\overset{S}{G}TXXXTXXXN\overset{T}{S}$, where X may be any amino acid residue. The second block, 21–37 residues downstream from this first conserved sequence, is even more prominent. It contains a fourteen-residue segment consisting of a gly-asp-asp (GDD) sequence flanked by hydrophobic residues at both sites (27). Additional sequence analyses have shown that these two conserved sequences also exist in the read-through parts of p170 of TRV (7) and p86 of CarMV (22), respectively. They are also present in the protein encoded upstream from the capsid protein cistron in tobacco etch virus (TEV, a potyvirus) RNA (5), as well as in the RNA1 product of beet necrotic yellow vein virus (a furovirus) (H. Guilley, personal communication). Hence these proteins, or protein domains, can also tentatively be regarded as viral core polymerases.

It should be noted that the 3a proteins of the tripartite viruses and p30 of TMV are less conserved proteins, for which no counterpart has been found among the Sindbis nonstructural proteins. While the function of the 3a proteins is still a matter of speculation, an accumulating amount of evidence indicates that p30 of TMV has a transport function and is involved in cell-to-cell movement of viral RNA (34, 46, 63). This protein is, however, not homologous to the 3a proteins of AlMV, BMV, and CMV, and its sequence can even vary remarkably among different TMV strains (38, 62). Yet, the 29K protein encoded by TRV RNA1 shows significant sequence homology to TMV p30 (7), which suggests that this TRV product is also involved in viral RNA transport. This fits nicely with the observation that TRV RNA1 can spread systemically through host plants in the absence of RNA2.

EVOLUTIONARY IMPLICATIONS OF THE INTERVIRAL HOMOLOGIES

Possible Evolutionary Mechanisms

The discovery that plant and animal viruses, previously regarded as evolutionarily unrelated, may encode proteins with similar amino acid sequences is exciting and throws a new light on virus evolution. How can viruses, which

infect organisms from different eukaryotic kingdoms that diverged 1.2×10^9 years ago, encode proteins with homologous sequences and, particularly in the case of como- and picornaviruses, how can such viruses have similar genetic arrangements?

Basically, three different evolutionary pathways have been proposed to account for the interviral relationships that have been detected thus far:

1. *Common ancestry*. The homologies observed may reflect a common evolutionary origin of the viruses. This would imply that plant and animal viruses (e.g. como- and picornaviruses), though nowadays ecologically separated by very different host ranges, have diverged from a common viral ancestor.
2. *Convergent evolution*. Alternatively, viruses of different origins may encode proteins with similar functions (for instance in RNA replication) and interact with the same highly conserved host proteins, and solely for this reason they may have evolved similar tertiary, and hence primary, structures.
3. *Transduction of host genes*. Similar viruses may have independently evolved from their host cells by adopting the same conserved genes from their host's chromosomes to employ for their own replication. The similarities in protein sequences observed would then be the result of a common gene transfer mechanism and a strong conservation of the genes so captured.

Both the second and third possibilities seem difficult to imagine since they do not explain the colinearities in the genetic maps of the various viruses considered (cf CPMV and poliovirus in Figure 4 and TMV and Sindbis virus in Figure 5). Therefore, I favor the hypothesis that the relationships observed are all based on the first pathway suggested above—i.e. common ancestors that underwent divergent evolution. Indeed, the information now available leaves little doubt that related groups of plant and animal viruses have a common evolutionary origin. As discussed in the following section, this is most evident for the como- and picornaviruses.

Plant Plus-Strand RNA Viruses May Be Classified into Two Super Groups

Obviously, the plant RNA viruses whose genome sequences have been determined so far can be divided into two sets, one containing those viruses related to Sindbis virus (i.e. AlMV, BMV, CMV, CarMV, TMV, and TRV), and one containing viruses related to the picornaviruses (i.e. CPMV).

Let us first consider the evolutionary relationship between CPMV and the picornaviruses. As discussed above, these viruses are similar in capsid structure, genome structure (with VPg and poly(A) tail), and genome expression (polyprotein processing). Also, both produce functionally equivalent proteins

that exhibit sequence homology. Moreover, if one places M-RNA to the left of B-RNA, the resulting genetic map is similar to that of picornaviruses. In fact, the only major difference seems to be the genome partition in the case of CPMV, which makes this virus a "split" picornavirus. Hence, in my opinion plant comoviruses and animal picornaviruses are derived from a common ancestral virus. Whether the comoviral genome became bipartite or, alternatively, the picornaviral genome became monopartite, I discuss in the next paragraph.

