Virology **286,** 72–82 (2001)

Analysis of a Three-Dimensional Structure of *Potato leafroll virus* Coat Protein Obtained by Homology Modeling

Laurent Terradot,*[,1,2] Michel Souchet,† Vinh Tran,‡ and Danièle Giblot Ducray-Bourdin*

*Institut National de la Recherche Agronomique, UMR Biologie des organismes et des populations appliquée à la protection des plantes, Station de Pathologie Végétale, BP 35327, 35653 Le Rheu Cédex, France; †SmithKline-Beecham Laboratoires Pharmaceutiques, 4 rue du Chesnay-Beauregard, 35760 Saint Grégoire, France; and ‡Unité de Physico-Chimie des Macromolécules (UPCM), BP 71627, 44316 Nantes Cédex, France

Received December 19, 2000; returned to author for revision January 18, 2001; accepted March 5, 2001

Viruses of the family *Luteoviridae* are ssRNA plant viruses that have particles that exhibit icosahedral symmetry. To identify the residues that might be exposed on the surface of the *Potato leafroll virus* (PLRV; genus *Polerovirus,* family *Luteoviridae*) capsid, and therefore involved in biological interactions, we performed a structural analysis of the PLRV coat protein (CP) on the basis of comparisons with protein sequences and known crystal structures of CPs of other viruses. The CP of PLRV displays 33% sequence similarity with that of *Rice yellow mottle virus* (genus *Sobemovirus*) when the sequences were aligned by using the hidden Markov model method. A structure model for PLRV CP was designed by protein homology modeling, using the crystal structure of RYMV as a template. The resulting model is consistent with immunological and site-directed mutagenesis data previously reported. On the basis of this model it is possible to predict some surface properties of the PLRV CP and also speculate about the structural evolution of small icosahedral viruses.

Key Words: virus capsid; homology modeling; viral coat; structural evolution; *Luteoviridae;* circulative virus; RNA virus.

INTRODUCTION

Potato leafroll virus (PLRV) infects potato crops worldwide. It is the typical member of the genus *Polerovirus* in the *Luteoviridae* family ("luteovirids"; Smith *et al.,* 2000). Luteovirids are spherical, phloem-limited, obligatory aphid transmitted plant viruses. This family is divided into three genera: *Enamovirus, Luteovirus,* and *Polerovirus* (Mayo and D'Arcy, 1999). The three genera display serological relatedness but induce different cytopathological effects. Members of the genera *Luteovirus* and *Polerovirus* have similar positive-sense single-stranded RNA genomes of around 6000 nucleotides. Their genomes are divided into two gene blocks separated by a non-coding region. The 5′ half contains open reading frames (ORFs) that encode the proteins involved in virus replication (RNA-dependent RNA polymerase; RdRp) and the 3′ half contains ORFs that encode for structural proteins. The 3′ half is translated from a subgenomic RNA (sgRNA1) of ca. 2500 nucleotides long (for a review, see Mayo and D'Arcy, 1999). The genome of the sole species in genus *Enamovirus, Pea enation mosaic virus* (PEMV-1), resembles that of poleroviruses but multiplies only in association with PEMV-2 (now classified as an *Umbravirus*).

Comparisons of RNA sequences of viruses within the *Luteoviridae* family with other viruses have revealed surprising evolutionary relationships. An apparent dichotomy exists between the 5′ half and the 3′ half part of the genome of luteovirids. The sequences of the RdRps of viruses in genus *Luteovirus* (luteoviruses) are closer to those of the RdRps of dianthoviruses or tombusviruses than they are to those of polerovirus RdRps, which resemble with sobemovirus RdRps (Dolja and Koonin, 1991; Habili and Symons, 1989). In contrast, amino acids sequences of coat proteins (CPs) of luteoviruses and poleroviruses are strongly conserved (Mayo *et al.,* 1989; Mayo and D'Arcy, 1999) but they are distant from those of other viruses in the "Luteovirus supergroup," which consists in genus *Sobemovirus,* and family *Tombusviridae,* as well as family *Luteoviridae* (Gibbs, 1995). However, significant alignments have been obtained (Dolja and Koonin, 1991; Gibbs, 1995) and suggest that secondary structural elements of luteovirids CPs are similar to those of *Southern cowpea mosaic virus* (SCPMV, genus *Sobemovirus;* previously described as the cowpea strain of *Southern bean mosaic virus*) or *Tomato bushy stunt virus* (TBSV genus *Tombusvirus;* Dolja and Koonin, 1991; Mayo and Ziegler-Graff, 1996; Torrance, 1992).

Capsids of TBSV and SCPMV consist of 180 CP subunits assembled according to $T = 3$ quasi symmetry (Rossman *et al.,* 1983; Abad-Zapatero *et al.,* 1980). Structures of other icosahedral plant viruses have been resolved by X-ray crystallography (Bhuvaneshwari *et al.,* 1995; Canady *et al.,* 1996; Krishna *et al.,* 1999) and have

[1] Current address: Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110.

[2] To whom correspondence and reprint requests should be addressed. Fax: (314) 362-7183. E-mail: terradot@biochem.wustl.edu.

0042-6822/01 $35.00

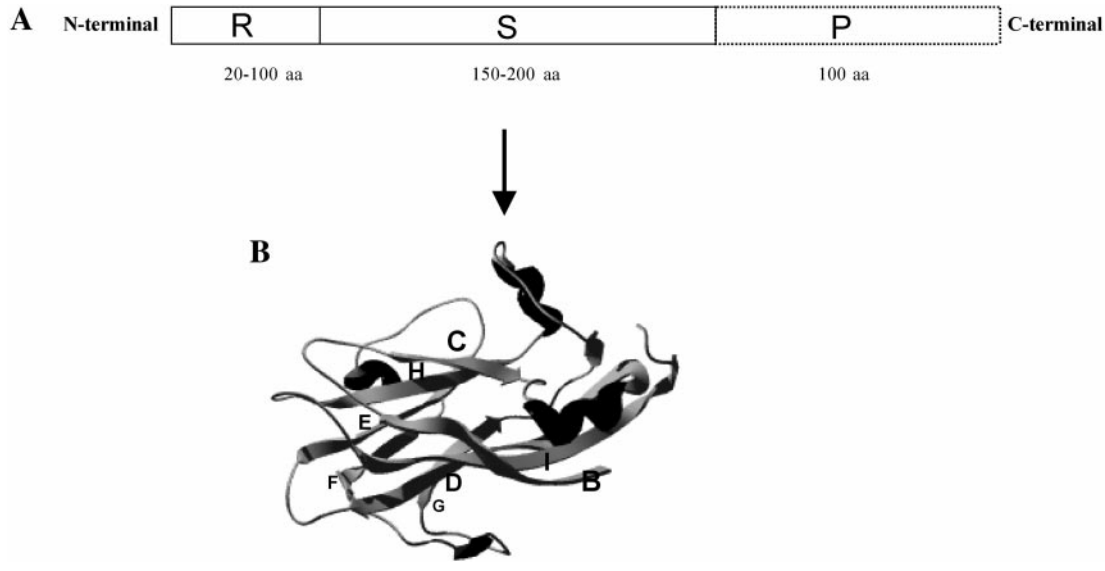


FIG. 1. Sequential distribution of the coat protein domains of small icosahedral plant viruses. (A) Schematic representation of R (for arginine rich) and S (for shell) domains of the coat protein. The box in dotted lines represents the P (for protruding) domain that is only encountered in TBSV. (B) Pov-ray rendering of the subunit A of RYMV CP illustrating the jelly roll scaffold. The nomenclature for the β-strands BIDG and CHEF follows Harrison *et al.* (1978).

revealed that the general architecture of their CP is based on two domains: the N-terminal arginine rich domain (R) and the shell domain (S) which forms the core of the capsid (Rossmann and Johnson, 1989; and Fig. 1). The R domain is found in the inner part of the capsid and is believed to interact with the virus RNA and be involved in capsid assembly (Harrison *et al.,* 1978; Abad-Zapatero *et al.,* 1980; Savithri and Erickson, 1983). The 3D structure of the S domain is very well conserved among plant and animal viruses (Rossmann and Johnson, 1989). The S domain exhibits a barrel of two β-sheets, each sheet consisting of four strands that form the so-called jelly roll structure (Harrison *et al.,* 1978). Following the nomenclature used by Harrison *et al.* (1978), the strands of the jelly roll are designated as BIDG and CHEF (Fig. 1). Two α-helices are observed, one connects strands C and D and another one is located between strands E and F.

