Review article

Highlights and prospects of potyvirus molecular biology

José Luis Riechmann,* Sonia Lain and Juan Antonio García

Centro de Biología Molecular (CSIC-UAM), Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain

Introduction

The potyvirus group [named after its type member, potato virus Y (PVY)] is the largest of the 34 plant virus groups and families currently recognized (Ward & Shukla, 1991). It contains at least 180 definitive and possible members (or 30% of all known plant viruses) which cause significant losses in agricultural, pasture, horticultural and ornamental crops (Ward & Shukla, 1991). These viruses are unique in the diversity of inclusion bodies that are formed during the infection cycle (see Lesemann, 1988). A feature shared by all potyviruses is the induction of characteristic pinwheel or scroll-shaped inclusion bodies in the cytoplasm of the infected cells (Edwardson, 1974). These cylindrical inclusion (CI) bodies are formed by a virus-encoded protein and can be considered as the most important phenotypic criterion for assigning viruses to the potyvirus group (Milne, 1988; Shukla et al., 1989; Ward & Shukla, 1991). Many potyviruses also induce cytoplasmic amorphous inclusion bodies and some form nuclear inclusions. Virions are flexuous and rod-shaped, 680 to 900 nm long and 11 to 15 nm wide, made up of about 2000 units of a single structural protein surrounding one molecule of ssRNA of approximately 10000 nucleotides and messenger polarity (Dougherty & Carrington, 1988). Although some of these viruses are transmitted by mites, and possibly by whiteflies, the predominant transmission of potyviruses is by aphids. In addition, there are viruses that share many characteristics with the typical potyviruses but have some important peculiarities. They are transmitted by fungi, have bipartite ssRNA genomes that are encapsidated in two different rod-shaped virions (approximately 275 and 550 nm long), and are not serologically related to potyviruses, differences that would set them apart from the potyvirus group (Kashiwazaki et al., 1990). On the other hand, their cistron order, strategy of genome expression and amino acid sequence homologies as well as the formation of similar cylindrical inclusions in infected cells indicate a common ancestry and a close evolutionary relationship between these viruses [whose type member is barley yellow mosaic virus (BaYMV)] and typical potyviruses (Kashiwazaki et al., 1990). The taxonomy of the potyvirus group, which has been studied extensively, may therefore be quite complex (reviewed in Milne, 1988; Ward & Shukla, 1991; for a proposal on potyvirus taxonomy, see Barnett, 1991).

A breakthrough in potyvirus studies was achieved in 1986 when the complete genome sequences of two members of this group, tobacco etch virus (TEV; Allison et al., 1986) and tobacco vein mottling virus (TVMV; Domier et al., 1986) were reported for the first time. Intensive research during the following years led to a greater understanding of the potyviral genome structure and expression. The complete genome sequences of two more potyviruses, PVY (Robaglia et al., 1989) and plum pox virus (PPV; Lain et al., 1989a; Maiss et al., 1989; Teycheney et al., 1989) are now available. Unsuspected relationships between potyviruses and other groups of plant and animal viruses arose from sequence comparisons (Domier et al., 1987; Lain et al., 1989b), allowing functions to be proposed for most of the potyviral gene products. However, actual knowledge about these functions is still scarce, and not all of the gene products have been identified in vivo.

A second breakthrough in potyvirus research is probably near, given the recent isolation of full-length cDNA clones of three potyviruses from which infectious transcripts can be synthesized: TVMV (Domier et al., 1989), PPV (Riechmann et al., 1990) and zucchini yellow mosaic virus (Gal-On et al., 1991). The ability to generate virus infection from cloned cDNA of potyviruses opens the possibility of applying genetic engineering techniques for the study of their biology at the molecular level. Aspects of the potyvirus life cycle such as translation, replication, symptom induction, and cell-to-cell and plant-to-plant propagation are now amenable to a new experimental approach, and the functions that have been proposed for potyviral gene products can now be tested in vivo. Also, discoveries have been made working with potyviruses that will influence our knowledge of the biology of most plus-stranded RNA viruses; on the other hand, results obtained by working with other viruses are
affecting our perception of the potyvirus group. For all these reasons, this is a timely moment to review progress in potyvirus research. The aim of this article is not to present a comprehensive review but to focus on the most relevant and novel aspects of potyvirus molecular biology and on the expected trends for the forthcoming years. Previous reviews that are more extensive or that cover mainly other topics of potyvirus biology are already available (Dougherty & Carrington, 1988; Shaw et al., 1990; Dougherty et al., 1990; Shukla et al., 1991).

Genome structure and organization

Potyviruses, like the majority of plant viruses, have a single-stranded, positive-sense RNA genome. The RNA is approximately 10000 nucleotides long, and has a 5'-terminal genome-linked protein (VPg) (Slaw et al., 1985; Riechmann et al., 1989; Murphy et al., 1990) and a 3' poly(A) tail (Hari et al., 1979). It contains one long open reading frame which is translated into a large polyprotein (ranging between 340K and 368K depending on the potyvirus) which is subsequently cleaved into smaller polypeptides. Thus, potyviruses are quite similar, in terms of their genomic structure and strategy of expression, to the plant bipartite como- and nepoviruses and to the animal picornaviruses. In addition, in the genomes of all these viruses a cluster of genes is found that encodes, in conserved order, a number of non-structural proteins exhibiting characteristic amino acid blocks which have been shown or suggested to be involved in RNA replication (Fig. 1). For all these reasons it has been proposed that the como-, nepo- and potyviruses may be arranged in a supergroup of picorna-like plant viruses (Goldbach, 1986, 1987; Goldbach et al., 1990). The possible genetic linkage between these viruses is interesting in terms of virus evolution because, in contrast with como- and nepoviruses, the genetic organization of the potyvirus genome cannot easily be inferred from the picornaviral genome (reviewed in Goldbach, 1987; Fig. 1). Despite the presence of the characteristic set of replication genes, the positions and functions of other genes are unique. Among these differences are the location of the coat protein cistron at the 3' end of the genome, the requirement of three proteolytic activities for polyprotein processing, and the sequence similarities between some non-structural and structural potyviral proteins and the equivalent proteins of some otherwise unrelated viruses (see below). These findings strongly support the idea that recombination of distinct gene sets occurred during virus evolution (the term ‘modular evolution’ refers to the mixing and joining of such modules; Zimmern, 1987; Gibbs, 1987), and indicate that these events can follow complex pathways (Goldbach, 1987).