It will be of great interest to extend the comparisons with the potyviruses (monopartite genome) and nepoviruses (bipartite genome) as soon as their genome sequences become known. These viruses also have genomes that are supplied with a VPg and a poly(A) tail (Table 1) and are expressed via the production of polyproteins (42, 59). Possibly they can all be classified, together with the comoviruses, in a "super group" of picornavirus-related plant viruses.

While the como-, nepo-, and potyviruses (see Table 1), and possibly additional plant virus groups, may be classified as "picorna-like" viruses, alfalfa mosaic, bromo-, cucumo-, tobra-, tombus-, and tobamoviruses (see Table 1) evidently form another cluster of evolutionarily related viruses; the "Sindbis-like" plant viruses. Clearly in this second "super group" the variation in genome structure and translation strategy is wider than that found within the picornavirus-related group. Still, common ancestry is the most likely explanation for the relationships in this group. The highest degree of similarity is found between AlMV, BMV, and CMV. Although AlMV differs from BMV and CMV both in having bacilliform instead of spherical particles and in its requirement of coat protein (or its mRNA) for infection, these three viruses are similar in genome structure (tripartite) and genetic organization, and their (three) nonstructural proteins exhibit sequence homology. Hence, these viruses have probably been derived from a common ancestor.

Although it has a nondivided genome contained in a rod-shaped particle, TMV is similar to the tripartite viruses in encoding four proteins. Moreover, the three distinct core domains conserved in the two largest proteins of the tripartite viruses are also found in proteins p126 and p183 of TMV. This, together with the similar order of cistrons encoding the conserved domains, provides strong evidence that TMV and the tripartite-genome viruses are related by common ancestry. This also holds for Sindbis virus, which encodes (in the same order) three nonstructural proteins homologous to proteins of AlMV, BMV, CMV, and TMV. In p183 of TMV, all three conserved domains are found in one polypeptide chain, while in the tripartite-genome viruses these domains are spread over two polypeptide chains. Moreover, in Sindbis virus the three domains are initially present in a single polyprotein but are all subsequently released from each other by proteolytic processing. These

differences, which reflect the various translation strategies developed during evolution, have not necessarily led to differences in function of the proteins. As noted by Ahlquist et al (4), protein domains that are physically separated in one virus (e.g. nsP1 and nsP2 of Sindbis virus) but linked in another virus (e.g. p126 of TMV and RNA1 product of AlMV, BMV, or CMV) possibly need to associate into a stable complex to become functional. This would imply that, despite the variation in genome structure and translation strategy employed, all Sindbis-related viruses end up with functionally equivalent proteins.

As with CPMV and the picornaviruses the structural proteins of the Sindbis-related viruses are not or are hardly conserved among the members of the group. Moreover, even the particle structure varies considerably, from isometric (BMV, CMV) and bacciliform (AlMV) to rod shaped (TMV), while the capsid of Sindbis virus is enveloped. This variation probably reflects the adaptation of these viruses to the special requirements of their different hosts and to different extracellular conditions. Moreover, animal viruses, but not plant viruses, have evolved under the constant pressure of the host's immune system. In view of these facts the relative heterogeneity among the structural proteins is not surprising. The intracellular conditions of eukaryotic cells will, on the other hand, change only slowly, and this may have been the reason why viral functions utilized within the cytoplasm of host cells (e.g. the viral polymerase) have been conserved so well during evolution (4, 16).

It should be mentioned that the related plant and animal RNA viruses described here do not possess completely identical sets of genes. CPMV, for instance, specifies two overlapping proteins of 58K and 48K for which no counterpart is found with poliovirus (Figure 4). Such differences, however, are also found among different members of the picornavirus group. FMDV, for instance, encodes two overlapping leader polypeptides but poliovirus does not. FMDV also differs from other picornaviruses in having three VPg genes, which probably arose by gene doubling. Similarly, the Sindbis-related plant viruses encode an ~ 30K protein that does not appear homologous to nsP3 of Sindbis virus. The extra genes coding for such unique proteins may have been introduced by recombination events. This is not unlikely, since for animal RNA viruses, recombination on the RNA level has been well documented (1, 29, 30), while Bujarski & Kaesberg have recently demonstrated that recombination occurs among the genomic RNAs of BMV (J. J. Bujarski, personal communication).