There are no crystallographic data available for luteovirid coat proteins although the general shape of the particle is icosahedral ca. 25 nm in diameter and probably consists of 180 subunits of CP arranged in $T = 3$ symmetry according to X-ray diffraction and molecular mass analysis (Waterhouse *et al.,* 1988). PLRV CP (or major CP, 23 kDa) and a second minor structural protein, the readthrough protein (RTP), are translated from sgRNA1. RTP is the fusion product of ORF3 (CP) and ORF5 by the readthrough of the CP amber termination codon (Bahner *et al.,* 1990). Although the capsid lattice consists mainly of CP subunits assembling, some RTP are likely incorporated into the viral capsid by their CP domain, the RT domain being exposed on particle surface. This "pseudo-protruding" domain could explain why some particles seem to have small projections at the vertices (Harrison, 1984).

One of the main features common to luteovirids is that they are transmitted by aphids in the persistent mode. This implies a passage through the aphid body in which the virus particles have to cross several aphid membranes, probably mediated by receptor-mediated endocytosis (for a review, see Gildow, 1999). The capsid of the virus is thought to be involved in these interactions since luteovirids can serve as helpers for other viruses or viroids (Falk *et al.,* 1999; Querci *et al.,* 1997). Some results suggest that the CP alone can mediate virus particle passage through the gut membrane (van den Heuvel *et al.,* 1993; Chay *et al.,* 1996; Gildow, 1999). Additionally, pseudo particles of PLRV which capsid consists only in CP are able to achieve the complete route of the virus within the vector (Gildow, 1999). In contrast, it has been shown that the RTP is necessary for the effective transmission of other luteovirids, BWYV and BYDV (Brault *et al.,* 1995; Filichkin *et al.,* 1994), possibly by protecting particles during their circulation in the aphid hemolymph (van den Heuvel *et al.,* 1994) or by initiating systemic infection in inoculated plants (Mutterer *et al.,* 1999). In recent work, the strain PLRV-14.2 was shown to be differentially transmitted by various clones of aphids (Bourdin *et al.,* 1998). The low transmission rates have been related to inefficient interaction of PLRV-14.2 capsid with the gut membrane of the insect (Rouzé-Jouan *et al.,* 2001). Moreover, the PLRV-14.2 CP and RTP amino acid sequences displayed several changes at specific positions that could be involved in its poor transmissibility (Rouzé-Jouan *et al.,* 2001).

To analyze the residues exposed on the surface of the PLRV-14.2 capsid, we performed structural analysis of the PLRV CP. PLRV CP displays identity scores below 20% when compared to coat protein sequences of viruses for which 3D structure has been resolved. At this low level of sequence identity (below the so-called twi-

light zone of 30%), modeling methods based on sequence homology may fail, since accurate alignments of the target sequence with potential structure templates are difficult to generate. Pairwise comparison of sequences, such as BLAST (Altschul *et al.,* 1990) and FASTA (Pearson and Lipman, 1988), are the most commonly used methods to search for structure-function (and evolution) relationships. It has been reported that when sequence identities of related proteins are below 30%, the chance of detecting their relationships by pairwise procedures becomes increasingly small (Park *et al.,* 1998). Brenner *et al.* (1998) have shown that, among the evolutionary relationships identified from structure, sequence and function in proteins with 20–30% identities, only half can be detected by pairwise sequence comparisons. For related proteins with less than 20% identity, the proportion detected is even smaller. To overcome these limitations, search procedures based on shared characteristics of sets of related sequences have been developed. Among them, the hidden Markov model (HMM; Krogh *et al.,* 1994) has been shown to be one of the most powerful method in predicting significant alignments of protein sequences with identity as low as 15% and below (Karplus *et al.,* 1997; Bystroff *et al.,* 2000).

Here we report the results of protein homology modeling of PLRV CP based on a HMM alignment. The proposed structure model uses the structure of the CP of *Rice yellow mottle virus* (RYMV, *Sobemovirus*), resolved at 2.9 Å, as a template (Qu *et al.,* 2000). We assumed that the PLRV CP shell domain shares common features with other already described $T = 3$ icosahedral plant viruses because of their general biological and physical characteristics (small icosahedral capsids of 180 subunits, ssRNA, sobemovirus-like RdRp) (Mayo and Ziegler-Graff, 1996). The obtained model was assessed by Profiles-3D calculation (Luthy *et al.,* 1992) and is satisfactorily consistent with immunological data and mutagenesis experiments previously reported on PLRV (Torrance, 1992) or related viruses (Mutterer, 1998; Gopinath *et al.,* 1994).

RESULTS AND DISCUSSION

Comparison of icosahedral viruses by structural alignments and choice of the structure template

Only the sequence of the S domain of PLRV-14.2 CP is considered in this study. Consequently, the 59 N-terminal residues (more than 33% were arginines) have been deleted and the present analysis starts at Thr_{60}. The remaining 149 C-terminal residues have been assumed to form the S domain of the PLRV capsid subunit.

PLRV-14.2 CP has been submitted to the automated fold recognition servers SAM-T99 (Karplus *et al.,* 1998), FoldFit (Kelley *et al.,* 2000) or Bioindgu (Fisher, 2000). The best hits have been SCPMV, *Sesbania mosaic virus* (*SeMV,* genus *Sobemovirus*) and TBSV indicating that the structure of PLRV CP is close to the jelly roll structure observed in coat proteins of these viruses. Hence, we have hypothesized that the general structure of PLRV CP is similar to that of sobemovirus or tombusvirus CPs. To determine the more suitable structure template to be used for homology modeling we have compared the sequence of PLRV-14.2 CP to those of CP of other plant viruses that share biological characteristics with PLRV, namely TBSV (Hopper *et al.,* 1984), SeMV (Bhuvaneshwari *et al.,* 1995), SCPMV (Abad-Zapatero *et al.,* 1980), *Physalis mottle virus* (PhMV, genus *Tymovirus;* Krishna *et al.,* 1999), and *Turnip yellow mosaic virus* (TYMV, genus *Tymovirus;* Canady *et al.,* 1996). We have also included the structure of RYMV which has recently been determined (Qu *et al.,* 2000). Main characteristics and Brookhaven protein databank (PDB; Bernstein *et al.,* 1977) codes of CP structures of these viruses are indicated in Table 1. A structural alignment of CP shell domains was generated between these viruses and the deduced multiple sequence alignment is shown in Fig. 2. PLRV-14.2 CP shell sequence has been added to the previous multiple structural alignment and aligned with RYMV CP using HMM (SAM-T99; Karplus *et al.,* 1998). The sequence identities corresponding to this alignment are indicated in Table 1.

All virus CPs analyzed are structurally remarkably similar as previously described by several authors (Qu *et al.,* 2000; Krishna *et al.,* 1999; Opalka *et al.,* 2000; Rossmann and Johnson, 1989) displaying a low root mean square deviation (RMSD) value based on their backbone superimposition (Table 1). Secondary elements are well conserved despite the observed variation in amino acid sequences. Comparisons between each pair of viruses show that there is no more than 16% of identity among members of different families (Fig. 2 and Table 1) although they share a similar 3D scaffold of eight β-strands. TBSV CP shows a loop inserted in strand B. Sobemoviruses show two extended loops: one connects strands G and H and another one contains an α-helix which joins strands F and G (Fig. 2). TYMV and PhMV CPs exhibit shorter loops that connect β-strands than those observed in the CPs of sobemoviruses and TBSV. Comparisons made among sobemoviruses show different relatedness among RYMV, SCPMV and SeMV. SCPMV and SeMV CP structures are clearly close (Gopinath *et al.,* 1994) and this is also supported by a 64% identity between the two polypeptide sequences used in this study. In contrast, RYMV CP is structurally equivalent to SCPMV and SeMV but shows differences in sequence composition, small deletions and insertions. This results in a low sequence identity (22%) as previously noticed by Qu *et al.* (2000), and small differences in particle diameters are observed (Opalka *et al.,* 2000).