The different gene products into which the potyviral polyprotein is cleaved are, proceeding from the N to the C terminus of the polyprotein, P1 (first protein; protease), the helper component/protease (HC-Pro) protein, P3 (third protein) [the denomination as P1 and P3 of proteins previously named by their Mr, has been proposed by Shukla et al. (1991)], a putative 6K peptide (6K1), the CI protein with RNA helicase activity, a second 6K peptide (6K2), the nuclear inclusion ‘a’ protein (NiA; VPg and protease), the nuclear inclusion ‘b’ protein (NiB; the presumed RNA polymerase) and the capsid protein (CP) (Fig. 1). The complete nucleotide sequence of the bipartite genome of the potyvirus-related BaYMV has recently been reported (Kashiwazaki et al., 1990, 1991; Davidson et al., 1991). Cistrons encoding an N-terminal protein, the putative 6K1 peptide, the CI, NiA, NiB proteins and CP have been localized in the larger RNA (RNA 1), arranged in the same order as that found in the middle and 3' regions of the potyvirus genome (Kashiwazaki et al., 1990; see Fig. 1). In contrast, differences in genetic organization have been found between BaYMV RNA 2 and the 5'-terminal region of the potyvirus genome (Kashiwazaki et al., 1991; Davidson et al., 1991; see Fig. 1).

The length of potyviral 5' non-coding regions ranges between 144 (TEV) and 205 (TVMV) nucleotides. They are especially rich in adenine residues and have few guanine residues. Alignment of the non-coding regions of PPV, TEV, TVMV and PVY revealed some nucleotide blocks conserved in the four viruses (Lain et al., 1989a; Turpen, 1989), although this is not accompanied by conservation of predicted secondary structures (Turpen, 1989). The conservation of these sequences in the four potyvirus RNAs suggests that they could play an important role in the virus life cycle, for example in the processes of encapsidation, translation or replication [in which they could either reflect or contribute to a different replication mechanism for the plus and minus strands, since sequences related to these conserved blocks were not found at the 5' end of the complementary strand (Lain et al., 1989a)].

The 3' non-coding regions of different potyviruses have been described as heterogeneous in size, sequence and predicted secondary structure (Lain et al., 1988; Turpen, 1989; Quemada et al., 1990a, b). Features that seem common to all of them are the presence of AU-rich segments and the fact that each sequence can be predicted to fold into stable secondary structures (Turpen et al., 1989). In contrast with the high sequence diversity found among the 3'-non-coding regions of different potyviruses, (i) among different strains of the same potyvirus they are clearly more conserved than for
Example: the CP coding region (Wetzel et al., 1991; M. T. Cervera & J. A. Garcia, unpublished results), another genetic marker that has been employed for potyvirus taxonomy (Shukla & Ward, 1989a; Ward & Shukla, 1991; Quemada et al., 1990) and (ii) several short segments displaying sequence homology among different potyviruses have been identified (Lain, 1990; Uyeda et al., 1990). The poly(A) tails have been found to be very heterogeneous in length (Allison et al., 1985; Lain et al., 1988); in PPV they have a very peculiar length distribution that might have some functional significance as it is rapidly recovered upon replication in plants of PPV infectious in vitro transcripts (Riechmann et al., 1990).

**Genome expression: polyprotein processing**

Potyvirus genomic RNA functions as mRNA, but very little is known about its translation process. It has been shown that the TEV 5′ non-coding region can function as an enhancer of translation (Carrington & Freed, 1990). This property could be due to the likely absence of strong secondary structure, as suggested for the enhancing leaders of other plant viruses (Jobling & Gehrke, 1987; Altmann et al., 1990). Potyvirus genomic RNA presents an atypical structure (VPg instead of a cap) at its 5′ end, and PPV RNA translation initiates at an internal AUG (Riechmann et al., 1991). Despite the genetic relationship among potyviruses, picornaviruses and comoviruses, the mechanism of internal ribosomal entry which has been demonstrated or suggested for these two last groups of viruses (Jackson et al., 1990; Thomas et al., 1991) has not been found to operate for translation of PPV RNA, of which the internal AUG seems to be recognized by a leaky scanning mechanism (Riechmann et al., 1991). Analysis of the proteolytic processing of the polyprotein into which potyvirus RNA is translated has been particularly illuminating because this process can be studied both in vitro (using cell-free translation systems).
and in heterologous (*Escherichia coli*) or homologous (transgenic plants) *in vivo* systems. Two virus-encoded proteases, Nla (Carrington & Dougherty, 1987a; Hellmann *et al.*, 1988; Chang *et al.*, 1988; García *et al.*, 1989a; Ghabrial *et al.*, 1990) and HC-Pro (Carrington *et al.*, 1989a), process co- and post-translationally the large viral precursor polyprotein, and it has been shown recently that the third proteolytic activity that is required for completion of the processing pathway (Carrington *et al.*, 1990) resides in the P1 protein (Verchot *et al.*, 1991).

Nla is responsible for cleavages in the C-terminal two-thirds of the polyprotein (Dougherty & Carrington, 1988; Fig. 2), whereas HC-Pro and P1 autocatalytically cleave at their respective C termini (Carrington *et al.*, 1989a, b; Verchot *et al.*, 1991; Fig. 2).