Divided vs Nondivided Genomes

Reviewing the positive-strand-RNA viruses of eukaryotes causes the question to arise why plant viruses—but not animal viruses—often have a segmented genome, each genome part being encapsidated separately. Before trying to

answer this question let us first consider the possible advantages of genome segmentation.

1. *Rapid recombination by RNA reassortment*. It is obvious that an advantage of genome segmentation may be easy interchange of genetic information upon mixed infection (or mixed transmission) of related virus strains.
2. *Increase of genome size*. Segmentation of the genome and separate encapsidation of the genome pieces allow viruses with spherical particles to enlarge their genome without encapsidation problems.
3. *Reduction of lethality*. As recently discussed by Reanney (49), segmentation reduces the target size for mutagenic agents and also decreases the lethality of progeny RNA caused by the high error level in RNA replication (RNA-dependent RNA polymerases are very error prone; see also next section).
4. *Regulation of translational expression*. Particularly for those viruses that do not produce subgenomic mRNAs, a segmented genome creates the possibility of separately regulating the expression of different groups of genes. This advantage may be illustrated with CPMV. The synthesis of a CPMV particle requires a single VPg molecule (encoded by B-RNA; see Figure 3) and 60 copies of each of the coat proteins (encoded by M-RNA). The divided genome of CPMV allows the production of coat proteins and VPg in the ratio in which these products are needed. This is not true for the nondivided picornaviral genome, which creates the consequence that 59 of each 60 VPg molecules synthesized are redundant.
5. *Facilitation of spreading through plants and/or vectors*. Separate encapsidation of smaller genome segments permits the virus to produce smaller capsids that may facilitate movements of the virus in the plant and/or the vector (8).

Contrasting to these advantages, there may be at least one disadvantage of a segmented genome, namely a reduced efficiency in host-to-host transmission. When the genome parts are separately encapsulated, two or more particles should enter a cell to initiate infection. As noticed by Nahmias & Reanney (44) and Reanney (49), this may be the clue to why plant viruses, but not animal viruses, can permit the luxury of a segmented genome. Compared to viruses infecting animals, plant viruses may have improved the fidelity of their transmission to overcome the disadvantage of multicomponent infection. Although plants cannot move, they often reach very high densities, which minimize the transmission distance. In addition, plant viruses are efficiently transmitted either by invertebrate vectors or by seed. In both cases large numbers of particles are simultaneously passed to new hosts. Since, moreover, specific ecological relationships may have developed between plants

and invertebrates, host-to-host transmission of plant viruses may indeed reach high degrees of fidelity.

In summary, the conclusion seems justified that a segmented genome has advantages over a nonsegmented genome, provided that the virus can utilize efficient transmission mechanisms. I would like to speculate, therefore, that segmented genomes are more specialized forms than nonsegmented genomes and, hence, that related plant and animal viruses have evolved from ancestor viruses with nonsegmented genomes.

Insects as Sources of Ancestral RNA Viruses

If indeed common ancestry underlies the relationships between plant and animal RNA viruses, then the question arises whether common ancestral viruses have predated the evolutionary separation of plant and animal cells that occurred approximately 1.2×10^9 years ago. As discussed by Franssen et al (16) and Ahlquist et al (4), the conservation of protein sequences would then reflect extreme selective pressures during evolution. This, however, appears to be in contradiction with the high mutation frequencies generally observed for viral RNA genomes (13, 25, 48). It is believed that these high mutation frequencies (ranging between 10^{-3} and 10^{-4} per nucleotide per doubling) are (at least partly) due to the lack of proofreading enzymes during RNA replication (see 49 for a review). As a result, the rate of viral RNA evolution can be more than a millionfold greater than the rate of chromosomal DNA evolution of their hosts (25).