Surprisingly, the sequence alignment generated with RYMV and PLRV-14.2 displays sequence identity and similarity of 17 and 33%, respectively (Table 1 and Fig. 2). Compared to the identity percentages obtained with the other viruses, the score between PLRV-14.2 and RYMV is identical to that of TBSV and SCPMV or similar to the score

TABLE 1

Sequence Identity and Similarity between CP Sequences of Small Icosahedral Viruses Considered in This Study

	PDB entry code	Capsid size (nm)	Family genus	SeMV	SCPMV	TBSV	TYMV	PhMV	RYMV	PLRV†
SeMV	1SMV	28	*Sobemovirus*		64 (82)	16 (30)	7 (24)	8 (23)	21 (41)	10 (22)
SCPMV	4SBV	28	*Sobemovirus*	**0.62** *784*		17 (33)	7 (22)	8 (22)	22 (41)	10 (24)
TBSV	2TBV	30	*Tombusviridae* *Tombusvirus*	**1.31** *568*	**1.42** *564*		10 (28)	6 (18)	12 (25)	7 (21)
TYMV	1AUY	28	— *Tymovirus*	**1.52** *648*	**1.64** *408*	**1.54** *396*		31 (44)	9 (24)	4 (16)
PhMV	1QJZ	28	— *Tymovirus*	**1.52** *384*	**1.65** *380*	**1.76** *332*	**1.16** *600*		8 (19)	4 (16)
RYMV	na	26	*Sobemovirus*	**1.12** *648*	**1.12** *648*	**1.44** *536*	**1.58** *384*	**1.59** *352*		17 (33)
PLRV		25	*Luteoviridae* *Polerovirus*	/ /	/ /	/ /	/ /	/ /	/ /	

Note. On the upper right half of the table, numbers indicate sequence identity followed by sequence similarity in parentheses. They were obtained upon structure superimposition except for PLRV† (sequence of PLRV-14.2) which was first aligned with RYMV by the HMM method and then added to the multiple alignment. On the bottom left half of the table, numbers in bold show root mean square deviations (RMSD in Å) and the number of atoms involved in the calculation is indicated in italic. SeMV, *Sesbania mosaic virus;* SCPMV, *Southern cowpea mosaic virus;* TBSV, *Tomato bushy stunt virus;* TYMV, *Turnip yellow mosaic virus;* PhMV, *Physalis mottle virus;* RYMV, *Rice yellow mottle virus;* PLRV, *Potato leafroll virus.* na, not available.

obtained with TBSV and SeMV. To determine the regions where RYMV and PLRV-14.2 sequences are the more divergent, the secondary elements of RYMV structure have been compared to those predicted for PLRV-14.2. Secondary structure prediction obtained by PROF (Ouali and King, 2000) was retained in the present work although other methods such as JPRED, PHD and PSI PRED gave very similar results (data not shown). The PLRV-14.2 CP shell domain was predicted to contain mainly β-strands and also two short α-helices (Fig. 2). The HMM alignment shows that all β-strands are correctly aligned to those of RYMV as are the helices α_{C-D} and α_{E-F} but with slight differences in length. Secondary structure predictions are similar to those reported by Mayo and Ziegler-Graff (1996) although the alignment with RYMV yields a different allocation of the strands. This discrepancy is likely due to the difference in the methods used to assign strands. Mayo and Ziegler-Graff (1996) have determined strands position by analogy with other virus CPs whereas in the present study, we have used an HMM alignment with RYMV CP. Moreover, the distribution of strands we have obtained is in agreement with that reported by Dolja and Koonin (1991) with an alignment between S domain of PLRV and TBSV CPs. It is also consistent with the comparison of $T = 3$ viruses highlighting that β-strands D, H, and I are particularly well aligned (Fig. 2). Differences in length of secondary structure elements between RYMV and PLRV-14.2 are not significant. Indeed, it is known that the core of a specified β-strand is predicted by the secondary structure prediction methods with an accuracy of more than 70% whereas it is not the case of its extremity (Rost and Sander, 2000). Therefore, it was assumed that the structure of PLRV CP shares structural features of other sobemovirus CPs. RYMV has been chosen as a structure template because it shows the closest amino acid sequence to PLRV CP. It was also assumed that, since RYMV is neither available in fold libraries nor in PDB, automated fold recognition servers used did not detect it as a suitable template for homology modeling.

PLRV CP: Model building and model assessment

Five models have been generated with the Modeler methodology using the subunit A of RYMV structure. Slight differences between the models have been observed on the helix connecting the β-strands E and F (Fig. 3). Assessment of the models with the Profiles-3D method (Luthy *et al.,* 1992) has led to fairly high score values for the five models (Fig. 3). Moreover, two models displayed no violations according to Profiles-3D when assessed with the more restricting window (10 residues). Regarding to the score values obtained with these models, all have been considered as equally valid (RMSD $<$ 0.90 for 138 C_α) and the model showing the highest value, PLRV3, was chosen (Fig. 3).

Although such a model cannot be compared to a 3D crystal structure (less accurate at the atomic level), it can give helpful insight of CP structural shape (up to residue level) to understand virus properties. Molecular modeling has proved valuable in numerous instances, including structures belonging to the immunoglobulin superfamily, which also consist of β-barrels. In the latter case, as in our study, sequence identity obtained with the template was lower than what is usually considered suitable for reliable model building (Bajorath and Linsley, 1997). However, comparison of the model with the experimentally determined structure has shown that the overall accuracy of the T cell receptor CD152 was sound and sufficient for a meaningful application of the model in experimental studies (Bajorath *et al.,* 1997).