Nla protease cleavage sites are defined by sequences of seven amino acids which have been characterized by a variety of experimental approaches (Carrington & Dougherty, 1987a, 1988; Carrington *et al.*, 1988; Dougherty *et al.*, 1988, 1989a; García *et al.*, 1989a, b). Six of these Nla cleavage sites are present in the potyvirus polyprotein, named A to F in Fig. 2. Cleavages at five of them (B to F) have been demonstrated by analysing the products of proteolytic processing *in vitro* or in *E. coli* (Carrington & Dougherty, 1987a; Hellmann *et al.*, 1988; Carrington *et al.*, 1988; García *et al.*, 1989a) and by N-terminal sequencing of several potyviral proteins (Robaglia *et al.*, 1989; Martin *et al.*, 1990; Ward & Shukla, 1991). Recently, partial cleavage at PPV site A has been shown to occur *in vitro* (J. A. García & J. L. Riechmann, unpublished results), although cleavage at the TEV site has not been detected (Parks *et al.*, 1992), partial processing at this place presumably happens *in vivo*: TVMV proteins of 37K and 42K (probably P3 and P3+6K1) have been detected in infected plants by using antibodies raised against a bacterially expressed TVMV 42K protein originating from the P3+6K1 cistron (Rodriguez-Cerezo & Shaw, 1991). Sequence comparison studies of the two products that cleavage at this site would generate, P3 and 6K1, also favour the hypothesis of its existence (Lain *et al.*, 1989a; Lain, 1990). Site A is therefore probably a suboptimal cleavage site that is only partially processed. Amino acids of the heptapeptide sequence (see Fig. 2) can either be essential for a functional cleavage site (Dougherty *et al.*, 1988) or affect the rate of cleavage (Dougherty *et al.*, 1989a), as demonstrated by biochemical and mutational analyses of TEV site F. It has been further shown that different heptapeptide sequences have, at least *in vitro*, different cleavage reaction profiles (Dougherty & Parks, 1989).

The requirement that potyviruses have for the conservation of an extended sequence motif at the Nla cleavage sites is a characteristic not shared by other virus groups using polyprotein processing for genome expression (i.e. comoviruses and picornaviruses; for a review, see Wellink & van Kammen, 1988). Information in the polyprotein for substrate specificity resides mainly at the amino acid sequence level, since Nlb–CP heptapeptide junctions inserted into a non-specific protein background were properly processed; additionally, the conformational context might contribute to substrate efficiency (Carrington & Dougherty, 1988; García *et al.*, 1989b). The conclusions obtained from all these experiments, performed mainly with TEV, will probably be true for other potyviruses because, although examination of the cleavage sites did not reveal an extended conserved cleavage motif used by all of them, a number of biochemical features are maintained (see Ghabrial *et al.*, 1990).

![Fig. 2. Deduced amino acid sequences of the demonstrated and putative cleavage sites for the Nla and HC-Pro proteases of four potyviruses and for the Nla-like protease of the potyvirus-related BaYMV. Scissile bonds are indicated by /]. A map of the potyviral polyprotein indicating the different protein products, the location of the cleavage sites and the protease responsible for each cleavage is shown in the upper part of the figure. Nla cleavage sites are named A to F, and V.](image-url)
However, and although NIa proteases are highly homologous (García et al., 1989a), PPV, TEV and TVMV cleavage sites are efficiently recognized only by their respective proteases (García et al., 1989b; García & Lain, 1991; Parks & Dougherty, 1991), indicating that very specific interactions must take place during the proteolytic reaction.

NIa protease also appears to catalyse post-translational proteolysis at an internal cleavage site which has been characterized recently for TEV (Dougherty & Parks, 1991). The sequence of this cleavage site (named V in Fig. 2), differs, although maintaining some characteristic features, from the consensus sequences of the other six NIa (A to F) sites; such differences might have the effect of promoting suboptimal cleavage and, in fact, this proteolytic event seems to happen in some but not all the NIa molecules (Dougherty & Parks, 1991).

This site is located between the VPg and protease domains of the NIa protein, and its cleavage might therefore be related to some step of the viral RNA replication process (see below).

Experiments using in vitro transcription and translation systems and E. coli expression systems have shown that NIa cleavage sites differ in their susceptibility to processing in cis (intramolecularly) or in trans (intermolecularly) (Carrington & Dougherty, 1987a, b; Carrington et al., 1988; Hellmann et al., 1988; García et al., 1989a, c, 1990). NIa protease is catalytically active within the polyprotein, excising itself by cis cleavages at sites D and E in a process that might happen as a primary event (maybe cotranslational) in the proteolytic cascade (Carrington & Dougherty, 1987a, b; Carrington et al., 1988; Hellmann et al., 1988; García et al., 1989a, c, 1990). In summary, NIa-catalysed proteolytic maturation of the potyviral polyprotein is a regulated process involving cis and trans cleavages at sites which are recognized with different efficiencies and processed at different rates. Therefore, viruses like the potyviruses, unable to regulate their gene expression differentially at the levels of transcription and translation, may have evolved to be capable of regulating the expression of gene products by sequential proteolytic events (Dougherty et al., 1989a).

This post-translational regulation could imply that products and precursors have different functions, or it may be a way in which the ratio of inactive : active viral moieties is regulated (Dougherty & Parks, 1989). The possibility that processing intermediates of the proteolytic pathway have a function is still an open question (see below).

Molecular genetic analyses have supported the hypothesis that NIa and other virus-encoded proteases (such as the picornavirus 3C and the cowpea mosaic virus (CPMV) 24K proteases) are related to the trypsin-like family of cellular serine proteases, but with the substitution of a Cys as the active site nucleophile (Bazan & Fletterick, 1988; Gorbalenya et al., 1989a). The proposed catalytic triad, composed of His, Asp and Cys residues, is located in the C-terminal half of the molecule (Dougherty et al., 1989b; García et al., 1990; Ghabrial et al., 1990) and, in fact, NIa protease is probably multifunctional as its N-terminal half is not relevant for its proteolytic activity in vitro or in E. coli (Carrington & Dougherty, 1987a; García & Lain, 1991; Dougherty & Parks, 1991) but has been found covalently attached to the genomic RNA (VPg; see below). Interestingly, a substitution of Glu for the catalytic Asp in the TEV and PPV NIa proteases resulted in anomalous proteolytic activities, the effect of these substitutions depending on the particular cleavage site analysed (Dougherty et al., 1989b; García et al., 1990). It could be inferred from these results that there are different structural requirements in the protease, probably similar in TEV and PPV, for processing at the various cleavage sites. The amino acid similarity among NIa proteases of different potyviruses is lower at their C-terminal end, suggesting that this region of the protease may be important in forming the substrate-binding pocket which allows each of the potyviral NIa proteases to recognize its own unique cleavage sequences (Ghabrial et al., 1990). Efforts have been made to delimit protease regions involved in catalysis and in substrate recognition through the construction of hybrid NIa proteases (García & Lain, 1991; Parks & Dougherty, 1991). The results obtained suggest that the recognition and catalytic sites of the NIa proteases are closely interlinked and that the main determinants for substrate specificity lie in the C-terminal one-third of the protease (García & Lain, 1991; Parks & Dougherty, 1991). It has also been suggested that the extreme carboxyl end of the protease could be involved in maintenance of the proper structure that would allow other parts of the protein to interact correctly with the various cleavage sites, since the effect caused by different mutations in this region depends on the cleavage site considered (Carrington & Dougherty, 1987a; García et al., 1990). Finally, it has been shown that the potyviral NIa protease processes a great deal of substrate efficiently (García et al., 1989c).