As proposed in a previous paper (16), it seems therefore more likely that common ancestors of plant and animal viruses existed in much more recent times. If this is true then there are a few strong arguments to support the contention that these ancestors were insect viruses:

1. *The host ranges of plant and animal viruses overlap in insects.* Many plant RNA viruses are transmitted by insects, and for a number of them (e.g. plant reoviruses and rhabdoviruses) it has been established that they are able to replicate in their insect vector (reviewed in 37). Additionally, some groups of insect RNA viruses are known to be capable of infecting mammalian cells (e.g. nodamura viruses).
2. *Insects are hosts for a considerable number of RNA viruses.* Hence, insects are indeed a potential source of RNA viruses. The small RNA viruses of insects appear to have a wide diversity of capsid morphology and genome size (36, 41). Among these viruses at least two have been definitely classified as picornaviruses, i.e. *Drosophila* C virus and Cricket paralysis virus (for review see 39, 40), while other viruses, like *Flaccherie* virus, *Rhopalosiphum padi* virus, and Tussock moth virus probably belong

to this group as well (N. F. Moore, personal communication). In view of argument 1, the plant comoviruses and the picornaviruses of vertebrates may originate from this group of insect viruses.

3. *Some positive-strand-RNA viruses of insects have a divided genome.* An almost fundamental difference between positive-strand viruses of plants and animals seems to be that plant viruses very often have divided genomes (see Table 1) and animal viruses do not. Among the animal viruses there is at least one exception, however, namely the nodamura-virus group (e.g. nodamura virus and black beetle virus) that is characterized by a bipartite single-stranded-RNA genome (39, 40). Strikingly, these viruses have insects as hosts!

In summary, I feel that insects have played a crucial role in virus evolution. In my opinion there are two attractive evolutionary pathways that can explain the observed similarities between RNA viruses of plants and higher animals. Either these viruses have all descended from insect viruses or, alternatively, insects have functioned as intermediate hosts upon the transfer and introduction of plant viruses into animals or vice versa. Hopefully further studies on insect viruses will tell us more about the phylogeny of eukaryotic RNA viruses.

ACKNOWLEDGMENTS

I thank Ab van Kammen and Joan Wellink for their critical readings of the text and Paul Ahlquist, Jozef Bujarski, Norman Moore, and David Zimmern for providing me with useful information and manuscripts. I would also like to thank Gré Heitkönig for her excellent typing.

GLOSSARY

Amber termination codon: See termination codons.

B-RNA: The largest of the two RNAs of comoviruses.

Capsid: The closed protein shell or tube of a virus.

Cistron: A region coding for one polypeptide.

Comoviruses: Group of plant viruses having a bipartite RNA genome, separately packaged in isometric protein capsids.

Conserved domains, conserved sequences: Distinct regions in different proteins that exhibit amino acid sequence homology as a result of (strong) conservation of these sequences during divergent evolution.

Downstream cistrons: Cistrons located at the 3'-side (for ribosomes, "downstream") of the 5'-proximal (for ribosomes, "first") cistron.

Encapsidation: The packaging of viral nucleic acid into protein shells or tubes.

Icosahedral capsid: Capsid with five-fold, three-fold, and two-fold rotational symmetry.

Initiation codons: Fist codon (triplet) of a protein-coding region in RNA. This codon has the base-sequence AUG and codes for methionine.

Kilodalton (K): Thousand dalton (mass).

Leaky termination codon: A termination codon that can be recognized (and hence suppressed) by a tRNA, leading to "read-through" of the ribosomes into a second cistron. As a result, a "read-through" protein is produced.

M-RNA: The smallest of the two RNAs of comoviruses.

Opal termination codon: See termination codons.

Picornaviruses: A group of isometric viruses, infecting animals, and having a small ("pico") RNA genome.

Polyprotein: A (large) primary translation product from which mature, functional proteins are generated by proteolytic cleavages.

Positive-stranded RNA: RNA with plus-polarity, i.e. upon entrance in a host cell directly translatable by ribosomes and infectious.

Primary structure: The sequence of amino acids in a polypeptide chain; the sequence of nucleotides in a nucleic acid.

5'-Proximal cistron: Cistron most proximal to the 5'-terminus of the RNA. Since ribosomes start translation from the 5'-terminus, this is the first cistron to be translated.

Read-through product: See leaky termination codon.

Termination codons: Codons (triplets) that do not code for any amino acid but follow a protein-coding region and serve as stop signals for protein synthesis. The three termination codons are UAG (amber), UGA (opal), and UAA (ochre).

Tertiary structure of a protein: The folding and coiling of a polypeptide chain to form a globular protein.

*VPg: V*iral *P*rotein *g*enome-linked; i.e. a (small) protein covalently bound to the 5'-end of some viral RNA genomes.