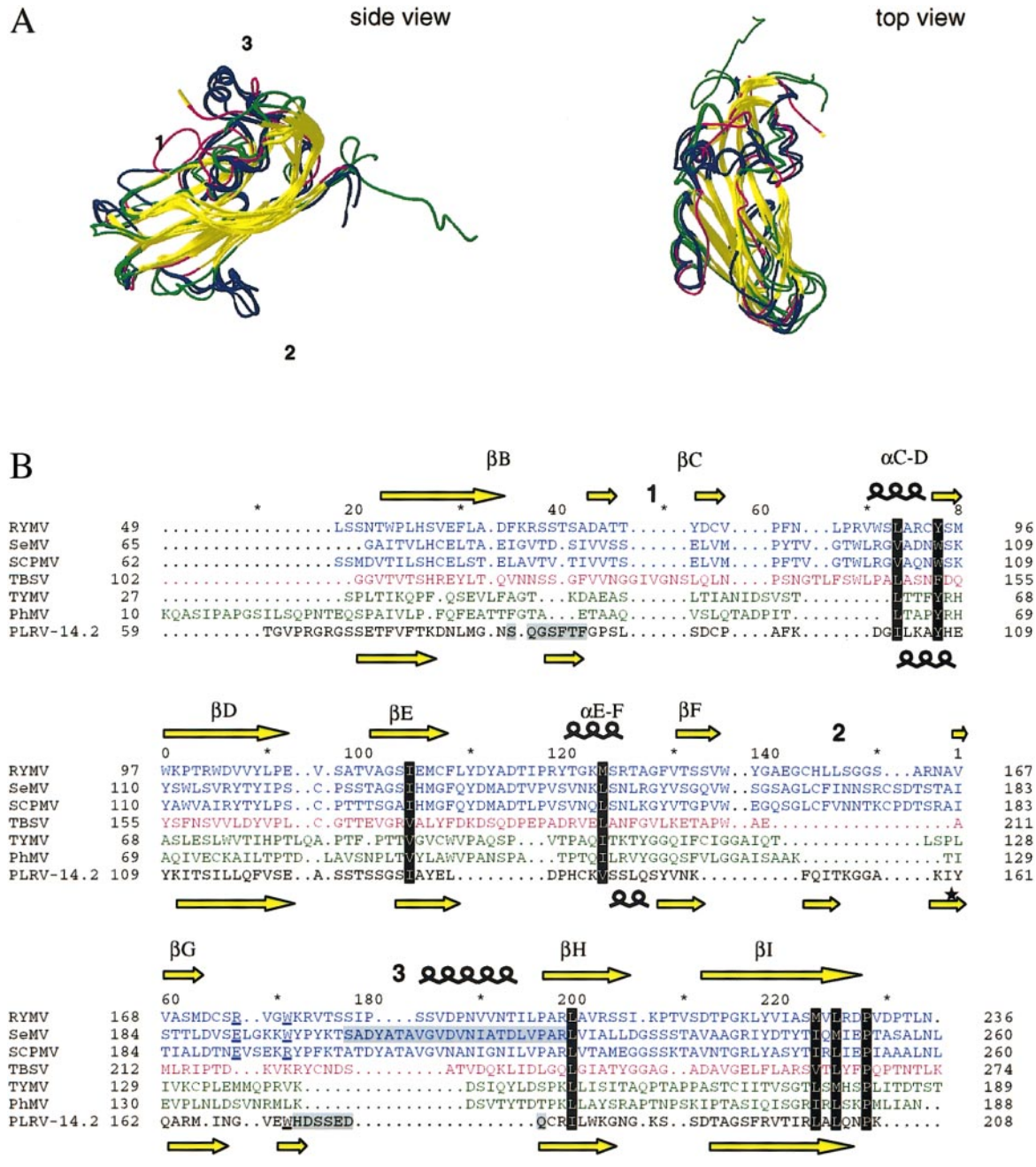
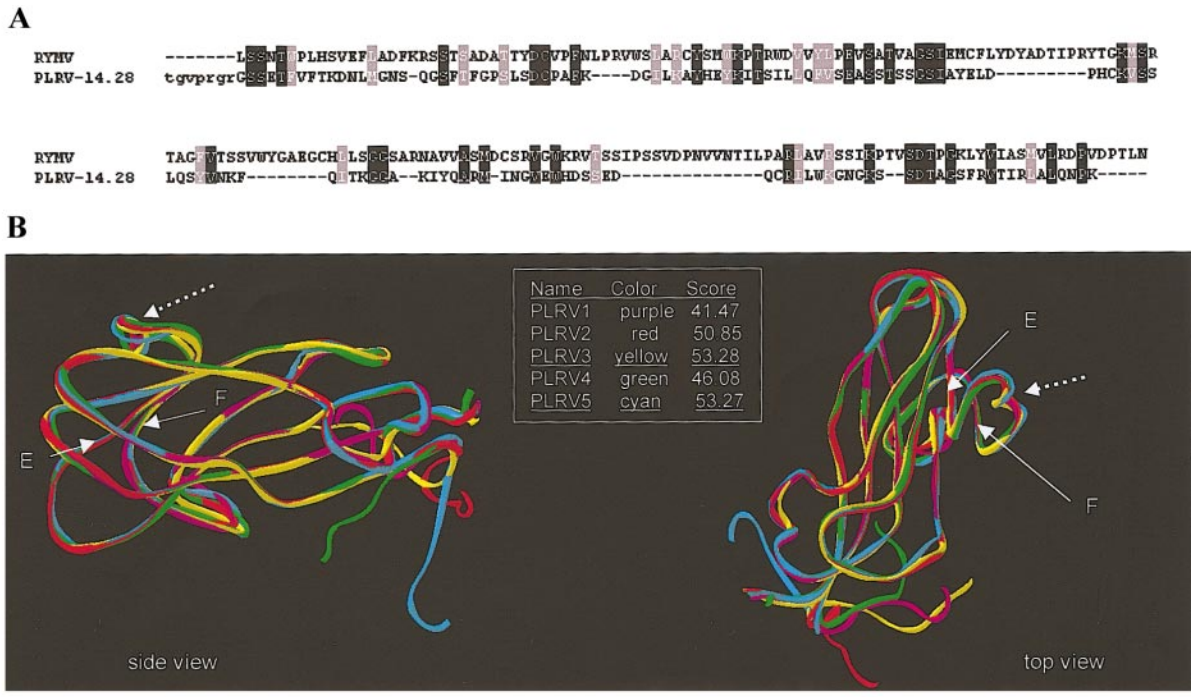


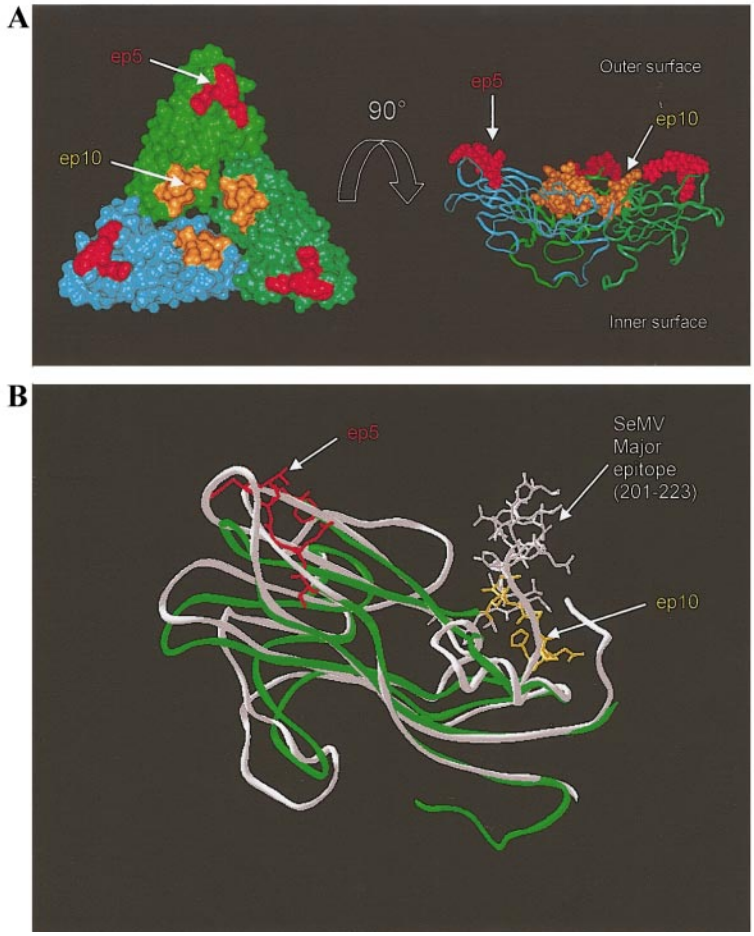
FIG. 2. Structural alignment of coat protein shell domains of subunit A for viruses belonging to genera *Sobemovirus* (RYMV, SCPMV, and SeMV), *Tymovirus* (PhMV and TYMV), *Tombusvirus* (TBSV) and resulting sequence comparisons with PLRV-14.2. (A) Two different views of the solid ribbons representation of CPs backbone superimposition. β-Strands are indicated by yellow flat ribbons. Besides β-strands, Cα trace of structures are indicated in blue, green and magenta for sobemoviruses, tymoviruses, and TBSV, respectively. Numbers indicate additional features for some structures. Label 1 points out an insertion in a TBSV β-strand whereas labels 2 and 3 indicate additional α-helices in sobemoviruses. Amino acids corresponding to these features are indicated in B. (B) Multiple alignment deduced from 3D structure superimposition. Conserved residues are shaded in black according to the Blosum62 matrix (Henikoff and Henikoff, 1992). β-Strands and α-helices of RYMV and PLRV are indicated with yellow arrows and repeats of black circles, respectively. These secondary structure elements correspond to those observed in RYMV_A crystal structure or in the case of PLRV-14.2, they have been predicted with PROF (Ouali and King, 2000). Segments corresponding to PLRV and SeMV epitopes are shaded in gray. Underlined residues correspond to those putatively involved in subunit interactions (see also Fig. 5). Star indicate I_{101}, a typical residue of PLRV-14.2 (Rouzé-Jouan *et al.,* 2001).