The helper component protease (HC-Pro; a protein formerly known as HC that was renamed HC-Pro for its two activities, insect transmission and protease (Carrington et al., 1989a)] is responsible for processing at its own C terminus via an autocatalytic mechanism and, in vitro, exhibits little or no proteolytic activity in trans (Carrington et al., 1989a, b). Cleavage in vitro occurs between a Gly–Gly dipeptide (Carrington et al., 1989a) and, although the primary sequence features that define HC-Pro cleavage sites have not been determined, there is a good consensus around the presumed Gly–Gly substrate
dipeptides of different potyviruses (Fig. 2). The HC-Pro autocatalytic processing event seems to be, at least in this particular in vitro system, extremely efficient, occurring immediately after ribosomes have traversed the cleavage site coding sequence (Carrington et al., 1989b). The proteolytically active domain of the TEV HC protein has been localized at its C-terminal half (Carrington et al., 1989a, b), and the identification of two essential residues in this protease is in support of the hypothesis that HC-Pro most closely resembles members of the cysteine-type family of proteases (Oh & Carrington, 1989). Although a cysteine-type protease comparable to HC-Pro is not encoded by the genomes of the otherwise related picornaviruses and comoviruses, possible analogues have been identified as products of the coronavirus (Oh & Carrington, 1989) and Sindbis virus (Hardy & Strauss, 1989) genomes. A protease that bears a striking resemblance to HC-Pro and that also cleaves between a Gly–Gly dipeptide is encoded by a dsRNA hypovirus-associated virus of the chestnut blight fungus (Choi et al., 1991a, b). In addition to its role in the processing of the polyprotein, the protease activity of HC-Pro could function during the aphid transmission process in an as yet unidentified manner (Carrington et al., 1989a).

By expressing TEV polyproteins in transgenic plants, Carrington et al. (1990) showed that a novel (caused by neither HC nor NIa proteases) proteolytic activity is required for processing at the P1 C terminus. Sequence similarities suggested that this process could be catalysed by P1 itself. In spite of the high variability found among the P1 proteins of TEV, TVMV, PVY and PPV, amino acids characteristic of the serine proteases have been found to be completely conserved among these (these are: HxsDx30 or 31GxSG) and their relative positioning is also coincident with that found in the serine proteases (Lain, 1990; Verchot et al., 1991). This hypothesis has recently been confirmed using an in vitro wheatgerm extract translation system and two of these conserved amino acids have been identified as essential for proteolytic processing (Verchot et al., 1991). The P1 protein thus behaves in a similar way to the L polypeptide of foot-and-mouth disease virus (a picornavirus), catalysing its own release from the N terminus of the nascent polypeptide (Strebel & Beck, 1986). However, another factor besides P1 might be required for cleavage at this site, since it does not reproducibly occur in an in vitro rabbit reticulocyte lysate system (Hiebert et al., 1984b; Carrington et al., 1989a). Alternatively, the restriction of processing in the reticulocyte-based system could be due to the presence of a protease inhibitor (Verchot et al., 1991).

Several questions remain unanswered concerning the proteolysis of the potyviral polyprotein; for instance, the existence and nature of cofactor(s) that might be involved in this process and the identity of the cleavage site that separates P1 from HC-Pro (for a discussion about its localization, see Ward & Shukla, 1991). The P3 protein could be one of these proteolytic cofactors: the assignment of a function to the P3 protein by sequence similarity searches is hampered by its low level of homology among the different potyviruses, but some similarity has been found between the N-terminal part of P3 (which is clearly more conserved than the rest of the protein) and the 32K protein of CPMV (comovirus group), as the potyviral sequence EPYx,SPrx,Lx-Ax,Nx,Gx,Ex,W resembles the comoviral EPFx,1,Vx-Ax,Nx,Gx,M (Lain, 1990). This sequence relationship and the fact that the proteins are located in equivalent positions in their corresponding genetic maps suggest that P3 protein could somehow be involved, by analogy with the 32K protein (Vos et al., 1988), in the regulation of the proteolytic processing of the potyviral polyprotein. Another aspect requiring further investigation is the participation of precursor forms of the NIa protease in the cleavage pathway, as reported for other viral systems (Palmenberg, 1990; de Groot et al., 1990). Recent in vitro studies with a variety of TEV polyproteins containing the NIa protease suggested that cleavage of the genome-derived polyprotein is mainly directed by the 49K form of the protease, although the TEV 50K–71K protein junction (site B in Fig. 2) was differentially processed by several of the protease forms (Parks et al., 1992). However, as the authors emphasized, it is unknown to what extent these cell-free studies might reflect the actual situation in the infected cells.

In summary, much work has been done aiming to unravel the proteolytic processing of the potyviral polyprotein. The picture emerging from these experiments reveals a complex process in which three different proteases and probably some other factor(s) are involved in the proteolysis at a variety of cleavage sites through an intricate pathway that involves both cis and trans processing events and whose initial steps are probably cotranslational and autocatalytic (Fig. 3). Most of these cleavage sites as well as the three virus-encoded proteases have been identified and characterized already. It should be noted, however, that the majority of the available data come from in vitro experiments and, although no qualitative differences between the in vitro results and the in vivo situation would be expected, topics like the temporal regulation and spatial localization of the proteolysis (aspects that could be important for the regulation of the virus replication cycle) cannot be addressed from these experiments.