Literature Cited

1. Agol, V. I., Grachev, V. P., Drozdov, S. G., Kolesnikova, M. S., Kozlov, V. G., et al. 1984. Construction and properties of intertypic poliovirus recombinants: first approximation mapping of the major determinants of neurovirulence. *Virology* 136:41–55
2. Ahlquist, P., Dasgupta, R., Kaesberg, P. 1984. Nucleotide sequence of the brome mosaic virus genome and its implications for viral replication. *J. Mol. Biol.* 172:369–83
3. Ahlquist, P., Luckow, V., Kaesberg, P. 1981. Complete nucleotide sequence of brome mosaic virus RNA3. *J. Mol. Biol.* 153:23–38
4. Ahlquist, P., Strauss, E. G., Rice, C. M., Straus, J. H., Haseloff, J., Zimmern, D. 1985. Sindbis virus proteins nsP1 and nsP2 contain homology to nonstructural proteins from several RNA plant viruses. *J. Virol.* 53:536-42
5. Allison, R. F., Sorenson, J. C., Kelly, M. E., Armstrong, F. B., Dougerthy, W. G. 1985. Sequence determination of the capsid protein gene and flanking regions of tobacco etch virus: Evidence for synthesis and processing of a polyprotein in potyvirus genome expression. *Proc. Natl. Acad. Sci. USA* 82:3969–72
6. Barker, R., Jarvis, N., Thompson, D.,

Loesch-Fries, L. and Hall, T. 1983. *Nucleic Acids Res.* 11:2881–91
7. Boccara, M., Hamilton, W. D. O., Baulcombe, D. C. 1986. The organization and interviral homology of genes at the 3'-end of tobacco rattle virus RNA1. *EMBO J.* 5:223–29
8. Bruening, G. 1977. Plant covirus systems: Two component systems. In *Comprehensive Virology,* ed. H. Fraenkel-Comat, R. R. Wagner, 11:55–141. New York: Plenum. 348 pp.
9. Cornelissen, B. J. C., Bol, J. F. 1984. Homology between proteins encoded by tobacco mosaic virus and two tricornaviruses. *Plant Mol. Biol.* 3:379–84
10. Cornelissen, B. J. C., Brederode, F. Th., Moorman, R. J. M., Bol, J. F. 1983. *Nucleic Acids Res.* 11:1253–65
11. Cornelissen, B. J. C., Brederode, F. Th., Veeneman, G. H., Van Boom, J. J., Bol, J. F. 1983. *Nucleic Acids Res.* 11:3019–25
12. Davies, J. W., Hull, R. 1982. Genome expression of plant positive-strand RNA viruses. *J. Gen. Virol.* 61:1–14
13. Domingo, E., Sabo, D., Taniguchi, T., Weissmann, C. 1978. Nucleotide sequence heterogeneity of an RNA phage population. *Cell* 13:735–44
14. Dougerthy, W. G., Hiebert, E. 1985. Genome structure and expression of plant RNA viruses. In *Molecular Plant Virology,* ed. J. W. Davies, 2:23–81. Boca Raton, Fla: CRC
15. Emini, E. A., Schleif, W. A., Colonno, R. J., Wimmer, E. 1985: Antigenic conservation and divergence between the viral-specific proteins of poliovirus type 1 and various picornaviruses. *Virology* 140:13–20
16. Franssen, H., Leunissen, J., Goldbach, R., Lomonossoff, G., Zimmern, D. 1984. Homologous sequences in nonstructural proteins from cowpea mosaic virus and picornaviruses. *EMBO J.* 3:855-61