Consistency of the model structure with biological data

Data currently available on PLRV and SeMV surface epitopes were applied to the chosen model. Pepscan analysis, used previously to map continuous epitopes on the PLRV capsid, suggested that two PLRV isolates shared several epitopes, two of them being particularly antigenic (Torrance, 1992). These epitopes consist of



3



4

FIG. 3. Model building and assessment. (A) Alignment of the RYMV and PLRV CP sequences used to generate the five models. Identical and similar residues are shaded in black and gray, respectively. (B) Backbone superimposition of the five models constructed with Modeler method using the template-sequence alignment. For each model structure, color and Profiles-3D calculation score are indicated in the center inset and the two models that generate no violation are underlined. White dotted arrows indicate the loop between β-strands E and F that was the most variable.

FIG. 4. Epitope mapping on a reconstituted trimer of PLRV CPs. (A, left) Representation of a solid surface of a reconstituted trimer of PLRV-14.2 capsid after superimposition on a reported crystallized RYMV trimer assembly. (A, right) Corresponding ribbon representation of the three subunits after a 90° rotation along the *Y* axis. PLRV CP subunits are in light green, medium green, and light blue. Epitope 5 (residues 83 to 89) is in red and epitope 10 (residues 172 to 178) is in orange. (B) Ribbon representation of the structural superimposition of the PLRV CP (green) with the SeMV subunit_A (white). All residue side chains composing the epitopes are displayed in a stick rendering. PLRV epitopes have the same colors as in A and the major epitope of SeMV is in white.

residues 83 to 89 (TQGSFTF, ep5) and of residues 172 to 178 (HDSSEDQ, ep10), respectively (Fig. 2). In the PLRV-14.2 model, the corresponding sequences of ep5 and ep10 are SQGSFTF and HDSSEDQ, respectively. They have been mapped onto a reconstituted trimer of PLRV-14.2 CPs obtained by superimposing three PLRV-14.2 CPs onto a crystallized assembly of three RYMV capsid subunits (RYMV_A, _B, and _C). As hypothesized by Torrance (1992) these two epitopes are located on two loops that are surface accessible in the case of our model structure (Fig. 4A). Ep5 is located on the loop connecting the β-strands B and C that has been described as an important antigenic site on the VP1 subunits of poliovirus (Hogle and Filman, 1989). Ep10 is located on the loop connecting β-strands G and H that is situated at the outer surface which makes it easily accessible. The latter structural feature is different from that reported by Mayo and Ziegler-Graff (1996) where ep10 is located between strands F and G, and is predicted not to be surface accessible. Moreover, Gopinath *et al.,* (1994) have identified a major epitope, consisting of residues 201 to 223, on SeMV capsid by using monoclonal antibodies (Fig. 2). The SeMV structure has then been superimposed onto the PLRV model and the major epitope position has been compared to PLRV ones. Interestingly, the geometry comparison has resulted in the co-location of the ep10 of PLRV and SeMV epitope on the G-H loop (Fig. 4B). These data also support the proposed model structure of PLRV.

Beet western yellows virus (BWYV, genus *Polerovirus,* family *Luteoviridae*) and PLRV are very closely related viruses and their CP sequences are more than 62% identical (Mayo and D'Arcy, 1999). Mutagenesis experiments have shown that when W_{166} of BWYV (corresponding to W_{171} in PLRV) is mutated to R, viral encapsidation is inhibited (Mutterer, 1998). Several contacts between subunits at two-, three-, and fivefold axis enable particles folding and stability in $T = 3$ viruses (for further details, see Johnson and Speir, 1997). In SeMV, one of these interactions occurs at the center of the trimer (Fig. 5A). Residue W_{196} of subunits A, B, and C interacts with E_{191} of subunits C, A, and B, respectively (Fig. 5). Similarly, SCPMV R_{196} interacts with E_{191} whereas in RYMV, W_{178} interacts with R_{175} (Fig. 5B). In our model structure, W_{171} corresponds topologically to W_{178} of RYMV, W_{196} of SeMV, or R_{196} of SCPMV (Figs. 2 and 5B). This suggests that W_{171} is involved in a similar stabilizing effect between subunits. That may explain that when W_{166} is mutated to R in the BWYV sequence, interactions between BWYV subunits are not sufficient for an efficient encapsidation (Mutterer, 1998).

Surface study of PLRV CP and Inter- and Intrafamily comparisons

PLRV-14.2 has been shown to differ from readily transmitted isolates of PLRV by changes at two sites in the CP, the first site being located in the R domain. The second site shows an isoleucine at position 160 instead of a threonine (Rouzé-Jouan *et al.,* 2001). On the basis of the deduced structure, I_{160} (Fig. 2) is believed to be buried in β-strand G and thus not surface accessible (Fig. 6). Therefore, although we cannot rule out the possible involvement of this change in capsid stability or RNA binding, it is unlikely that it can explain the PLRV-14.2 phenotype by any difference in capsid affinity for an aphid receptor.

To compare the surface properties of the PLRV-14.2 CP with those of SCPMV, RYMV, and PhMV, the structure surface of the trimer was colored according to the properties of amino acid residues (Fig. 6). The PLRV-14.2 model displays an interesting patch of acidic residues at the center of the trimer. This "acidic patch" is made of residues E_{109}, E_{170}, E_{176}, D_{173}, and D_{177} on the model. Sequences of other poleroviruses and luteoviruses have been mapped onto the PLRV-14.2 model structure, assuming that other luteovirid CPs (Mayo and D'Arcy, 1999) have structures similar to that of PLRV CP (as they share more than 60% of sequence identities). As shown in Fig. 6, the acidic patch is well conserved within the poleroviruses *Cucurbit aphid-borne yellows virus* (CABYV) and BWYV. In contrast, the acidic property of the patch is reduced when sequences of luteoviruses *Barley yellow dwarf virus-MAV* (BYDV-MAV, genus *Luteovirus,* family *Luteoviridae*) and BYDV-PAV (genus *Luteovirus,* family *Luteoviridae*) are mapped onto the PLRV CP model. Interestingly, this common feature of luteovirids is located at the center of the trimer which exhibits highly accessible residues such as epitope 10. As the CP surface is involved in several parts of virus life cycle, we suggest that this patch may be important for molecular recognition and could confer specific properties for each genus and/or species within *Luteoviridae*.

Structural evolution

Three-dimensional biological structures are generally better conserved than are protein sequences (Rossmann and Johnson, 1989). Viruses in genera *Tombusvirus* (TBSV), *Sobemovirus* (RYMV, SCPMV and SeMV), and *Tymovirus* (PhMV and TYMV) exemplify this observation (Gopinath *et al.,* 1994; Krishna *et al.,* 1999; Rossmann and Johnson, 1989). The major differences between homologous structures usually correspond to deletions and insertions of residues, while the essential core of the polypeptide fold is maintained. Interestingly, Krishna *et al.* (1999) have observed differences mainly in the length of loops connecting β-strands between the structures of the CP of the tymovirus PhMV and the sobemovirus SeMV. Our results suggest that RYMV and PLRV CPs have a common ancestor and that divergence between these proteins may have arisen because of deletion events at loops connecting highly conserved β-strands, as seems to be the case for PhMV or TYMV. Argos (1981)

used secondary structure prediction methods to propose that, despite a lack of sequence homology, the 3D structure of TYMV CP was similar to that of TBSV and SCPMV CPs and that these viruses may have a common ancestor. Fifteen years later, the TYMV structure was resolved (Canady *et al.,* 1996) and was shown to be in good agreement with the earlier prediction (Argos, 1981).