**Genome replication**

Since almost nothing is known about the potyvirus RNA replication process, the current ideas dealing with this
Fig. 3. Diagrammatic representation of the potyviral polyprotein processing (based on Carrington et al., 1989a, with modifications) relying on the results obtained in vitro and in E. coli and on amino acid sequence features of several proteins. The primary events in the proteolytic pathway are probably cotranslational and autocatalytic, yielding both precursors and mature products (indicated by their corresponding names). These initial steps are followed by trans (and possibly cis) cleavages, although there is no information about the sequential order in vivo of all these processing events. The possible existence of an uncleaved Nla–NIb product is indicated, as is the protease responsible for each cleavage. Not all the possible intermediate products are indicated. A cistron map of the potyvirus genomic RNA is shown in the upper part of the figure; abbreviations and symbols are the same as in Fig. 1 and 2. Hel, helicase; Pro, protease; Rep, replicase. See text for details.

step of the viral life cycle are mainly based on the similarities in genome structure and strategy of expression among potyviruses and other more studied viruses, like the picornaviruses. In contrast to this scarcity of data on the mechanism of RNA replication, amino acid sequence comparisons have provided some clues about which proteins are involved in this process and the functions that they could perform. These proteins, encoded by cistrons that form the ‘RNA replication module’ (Fig. 1) are the CI protein, the 6K₁ and 6K₂ peptides, Nla and NIb proteins, and, perhaps, the P₃ protein.

The most prominent sequence feature of the CI protein is the presence of a nucleotide-binding motif (NTBM) (Domier et al., 1987; Lain et al., 1989 a; Robaglia et al., 1989) composed of cassettes of amino acids that are common to many proteins for which NTP-binding has been demonstrated or postulated (Gorbunova & Koonin, 1989). These cassettes of residues are found in proteins encoded by most positive-strand RNA viruses, and three groups of NTBM-containing viral proteins have been distinguished that are encoded by ‘Sindbis-like’ viruses, ‘picorna-like’ viruses and poty-viruses (Gorbunova & Koonin, 1989), in agreement with a previous classification of plant positive-sense RNA viruses (Goldbach, 1986, 1987). Moreover, extensive amino acid sequence similarity has been found among the NTBM-containing proteins encoded by plant potyviruses (CI) and animal flaviviruses (NS3), pestiviruses and hepatitis C virus (Lain et al., 1989 b; Gorbunova et al., 1989 b; Lain, 1990; Miller & Purcell, 1990). Further amino acid sequence comparisons with several proteins of viral and cellular origin, like for example eIF-4A, suggested that these NTBM-containing viral proteins possess a helicase activity (Gorbunova et al., 1988, 1989 b; Hodgman, 1988; Lain et al., 1989 b; Company et al., 1991).

As suggested from these sequence comparisons, the CI protein of PPV has been shown to have a nucleic acid-stimulated ATPase activity and to be able to unwind RNA duplexes (Lain et al., 1990, 1991). This unwinding activity is dependent on the hydrolysis of NTP to NDP plus P, and thus can be considered to be an RNA helicase activity. In the in vitro assay used, the PPV CI protein was able to unwind only dsRNA substrates with 3' single-strand overhangs, indicating that the helicase
activity functions in the 3’ to 5’ direction (Lain et al., 1990). This was the first report on a helicase activity associated with a protein encoded by an RNA virus. Although still scarce, experimental data indicate that the NTBM-containing proteins which the majority of positive-strand RNA virus genomes encode are involved in their replication. The genome of bromovirus (BMV; bromovirus group; 'Sindbis-like' supergroup of plant viruses) encodes a helicase-like protein, named the 1a protein (Hodgman, 1988), which is required for continued accumulation of all classes of BMV RNA (positive-strand, negative-strand and subgenomic) (Kroner et al., 1990). In agreement with the presumed necessity of an unwinding activity for RNA replication, it has been found that an RNA helicase activity is present in RNA polymerase preparations of encephalomyocarditis virus (a picornavirus; Dmitrieva et al., 1991) and alfalfa mosaic virus (ilarvirus group; Quadt et al., 1991). Moreover, an in vitro replication system able to catalyse complete replication of cucumber mosaic virus (CMV) RNAs has been developed recently (Hayes & Buck, 1990), and it was shown that the helicase-like CMV protein is an essential subunit of the purified RNA-dependent RNA polymerase (Hayes & Buck, 1990). In addition, the 2C protein, the NTBM-containing protein of poliovirus, seems to be involved in the release of newly synthesized positive-sense RNA molecules from the replication complex (Bienz et al., 1987; Li & Baltimore, 1988), an event which could be nicely explained by the action of an RNA helicase activity. It should be noted, however, that only the N-terminal region of the CI protein is related by sequence comparisons with the helicase activity (Lain et al., 1989b). Therefore, the C-terminal region of about 230 amino acids might be dispensable for the helicase activity. It could be involved in some of the other functions which have been proposed for the CI protein, for example the attachment of the replication complex to membranes (Domier et al., 1987) in a similar way to the NTBM-containing protein, 2C, of poliovirus (Bienz et al., 1987), the induction of vesicles in the infected cells (Calder & Ingerfeld, 1990), or the virus transport from cell to cell (Langenberg, 1986; see below). Given the aforementioned ubiquity of helicase-like proteins among positive-strand RNA viruses, suggesting that RNA unwinding is probably a general requirement for genomic replication, the study of the enzymic properties of the potyvirus CI protein will be of interest in the further understanding of RNA replication mechanisms and in the design of strategies to interfere with positive-strand RNA virus infections.

The two small peptides 6K1 and 6K2 that are predicted from the amino acid sequence of the potyviral polyprotein might also play a role in RNA replication, although they have not yet been identified in vivo. The presence of a stretch of hydrophobic amino acids in the 6K1 and in the 6K2 peptides, and the relative position of their cistrons in the potyviral genetic map (see Fig. 1) suggest that they could be analogous to the picornaviral 2B and 3A peptides, respectively (Lain et al., 1989a). Protein 2B seems to be involved in RNA replication (Johnson & Sarnow, 1991), and the function that has been proposed for 3A is the membrane binding of the poliovirus VPg precursor, protein 3AB (Giachetti & Semler, 1991). It is thus conceivable that 6K2 could anchor the NIa/VPg (see below) to the membranes and that processing at the 6K2–NIa junction might be related to some step of the RNA replication process, although cleavage at this site might not be the last processing event to release the VPg from the polyprotein precursor (see below).