17. Franssen, H., Moerman, M., Rezelman, G., Goldbach, R. 1984. Evidence that the 32,000-dalton protein encoded by the bottom-component of cowpea mosaic virus is a proteolytic processing enzyme. *J. Virol.* 50:183–90
18. Goelet, P., Lomonossoff, G. P., Butler, P. J. G., Akam, M. E., Gait, M. J., Karn, J. 1982. Nucleotide sequence of tobacco mosaic virus RNA. *Proc. Natl. Acad. Sci. USA* 79:5818–22
19. Goldbach, R. 1986. Comoviruses: Molecular biology and replication. In *The Plant Viruses: Viruses with Bipartite RNA Genomes and Isometric Particles,* ed. B. D. Harrison, A. F. Murant. New York: Plenum. In press
20. Goldbach, R., Rezelman, G. 1983. Orientation of the cleavage map of the 200 kilodalton polyprotein encoded by the bottom-component RNA of cowpea mosaic virus. *J. Virol.* 46:614–19
21. Goldbach, R., Van Kammen, A. 1985. Structure, replication and expression of the bipartite genome of cowpea mosaic virus. See Ref. 14, pp. 83–120
22. Guilley, H., Carrington, J. C., Balàzs, E., Jonard, G., Richards, K., Morris, T. J. 1985. Nucleotide sequence and genome organization of carnation mottle virus RNA. *Nucleic Acids Res.* 13: 6663–77
23. Haseloff, J., Goelet, P., Zimmern, D., Ahlquist, P., Dasgupta, R., Kaesberg, P. 1984. Striking similarities in amino acid sequence among nonstructural proteins encoded by RNA viruses that have dissimilar genomic organization. *Proc. Natl. Acad. Sci. USA* 81:4358–62
24. Hogle, J. M., Chow, M., Filman, D. J. 1985. Three-dimensional structure of poliovirus at 2.9 Å resolution. *Science* 229:1358–67
25. Holland, J., Spindler, K., Horodyski, F., Graban, E., Nichol, S., VandePol, S. 1982. Rapid evolution of RNA genomes. *Science* 215:1577–85
26. Joshi, S., Haenni, A.-L. 1984. Plant RNA viruses: strategies of expression and regulation of viral genes. *FEBS Lett.* 177:163–74
27. Kamer, G., Argos, P. 1984. Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Res.* 12:7269–82
28. Kiberstis, P. A., Loesch-Fries, L. S., Hall, T. C. 1981. Viral protein synthesis in barley protoplasts inoculated with native and fractionated brome mosaic virus RNA. *Virology* 112:804–8
29. King, A. M. Q., McCahon, D., Saunders, K., Newman, J. W. I., Slade, W. R. 1985. Multiple sites of recombination within the RNA of foot-and-mouth disease virus. *Virus Res.* 3:373–84
30. King, A. M. Q., McCahon, D., Slade, W. R., Newman, J. W. I. 1982. Recombination in RNA. *Cell* 29:921–28
31. Kozak, M. 1981. Mechanism of mRNA recognition by eucaryotic ribosomes during initiation of protein synthesis. *Curr. Top. Microbiol. Immunol.* 93:81–123
32. Kozak, M. 1981. Possible role of flanking nucleotides in recognition of the

AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Res.* 9:5233–52
33. Kozak, M. 1984. Compilation and analysis of sequences upstream from the translational start site in eucaryotic mRNAs. *Nucleic Acids Res.* 12:857–72
34. Leonard, D. A., Zaitlin, M. 1982. A temperature-sensitive strain of tobacco mosaic virus defective in cell-to-cell movement generates an altered viral-coded protein. *Virology* 117:416–24
35. Lomonossoff, G., Shanks, M. 1983. The nucleotide sequence of cowpea mosaic virus B RNA. *EMBO J.* 2:2253–58
36. Longworth, J. F. 1978. Small isometric viruses of invertebrates. *Adv. Virus Res.* 23:103–57
37. Matthews, R. E. F. 1981. *Plant Virology.* New York: Academic. 897 pp. 2nd ed.
38. Meshi, T., Ohno, T., Okada, Y. 1982. Nucleotide sequence of the 30K protein cistron of cowpea strain of tobacco mosaic virus. *Nucleic Acid Res.* 10: 6111–17
39. Moore, N. F. 1985. The replication schemes of insect viruses in host cells. In *Viral Insecticides for Biological Control,* ed. K. Maramorosch, K. E. Sherman, pp. 635–74. New York: Academic. 809 pp.
40. Moore, N. F., Reavy, B., King, L. A. 1985. General characteristics, gene organization and expression of small RNA viruses of insects. *J. Gen. Virol.* 66:647–59
41. Moore, N. F., Tinsley, T. W. 1982. The small RNA-viruses of insects. *Arch. Virol.* 72:229–45
42. Morris-Krsinich, B. A. M., Forster, R. L. S., Mossop, D. W. 1983. The synthesis and processing of the nepovirus grapevine fanleaf virus proteins in rabbit reticulocyte lysate. *Virology* 130:523–26
43. Murthy, M. R. N. 1983. Comparison of the nucleotide sequences of cucumber mosaic virus and brome mosaic virus. *J. Mol. Biol.* 168:469–75
44. Nahmias, A. J., Reanney, D. C. 1977. The evolution of viruses. *Ann. Rev. Ecol. Syst.* 8:29–49
45. Nassuth, A., Alblas, F., Bol, J. F. 1981. Localization of genetic information involved in the replication of alfalfa mosaic virus. *J. Gen. Virol.* 53:207–14
46. Ohno, T., Takamatsu, N., Meshi, T., Okada, Y., Nishiguchi, M., Kiho, Y. 1983. Single amino acid substitution in 30K protein of TMV, defective in virus transport function. *Virology* 131:255–58
47. Pincus, S. E., Diamond, D. C., Emini, E. A., Wimmer, E. 1986. Guanidine selected mutants of poliovirus: Mapping of point mutations to polypeptide 2C. *J. Virol.* In press
48. Reanney, D. C. 1982. The evolution of RNA viruses. *Ann. Rev. Microbiol.* 36:47–73
49. Reanney, D. C. 1984. The molecular evolution of viruses. In *The Microbe 1984.* Part I, ed. B. W. J. Mahy, J. R. Pattison, pp. 175–96. London: Cambridge Univ. Press
50. Rezaian, M. A., Williams, R. H. V., Gordon, K. H. J., Gould, A. R., Symons, R. H. 1984. Nucleotide sequence of cucumber-mosaic-virus RNA2 reveals a translation product significantly homologous to corresponding proteins of other viruses. *Eur. J. Biochem.* 143:277–84
51. Rezaian, M. A., Williams, R. H. V., Symons, R. H. 1985. Nucleotide sequence of cucumber mosaic virus RNA1. *Eur. J. Biochem.* 150:331–39
52. Rossmann, M. G., Arnold, E., Erickson, J. W., Frankenberger, E. A., Griffith, J. P., et al. 1985. Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* 317:145–53
53. Savithri, H. S., Murthy, M. R. N. 1983. Evolutionary relationship of alfalfa mosaic virus with cucumber mosaic virus and brome mosaic virus. *J. Biosci.* 5:183–87
54. Strauss, E. G., Rice, C. M., Strauss, J. H. 1983. The sequence coding for the alphavirus nonstructural proteins is interrupted by an opal termination codon. *Proc. Natl. Acad. Sci. USA* 80:5271–75
55. Strauss, E. G., Rice, C. M., Strauss, J. H. 1984. Complete nucleotide sequence of the genomic RNA of Sindbis virus. *Virology* 133:92–110
56. Strauss, E. G., Strauss, J. H. 1983. Replication strategies of the single stranded RNA viruses of eukaryotes. *Curr. Top. Microbiol. Immunol.* 105:1–98
57. Toh, H., Hayashida, H., Miyata, T. 1983. Sequence homology between retroviral reverse transcriptase and putative polymerases of hepatitis B virus and cauliflower mosaic virus. *Nature* 305: 827–29
58. Toyoda, H., Dunn, J. J., Studier, F. W., Murray, M. G., Nicklin, M. J. H., Wimmer, E. 1985. Investigation of the proteinase that cleaves the polyprotein of poliovirus at the specific tyr-gly sites. *Fourth Meet. Eur. Group Mol. Biol.*

Picornaviruses, Abstr. H7. Seillac, France
59. Vance, V., Beachy, R. N. 1984. Translation of soybean mosaic virus RNA in vitro: evidence of protein processing. *Virology* 132:271–81
60. Volovitch, M., Modjtahedi, N., Yot, P., Brun, G. 1984. RNA-dependent DNA polymerase activity in cauliflower mosaic virus–infected plant leaves. *EMBO J.* 3:309–14
61. Van Wezenbeek, P., Verver, J., Harmsen, J., Vos, P., Van Kammen, A. 1983. Primary structure and gene organization of the middle-component RNA of cowpea mosaic virus. *EMBO J.* 2:941–46
62. Zimmern, D. 1983. Homologous proteins encoded by yeast mitochondrial introns and by a group of RNA viruses from plants. *J. Mol. Biol.* 171:345–52
63. Zimmern, D., Hunter, T. 1983. Point mutation in the 30K open reading frame of TMV implicated in temperature-sensitive assembly and local lesion spreading of mutant Ni 2519. *EMBO J.* 2:1893–1900