Gibbs (1995) has proposed that poleroviruses and sobemoviruses are related in the 5′ half of their genome but that the module encoding structural genes has come from different ancestors. This view is supported by greater sequence homologies between poleroviruses and sobemoviruses in the RdRp than in other virus proteins (Dolja and Koonin, 1991; Habili and Symons, 1989). However, our findings suggest that the CP genes of PLRV and RYMV are probably more related than previously believed. Comparison of their sequences has revealed that PLRV is not more divergent from RYMV (17% identity and 33% similarity) than is TBSV from SCPMV (17% identity and 33% similarity) or RYMV from SCPMV (21% identity and 41% similarity; this study and Opalka *et al.,* 2000). Thus, the discrepancy between the 5′ half and the 3′ half part of luteovirid genomes may also reflect more the different selection pressures than any different origins. While virus RdRp genes need only to adapt to the host plant, structural proteins of these viruses must be adapted both to the host plant and to the insect vector. Therefore, it seems reasonable to assume that the structural genes have evolved independently from the RdRp gene. These assumptions are supported by several observations of other RNA viruses in which coat protein genes are more divergent than any other virus gene (Rossmann and Johnson, 1989; and references cited therein).

In summary, data reported in the present work show a good agreement between the model structure for PLRV-14.2 CP and the results of immunological and mutagenesis studies. Therefore the methodology used here provides a suitable approach to gain insight into the capsid of PLRV. Possibly it may also be used to generate useful 3D models for other $T = 3$ virus CPs.

MATERIALS AND METHODS

Virus strain and sequence

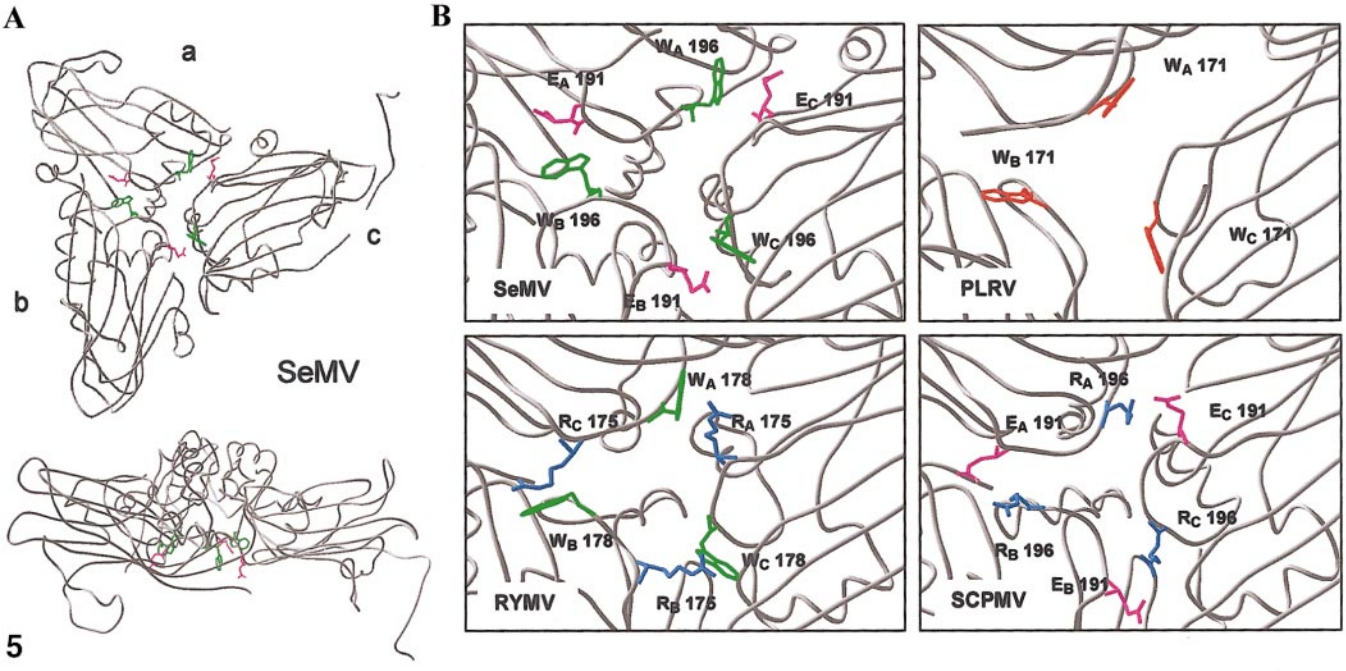
A collection of PLRV strains originating from different countries has been maintained for years on *Physalis floridana* (Rydb.) through vegetative propagation (cuttings). The isolate PLRV-14.2, collected in the North of France on potato in 1985 was used in this study and has already been described elsewhere (Bourdin *et al.,* 1998; Rouzé-Jouan *et al.,* 2001). Total RNAs from infected plant tissues were extracted using the RNeasy plant Mini kit (Qiagen), according to manufacturer's instructions. For cDNA synthesis, 10 μl of total RNAs was used. Reverse transcription was primed with an oligonucleotide complementary to nucleotides CCAAATAGGTAGACTCCG and a PCR product of approximately 630 bp was synthesized with the forward primer GTTAATGAGTACGGTCGT (underlined letters indicate the start and the stop codons, respectively). The PCR product containing ORF3 was cloned into pCR-script vector (Stratagene). Four hundred nanograms of the clone pCRPLRV14.28 was used as matrix for sequencing reactions (ABI Prism Big Dye dRhodamine terminator cycle Sequencing Ready Reaction kit, Perkin–Elmer Biosystems) using the primers T3 and T7. Sequence was performed in a ABI prism 310 (Perkin–Elmer). The accession number of the PLRV-14.28 CP is AY007727. This sequence displays one particular change with the original PLRV-14.2 isolate (Rouzé-Jouan *et al.,* 2001) resulting in leucine instead of a valine at position 202.

Protein sequence analysis

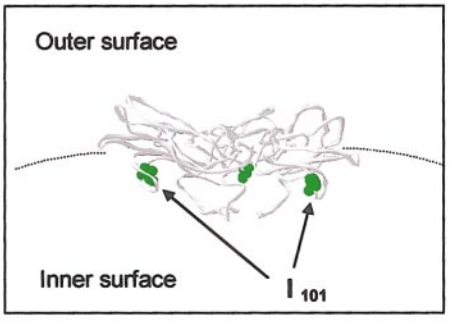
Protein sequence analysis was achieved with the program BioEdit (Hall, 1999) and GeneDoc version 2.6 available at www.cris.com/~ketchup/genedoc.htmls. The SAM-T99 HMM server version 3.0 (http://www.cse.ucsc.edu/resesarch/compbio/HMM-apps) was used to generate the alignment of PLRV with RYMV (Karplus *et al.,* 1998). Secondary structure prediction of PLRV CP was carried out by several methods including JPRED (Cuff *et al.,* 1998), PROF (Ouali and King, 2000), PHD (Rost *et al.,* 1994), and psi-pred (Altschul *et al.,* 1997).