NIa and NIb must be involved in RNA replication; the NIb protein has been postulated as the potyviral RNA-dependent RNA polymerase based on the presence of conserved sequence motifs characteristic of these enzymes (Domier et al., 1987; Lain et al., 1989a; Robaglia et al., 1989; Poch et al., 1989). In addition to its role in the proteolytic processing of the potyviral polyprotein, for which only its carboxyl half is required (see above), the NIa protein (or its N-terminal part) has been shown to be the VPg. The VPgs of TVMV (Siaw et al., 1985) and PPV (Riechmann et al., 1989) have been identified as proteins of 24K and 22K, respectively. The TVMV VPg cistron has been mapped showing that the VPg is the N-terminal 24K portion of the NIa protein (Shahabuddin et al., 1988), and the TEV VPg has been found to be either the 49K NIa protein of its N-terminal 24K half (Murphy et al., 1990). It would be interesting to see whether the complete TVMV and PPV NIa proteins could be found attached to their respective RNAs, and whether the presence of the two different proteins at the 5' end of TEV RNA has any functional significance. The cleavage site that separates the VPg and protease domains of NIa protein has been identified (Dougherty & Parks, 1991; see above), and its proteolysis might be temporarily or functionally connected with the RNA replication process. It is interesting to note that this is a suboptimal cleavage site, and the final VPg precursor might therefore be the whole NIa molecule. In contrast, the VPg precursors of the related comov- and picorna-viruses are the 60K (58K protein + VPg) and 3AB polyproteins, respectively (Vos et al., 1988; Giachetti & Semler, 1991; see Fig. 1), indicating that different mechanisms might accomplish the necessary tuning between proteolytic and replication processes that must be similar to a certain extent among these viruses. A Tyr residue in the NIa protein of TVMV has recently been shown to link the TVMV VPg to the genomic RNA.
membrane, thereby sometimes introducing cytoplasm recognition motifs of proteins that interact with RNA of this transport is not understood. Potyviruses are mosaic virus induces the invagination of the nuclear trepo terminal part of the NIa proteins of PPV, TEV, TVMV and PVY contains a high content of aromatic amino acids (Tyr and Phe), and Gly and basic amino acids (Lain, 1990); this composition is typical of the consensus structural proteins present within the cytoplasm (Restrepo 1990). Association of TEV negative-strand RNA with chloroplasts in extracts of infected plants has been reported (Gadh & Hari, 1986) and a crude membrane fraction from PPV-infected plants has been shown to be able to catalyse specific PPV RNA synthesis (Martin & García, 1991). Nevertheless, we are still lacking clear data on the subcellular localization of potyviral replication, data that might be difficult to obtain given the varied morphological alterations that are induced in potyvirus-infected cells (for example, wheat streak mosaic virus induces the invagination of the nuclear membrane, thereby sometimes introducing cytoplasm into the nucleus; A Nassuth, personal communication). Timing of specific events (e.g. translation, RNA synthesis, encapsidation) in the replication cycle may be controlled, at least partially, by the levels of non-structural proteins present within the cytoplasm (Restrepo et al., 1990). It should be remembered, however, that not all potyviruses induce the formation of nuclear inclusions. In addition to NIa and N Ib, a high Mr product (94K to 110K) formed by both proteins has been observed in the nuclear inclusions (Knuhtsen et al., 1974; Hiebert et al., 1984a; Martin et al., 1990). Terminal amino acid sequencing of the PPV product (110K) has suggested that it is a non-processed NIa–N Ib polypeptide (Martin et al., 1990). But whether this potyvirus polypeptide is processed or not, its functional significance remains to be confirmed. In this regard, it is interesting to note that an uncleaved protease–polymerase precursor of CPMV has been shown to be involved in RNA replication (Dorssers et al., 1984). It has also been found that a functional ribonucleoprotein complex, composed of a cellular protein and the two poliovirus proteins (3Cpre and 3Dpol) equivalent to potyviral NIa and N Ib, forms around the 5′ end of poliovirus RNA and is required for positive-strand RNA synthesis but dispensable for negative-strand synthesis (Andino et al., 1990). It is speculated that these two proteins might be present in that complex as an uncleaved precursor (Andino et al., 1990).

In summary, the potyviral replication process is poorly understood. However, the recent preparation of a membrane fraction from infected cells able to catalyse PPV RNA synthesis (Martin & García, 1991) might be a first step towards the development of well-defined in vitro RNA replication systems that will greatly help in the unravelling of this process.

Virus spread

Four viral proteins, P1, CI, HC and CP, have been suggested or demonstrated to be involved in either cell-to-cell or plant-to-plant virus propagation.

The P1 protein has been proposed to be involved in the cell-to-cell spread of the infection on the basis of sequence similarity between P1 of TMV (but not of TEV, PPV and PVY) and the 30K movement protein of tobacco mosaic virus (TMV; Domier et al., 1987; Lain et al., 1989a; Robaglia et al., 1989). This sequence similarity involves the 30K TMV protein domain that has been implicated by Citovsky et al. (1990) in RNA binding. Since proteins presumably involved in cell-to-cell movement appear to display very little similarity, even among members of the same group (reviewed in Hull, 1989), a fact that could reflect the specificity of the virus–host interactions, the lack of sequence relationships does not permit exclusion of a cell-to-cell spread function for the potyviral N-terminal protein. Also this function cannot be ruled out by the recent finding that P1 has a proteolytic activity, because, since the proposed protease domain of P1 is confined to its carboxyl region (Verchot et al., 1991), additional function(s) might be
performed by this protein. In fact, its low conservation and its localization at the S' end of the polyprotein, as in the case of the proposed movement proteins of the related como- (Wellink & van Kammen, 1989; van Lent et al., 1990) and nepoviruses, support this hypothesis (Lain et al., 1989a).