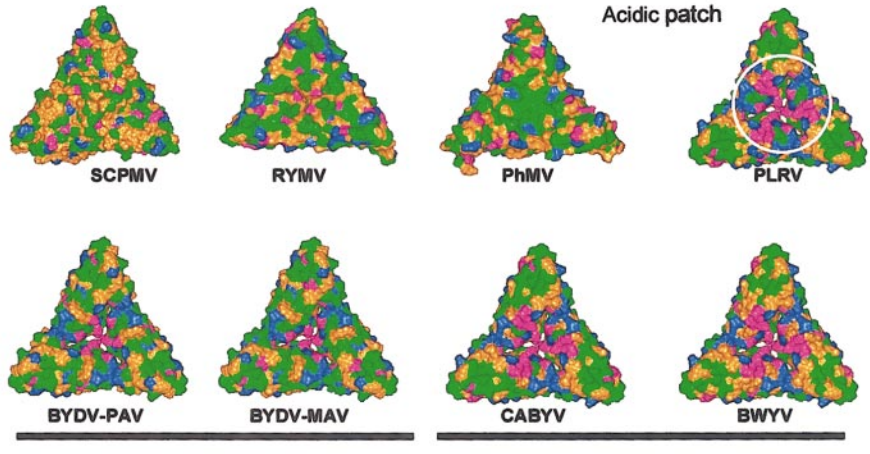
3D protein structure

Structure superpositions and 3D manipulations were done with Swiss-PdbViewer version 3.6b2 (available at http://www.expasy.ch/spdb/mainpage.html) and InsightII 980 package (Molecular Simulations Inc., San Diego, CA). The alignment between PLRV and RYMV CPs generated by SAM-T99 has been finally slightly modified manually to match residues of PLRV according to the beginning or end of secondary structure elements. PLRV CP 3D model was generated by the Modeler module of the InsightII 980 package using the high level of refinement (Sali and Blundell, 1990; Sali *et al.,* 1995). Coordinates of the structure template RYMV have been obtained from V. S. Reddy and J. E. Johnson (Qu *et al.,* 2000). Model assessment was carried out following the method of Profiles-3D implemented in InsightII (Luthy *et al.,* 1992) using a 10-residue window. Minimum and maximum theoretical values obtained for PLRV-14.2 sequence have been 28.9 and 64.23, respectively. RMSD have been calculated with Swiss-PdbViewer3.6b2 from backbone superimposition. Connoly surfaces of protein and assembly of multimer were performed with the InsightII package. To color surface of other members of the *Luteoviridae* family, an alignment has been generated by ClustalW as sequences used displayed around 60% identities (Mayo and D'Arcy, 1999). EMBL database accession numbers for BWYV, CABYV, BYDV-PAV, and BYDV-MAV are X13062, X76931, X07653, and D01213, re-



5





6

Luteovirus *Polerovirus*

FIG. 5. Location of important residues putatively involved in the interaction of CP subunits of sobemoviruses and comparison with the PLRV model structure. (A) Two different views of a gray solid ribbon representation of the SeMV CP trimer (subunits_A, _B, and _C in interaction). Side chains of important residues are displayed in colored stick. (B) Comparison of SeMV with RYMV, SCPMV, and PLRV in the region of subunit interaction. All structures are represented with a gray solid ribbons. For each virus, heavy atoms of residues putatively involved in subunits interactions are displayed in a stick rendering. Tryptophane, arginine, and glutamate residues are indicated in green, blue, and magenta, respectively, except for PLRV where W171 is in red.

FIG. 6. Surface study of PLRV capsid and comparison with other small icosahedral viruses. In the inset, backbone of a reconstituted trimer of PLRV CP is represented with a white ribbon. Heavy atoms of I101 in each subunit are displayed in green balls that enable to show the residue in the inner part of the capsid. Surfaces of CP trimeric structures are colored according to residue properties: neutral and hydrophobic (A, V, L, I, P, M, F, and W) in orange, neutral and polar (G, S, T, C, N, Q, and Y) in green, acidic (D and E) in magenta, and basic (R, H, and K) in blue. PhMV, SCPMV, and RYMV images were generated from their PDB coordinates. PLRV image corresponds to the trimeric assembly of the PLRV-14.2 CP structure model. A white circle indicates the patch of acidic residues at the center of the trimeric assembly. For luteoviruses and poleroviruses, the PLRV structure model was used as a template.

spectively. Therefore, the PLRV model structure was used as a template to color residues according to each of the CP sequences aligned.

ACKNOWLEDGMENTS

We are indebted to C. Qu, V. S. Reddy, and J. E. Johnson for providing us with the coordinates of the RYMV 3D structure as well as information about the RYMV structure. We also thank Dr. Mike A. Mayo for his helpful contribution in reviewing the manuscript. Laurent Terradot was supported by an "INRA-Région Bretagne" grant.

REFERENCES

Abad-Zapatero, C., Abdel-Meguid, S. S., Johnson, J. E., Leslie, A. G. W., Rayment, I., Rossmann, M. G., Sick, D., and Tsukihara, T. (1980).

Structure of southern bean mosaic virus at 2.8 Å resolution. *Nature* **286,** 33–39.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402.

Argos, P. (1981). Secondary structure prediction of plant virus coat proteins. *Virology* **110,** 55–62.

Bahner, I., Lamb, J., Mayo, M. A., and Hay, R. T. (1990). Expression of the genome of potato leafroll virus: Readthrough of the coat protein termination codon *in vivo*. *J. Gen. Virol.* **71,** 2251–2256.

Bajorath, J., and Linsley, P. S. (1997). Molecular modeling of Immunoglobulin Superfamily Proteins: Predicting the three-dimensional structure of the extracellular domain of CTLA-4 (CD152). *J. Mol. Model.* **3,** 117–123.

Bajorath, J., Linsley, P. S., and Metzler, W. J. (1997). Molecular modeling of Immunoglobin Superfamily Proteins: CTLA-4 (CD152)—Comparison of a predicted and experimentally determined three-dimensional structure. *J. Mol. Model.* **3,** 287–293.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977). The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112,** 535.

Bhuvaneshwari, M., Subramanya, H. S., Gopinath, K., Savithri, H. S., Nayudu, M. V., and Murthy, M. R. (1995). Structure of sesbania mosaic virus at 3 Å resolution. *Structure* **3,** 1021–1030.

Bourdin, D., Rouzé, J., Tanguy, S., and Robert, Y. (1998). Variation among clones of *Myzus persicae* (Sulzer) and *Myzus nicotianae* (Blackman) in the transmission of a poorly- and a highly-aphid transmissible isolate of potato leafroll luteovirus (PLRV). *Plant Pathol.* **47,** 794–800.

Brault, V., van den Heuvel, J. F. J. M., Verbeek, M., Ziegler-Graff, V., Reutenauer, A., Herrbach, E., Garaud, J. C., Guilley, H., Richards, K., and Jonard, G. (1995). Aphid transmission of beet western yellows luteovirus requires the minor capsid read-through protein P74. *EMBO J.* **14,** 650–659.

Brenner, S. E., Chothia, C., and Hubbard, T. J. (1998). Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships. *Proc. Natl. Acad. Sci. USA* **95,** 6073–6078.

Bystroff, C., Thorsson, V., and Baker, D. (2000). HMMSTR: A hidden Markov model for local sequence–structure correlations in proteins. *J. Mol. Biol.* **301,** 173–190.

Canady, M. A., Larson, S. B., Day, J., and McPherson, A. (1996). Crystal structure of turnip yellow mosaic virus. *Nat. Struct. Biol.* **3,** 771–781.

Chay, C. A., Gunasinge, U. B., Dinesh-Kumar, S. P., Miller, W. A., and Gray, S. M. (1996). Aphid transmission and systemic plant infection determinants of barley yellow dwarf luteovirus-PAV are contained in the coat protein readthrough domain and 17-kDa protein, respectively. *Virology* **219,** 57–65.

Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M., and Barton, G. J. (1998). JPred: A consensus secondary structure prediction server. *Bioinformatics* **14,** 892–893.

Dolja, V. V., and Koonin, E. V. (1991). Phylogeny of capsid proteins of small icosahedral RNA plant viruses. *J. Gen. Virol.* **72,** 1481–1486.

Falk, B. W., Tian, T., and Yeh, H. H. (1999). Luteovirus-associated viruses and subviral RNAs. *In* "Satellites and Defective Viral RNAs" (P. K. Vogt and A. O. Jackson, Eds.), pp. 159–175. Springer-Verlag, Berlin, Heidelberg.

Filichkin, S. A., Lister, R. M., McGrath, P. F., and Young, M. J. (1994). *In vivo* expression and mutational analysis of the barley yellow dwarf virus readthrough gene. *Virology* **205,** 290–299.

Fisher, D. (2000). Hybrid fold recognition: Combining sequence derived properties with evolutionary information. *Pacific Symp. Biocomput. Hawaii,* 119–130.

Gibbs, M. (1995). The luteovirus supergroup: rampant recombination and persistent partnerships. *In* "Molecular Basis of Virus Evolution" (A. J. Gibbs, C. H. Calisher, and F. Garcia-Arenal, Eds.), pp. 351–368. Cambridge Univ. Press, Cambridge.

Gildow, F. E. (1999). Luteovirus transmission mechanisms regulating vector specificity. *In* "The Luteoviridae" (H. G. Smith and H. Barker, Eds.), pp. 88–111. CABI Publishing, CAB International, Wallingford.

Gopinath, K., Sundareshan, S., Bhuvaneswari, M., Karande, A., Murthy, M. R., Nayudu, M. V., and Savithri, H. S. (1994). Primary structure of sesbania mosaic virus coat protein: its implications to the assembly and architecture of the virus. *Indian J. Biochem. Biophys.* **31,** 322–328.

Habili, N., and Symons, R. H. (1989). Evolutionary relationship between luteoviruses and other RNA plant viruses based on sequence motifs in their putative RNA polymerases and nucleic acid helicases. *Nucleic Acids Res.* **17,** 9543–9555.

Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41,** 95–98.

Harrison, B. D. (1984). Potato leafroll virus. "CMI/AAB Descriptions of Plant Viruses," No. 291, 6 pp.

Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K., and Bricogne, G. (1978). Tomato bushy stunt virus at 2.9 Å resolution. *Nature* **276,** 368–373.

Henikoff, S., and Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* **89,** 10915–10919.

Hogle, J. M., and Filman, D. J. (1989). The antigenic structure of poliovirus. *Philos. Trans. R. Soc. London B Biol. Sci.* **323,** 467–478.

Hopper, P., Harrison, S. C., and Sauer, R. T. (1984). Structure of tomato bushy stunt virus. V. Coat protein sequence determination and its structural implications. *J. Mol. Biol.* **177,** 701–713.

Johnson, J. E., and Speir, J. A. (1997). Quasi-equivalent viruses: A paradigm for protein assemblies. *J. Mol. Biol.* **269,** 665–675.

Karplus, K., Barrett, C., and Hughey, R. (1998). Hidden Markov models for detecting remote protein homologies. *Bioinformatics* **14,** 846–856.

Karplus, K., Sjolander, K., Barrett, C., Cline, M., Haussler, D., Hughey, R., Holm, L., and Sander, C. (1997). Predicting protein structure using hidden Markov models. *Proteins* **1,** 134–139.

Kelley, L. A., MacCallum, R. M., and Sternberg, M. J. (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.* **299,** 499–520.

Krishna, S. S., Hiremath, C. N., Munshi, S. K., Prahadeeswaran, D., Sastri, M., Savithri, H. S., and Murthy, M. R. (1999). Three-dimensional structure of *Physalis mottle virus:* Implications for the viral assembly. *J. Mol. Biol.* **289,** 919–934.

Krogh, A., Brown, M., Mian, I. S., Sjolander, K., and Haussler, D. (1994). Hidden Markov models in computational biology. Applications to protein modeling. *J. Mol. Biol.* **235,** 1501–1531.

Luthy, R., Bowie, J. U., and Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature* **356,** 83–85.

Mayo, M. A., and D'Arcy, C. J. (1999). Family *Luteoviridae:* A reclassification of Luteoviruses. *In* "The Luteoviridae" (H. G. Smith and H. Barker, Eds.), pp. 15–22. CABI Publishing, CAB International, Wallingford.

Mayo, M. A., Robinson, D. J., Jolly, C. A., and Hyman, L. (1989). Nucleotide sequence of potato leafroll luteovirus RNA. *J. Gen. Virol.* **70,** 1037–1051.

Mayo, M. A., and Ziegler-Graff, V. (1996). Molecular biology of luteoviruses. *Adv. Virus Res.* **46,** 413–460.

Mutterer, J. (1998). "Etude des déterminants viraux impliqués dans le mouvement et la transmission du virus des jaunisses occidentales de la betterave ou BWYV." Thèse de Doctorat, Univ. Louis Pasteur, Strasbourg.

Mutterer, J. D., Stussi-Garaud, C., Michler, P., Richards, K. E., Jonard, G., and Ziegler-Graff, V. (1999). Role of the beet western yellows virus readthrough protein in virus movement in *Nicotiana clevelandii*. *J. Gen. Virol.* **80,** 2771–2778.

Opalka, N., Tihova, M., Brugidou, C., Kumar, A., Beachy, R. N., Fauquet, C. M., and Yeager, M. (2000). Structure of native and expanded sobemoviruses by electron cryo-microscopy and image reconstruction. *J. Mol. Biol.* **303,** 197–211.

Ouali, M., and King, R. D. (2000). Cascaded multiple classifiers for secondary structure prediction. *Protein Sci.* **9,** 1162–1176.

Park, J., Karplus, K., Barrett, C., Hughey, R., Haussler, D., Hubbard, T., and Chothia, C. (1998). Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods. *J. Mol. Biol.* **284,** 1201–1210.

Pearson, W. R., and Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85,** 2444–2448.

Qu, C., Liljas, L., Opalka, N., Brugidou, C., Yeager, M., Beachy, R. N., Fauquet, C. M., and Johnson, J. E. (2000). 3D domain swapping modulates the stability of members of an icosahedral virus group. *Structure* **8,** 1095–1103.

Querci, M., Owens, R. A., Bartolini, I., Lazarte, V., and Salazar, L. F. (1997). Evidence for heterologous encapsidation of potato spindle tuber viroid in particles of potato leafroll virus. *J. Gen. Virol.* **78,** 1207–1211.

Rossmann, M. G., and Johnson, J. E. (1989). Icosahedral RNA virus structure. *Annu. Rev. Biochem.* **58,** 533–573.

Rossmann, M. G., Abad-Zapatero, C., Murthy, M. R., Liljas, L., Jones, T. A., and Strandberg, B. (1983). Structural comparisons of some small spherical plant viruses. *J. Mol. Biol.* **165,** 711–736.

Rost, B., and Sander, C. (2000). Third generation prediction of secondary structures. *Methods Mol. Biol.* **143,** 71–95.

Rost, B., Sander, C., and Schneider, R. (1994). PHD—An automatic mail server for protein secondary structure prediction. *Comput. Appl. Biosci.* **10,** 53–60.

Rouzé-Jouan, J., Terradot, L., Pasquer F., Tanguy, S., and Giblot Ducray-Bourdin, D. (2001). The passage of *Potato leafroll virus* through *Myzus persicae* gut membrane regulates transmission efficiency. *J. Gen. Virol.* **82,** 17–23.

Sali, A., and Blundell, T. L. (1990). Definition of general topological equivalence in protein structures. A procedure involving comparison of properties and relationships through simulated annealing and dynamic programming. *J. Mol. Biol.* **212,** 403–428.

Sali, A., Potterton, L., Yuan, F., van Vlijmen, H., and Karplus, M. (1995). Evaluation of comparative protein modeling by MODELLER. *Proteins* **23,** 318–326.

Savithri, H. S., and Erickson J. W. (1983). The self assembly of the cowpea strain of southern bean mosaic virus: Formation of the T = 1 and T = 3 nucleoprotein particles. *Virology* **126,** 328–335.

Smith, G. R., Borg, Z., Lockhart, B. E., Braithwaite, K. S., and Gibbs, M. J. (2000). Sugarcane yellow leaf virus: A novel member of the *Luteoviridae* that probably arose by inter-species recombination. *J. Gen. Virol.* **81,** 1865–1869.

Torrance, L. (1992). Analysis of epitopes on potato leafroll virus capsid protein. *Virology* **191,** 485–489.

van den Heuvel, J. F., Verbeek, M., and Peters, D. (1993). The relationship between aphid-transmissibility of potato leafroll virus and surface epitopes of the viral capsid. *Phytopathology* **83,** 1125–1129.

van den Heuvel, J. F., Verbeek, M., and van der Wilk, F. (1994). Endosymbiotic bacteria associated with circulative transmission of potato leafroll virus by *Myzus persicae*. *J. Gen. Virol.* **75,** 2559–2565.

Waterhouse, P. M., Gildow, F. E., and Johnstone, G. R. (1988). Luteovirus group. *In* "CMI/AAB Descriptions of Plant Viruses," No. 339, 9 pp.