Another protein that might be involved in cell-to-cell spread is the CI protein, as suggested on the basis of electron microscopy observations. The CI protein forms the cytoplasmic pinwheel or scroll-shaped inclusions that are induced in all potyvirus-infected plants (Edwardson, 1974; Dougherty & Hiebert, 1980). These inclusions have been found associated with the plasma membrane, with plasmodesmata, with the endoplasmic reticulum, and free in the cytoplasm (Lawson & Hearon, 1971; Langenberg, 1986; Lesemann, 1988; Baunoch et al., 1991). Additionally, CP or virions of several viruses have been detected by immunogold labelling techniques associated with these inclusion bodies (Langenberg, 1986; Langenberg & Purcifull, 1989; Langenberg et al., 1989). Based only on these ultrastructural observations, it has been proposed that the cylindrical inclusions, and therefore the CI protein, could be involved in virus transport from cell to cell through plasmodesmata (Lawson & Hearon, 1971; Langenberg, 1986; Calder & Ingerfeld, 1990). It has also been suggested that the CI protein could serve a structural function within the inclusion body, possibly to modify the cell cytoskeleton to adapt it for virus synthesis and/or translocation (Brakke et al., 1987).

As is the case with most plant viruses, natural plant-to-plant spread is accomplished by specific vectors, which for the majority of potyviruses are aphids. The HC protein is involved in the aphid transmission of potyviruses and must be acquired by the insect in conjunction with the virus (Pirone & Thornbury, 1983; Thornbury & Pirone, 1983; Hiebert et al., 1984b; Thornbury et al., 1985; Berger & Pirone, 1986). The HC protein has been associated with the amorphous inclusions that are induced by some potyviruses by several criteria (Hiebert et al., 1984b; De Mejia et al., 1985a, b; Baunoch et al., 1990), although functional studies have suggested that either the inclusion-bound form of this protein has been inactivated or, alternatively, the helper component activity is associated with a modified form of the inclusion protein (Thornbury & Pirone, 1983; Thornbury et al., 1985; Dougherty & Carrington, 1988). The biologically active form of the HC protein is believed to be a dimer (Hellmann et al., 1983; Thornbury et al., 1985). As Berger & Pirone (1986) have demonstrated, the HC protein mediates a selective localization of the virus into aphid mouthparts, suggesting a binding mechanism for the mode of action of HC. Since the proteolytically active domain has been localized to the C-terminal half of the TEV HC protein (Carrington et al., 1989b; see above), it seems likely that the sequences responsible for its role in transmission might be localized at its N-terminal region, which has a level of conservation among different potyviruses similar to that of the carboxyl region. The HC amino acid sequence of PVC (an aphid non-transmissible strain of PVY) and an aphid-transmissible strain of PVY have been determined and compared, and two of the differences that exist between these two HCs are localized in the homologous regions of the N-terminal part of HC which are conserved among all potyviruses sequenced to date (Thornbury et al., 1990). One of these two changes is found in a cysteine-rich region that might allow the dimerization of the molecule by the common fixation of metal atoms (Robaglia et al., 1989). These two amino acids are prime targets for mutational analyses of HC activity (Thornbury et al., 1990), studies that could be performed using transgenic plants expressing HC, although the use of full-length clones should be much more convenient (Berger et al., 1989).

The CP has been the most extensively characterized potyviral gene product in terms of sequence data (Ward & Shukla, 1991 and references therein). The interest in CP comes mainly from its usefulness in taxonomic and evolutionary studies, in diagnosis and in the so-called CP-mediated resistance, topics that have been extensively reviewed elsewhere [see Shukla & Ward (1988, 1989a, b), Shukla et al. (1991) and Ward & Shukla (1991), for taxonomy and diagnosis, and Beachy et al. (1990) for CP-mediated resistance]. For an extensive review on CP structure, see Shukla & Ward (1989b). The primary function of CP is, obviously, to encapsidate viral RNA, a process in which the highly conserved central and C-terminal regions characteristic of potyviral CPs are likely to be involved (Shukla & Ward, 1989b). In fact, the C-terminal part of potyvirus CPs shares sequence similarity with the CPs of the otherwise unrelated (in terms of genomic structure and strategy of expression) filamentous plant viruses potex-, carla-, clostero- and tombamoviruses (Morozov et al., 1987; Lain et al., 1988; Dinant et al., 1991). The molecular characterization of particle assembly and genome encapsidation processes would be facilitated by the recent development of microbial (E. coli and yeast) assembly systems (Jagadish et al., 1991). The high variability observed in the surface-exposed N-terminal region of potyvirus CP (Ward & Shukla, 1991) suggests that it is involved in virus-specific functions or host–vector–virus interactions. The CP could, for example, affect host range and symptom induction: the CP gene of a rod-shaped plant RNA virus, TMV, encodes a host response determinant (Saito et al., 1987). Sequence data have accumulated to support the hypothesis (Harrison & Robinson, 1988; Lain et al.,
### Table 1. Functions of potyviral gene products

<table>
<thead>
<tr>
<th>Gene product</th>
<th>Amino acid sequence feature</th>
<th>Function*</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Similarity between TVMV P1 and TMV 30K protein</td>
<td>Cell-to-cell movement? Polyprotein processing (Protease)</td>
</tr>
<tr>
<td></td>
<td>Amino acids typical of serine proteases</td>
<td></td>
</tr>
<tr>
<td>HC-Pro</td>
<td>Cysteine-rich region</td>
<td>Vector transmission Polyprotein processing (Protease)</td>
</tr>
<tr>
<td></td>
<td>Amino acids typical of cysteine proteases</td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td>Similarity with 32K CPMV protein</td>
<td>Polyprotein processing? (Cofactor?)</td>
</tr>
<tr>
<td>6K₁</td>
<td>Stretch of hydrophobic amino acids</td>
<td>Replication? Polyprotein processing (Protease)</td>
</tr>
<tr>
<td>CI</td>
<td>Nucleotide-binding motif</td>
<td>Replication? (RNA helicase)</td>
</tr>
<tr>
<td></td>
<td>Similarity with helicases</td>
<td></td>
</tr>
<tr>
<td>6K₂</td>
<td>Stretch of hydrophobic amino acids</td>
<td>Replication?</td>
</tr>
<tr>
<td>Nla</td>
<td>Amino acids typical of serine proteases</td>
<td>Polyprotein processing (Protease) Replication? (VPg)</td>
</tr>
<tr>
<td>Nlb</td>
<td>Motifs of RNA-dependent RNA polymerases</td>
<td>Replication? (RNA-dependent RNA polymerase?)</td>
</tr>
<tr>
<td>CP</td>
<td>(I/V)DAG block</td>
<td>Aphid transmission RNA encapsidation</td>
</tr>
</tbody>
</table>

* Functions that have not yet been supported by experimental evidence are indicated with a question mark, and those which have been are written in italics.

1988) that there is a specific involvement of a conserved (I/V)DAG block present at the N terminus of CP in virus transmission by aphids (Maiss et al., 1989; Eggenberger et al., 1989; Gal-On et al., 1990). Moreover, this hypothesis has been recently confirmed experimentally through the use of a full-length TVMV cDNA clone: a single amino acid change that alters the VDAG block to VDAE rendered a virus non-transmissible by aphids (Atreya et al., 1990). Additional mutants obtained by site-directed mutagenesis have demonstrated the relevance of the DAG block and other amino acids at the N terminus of the CP in aphid transmissibility (Atreya et al., 1991). No data are available on the function of this region of the protein for the transmission and it remains unknown whether it is involved in the maintenance of a conformation of the coat essential for the aphid transmission or whether it directly interacts with the HC or with any structure of the aphid. In addition to its natural functions, the potyviral CP has been used to obtain genetically engineered CP-mediated resistance (reviewed in Beachy et al., 1990). In contrast with what has generally been found for other viruses, there was no clear correlation between the levels of CP expression and the levels of protection (Lawson et al., 1990). However, (i) efficient genetically engineered protection against the natural insect transmission of the virus (Lawson et al., 1990) and (ii) protection against infection by heterologous potyviruses were obtained (Stark & Beachy, 1989; Ling et al., 1991).

### Prospects on potyvirus molecular biology

The seminal work by Ahlquist et al. (1984) provided RNA virologists with a novel tool to analyse these pathogens. The in vitro synthesis of biologically functional RNA transcripts from full-length cDNA clones has proved to be a powerful way to circumvent difficulties in the study of the molecular biology of RNA viruses caused by the lack of well-defined mutants and by the inability to engineer changes in their genomes. This system has recently become available for three members of the potyvirus group, TVMV (Domier et al., 1989), PPV (Riechmann et al., 1990) and ZYMV (Gal-On et al., 1991). Full-length cDNA copies of their genomes were cloned downstream from modified T7 RNA polymerase promoters and capped RNAs synthesized in vitro were infectious. On replication initiated by PPV transcripts with additional nucleotides on either end, the progeny viral RNA did not contain the extra nucleotides and regained the typical poly(A) chain length distribution of native viral RNA (Riechmann et al., 1990). Although the specific infectivities of the synthetic transcripts are rather low when compared with that of the native viral RNAs, 0.2% in the case of TVMV transcripts (Domier et al., 1989) and up to 5% in the case of PPV (Riechmann et al., 1990; unpublished results), they are good enough to analyse in vivo the effects of mutations introduced into TVMV and PPV genomes by genetic engineering (Atreya et al., 1990, 1991; Riechmann et al., 1991;
Rodriguez-Cerezo et al., 1991). As an alternative approach to the in vitro synthesis of viral RNA, it has recently been reported that potyvirus infection can be generated by plasmid DNA when a 35S promoter sequence upstream from the viral cDNA is provided (Maiss et al., 1992). A good example of the important and unexpected results that can be obtained by the use of cDNA clones is the recent discovery that a determinant of disease symptom severity is located in the 3'-terminal non-coding region of TVMV RNA (Rodriguez-Cerezo et al., 1991). Although the construction and manipulation of full-length potyviral cDNA clones has proved to be a substantial task, it is expected that this system will be developed for more potyviruses in the forthcoming years and that, coupled with protoplast systems (Luciano et al., 1987; Domier et al., 1989; Atreya et al., 1990), which have been underemployed in potyvirus research owing to experimental difficulties, and with transgenic plants expressing potyviral gene products (Berger et al., 1989; Graybosch et al., 1989), will greatly help in the further unravelling of the potyvirus life cycle.

This review illustrates how both our understanding of potyviruses and the experimental approaches to study them have enlarged and changed during the last 5 years. Functions have been proposed for the different potyviral proteins based on sequence data and some of them have been tested using in vitro and in vivo systems (Table 1). It is a recurrent suggestion that some proteins could perform more than one function and might therefore possess two-domain structures. These functions can now be studied in vivo using molecular biological techniques, since the development of cDNA clones from which infectious in vitro transcripts can be obtained has greatly increased the scope of possible experiments with which to examine the life cycle of these viruses. It is possible to modify their genomes by site-directed mutagenesis and mapping of specific mutant phenotypes should also be feasible. Lethal mutations could also be studied and a uniformity of inoculum not previously available is provided. The possibility to perform investigations that have not been possible hitherto is therefore already open. And although almost nothing is known about the regulation of the potyviral life cycle, it should also be amenable to the currently available genetic engineering techniques. It is therefore easy to predict that the next 5 years will be even more fruitful and exciting.

We are grateful to Drs E. Rodriguez-Cerezo, J. G. Shaw, J. C. Carrington, W. G. Dougherty, A. Nassuth, D. D. Shukla, B. Raccah and E. Maiss for communicating results and providing manuscripts prior to publication, and to Drs F. Rojo and F. Gebauer for valuable comments on the manuscript. Work carried out in the authors' laboratory was supported by grants BIO90-1127-E from CICYT and C168/90 from Comunidad de Madrid, and by an institutional grant from Fundación Ramón Areces. J.L.R. and S.L. were recipients of fellowships from CSIC.

References


the 5' terminus of tobacco vein mottling virus RNA. *Virology* **142**, 134–143.


