Comparison of the predicted secondary structure of aphid transmission factor for transmissible and non-transmissible cauliflower mosaic virus strains

N. Modjtahedi*, M. Volovitch, L. Mazzolini+ and P. Yot+

Section de Biologie, Institut Curie, 75231 Paris Cedex 05, and Laboratoire de Biochimie des Acides Nucléiques, Ecole Normale Supérieure de Saint-Cloud, 92211 St Cloud Cedex, and *+Laboratoire de Biologie Moléculaire, CNRS-INRA, Groupement Scientifique 'Microbiologie' de Toulouse, BP 27, 31326 Castanet-Tolosan Cedex, France*

Received 4 December 1984; revised version received 18 December 1984

Cauliflower Mosaic Virus (CaMV) aphid transmissibility depends on the concomitant presence of an 18 kDa polypeptide and an intact gene II coding for this P18 protein. The molecular basis of this phenomenon can now be analyzed since the complete nucleotide sequence is known for two transmissible (S and D/H) and one non-transmissible strain (CM1841). In order to get more detail on this topic, we have determined the nucleotide sequence of gene II and flanking regions for an additional strain PV147. The DNAs of strains CM1841 and PV147 show considerable homology (more than with strains S and D/H) but these strains have opposite phenotypes. In consequence, strain PV147 allows us to specify further the putative point mutations responsible for the lack of aphid transmission observed with CM1841. The results obtained show that substitutions possibly involved in this defective phenotype are only two, instead of five on the basis of preexisting data. More interestingly, analysis of the potential effect of such substitutions upon predicted secondary structures of the different P18 polypeptides led us to a new hypothesis. Strain CM1841 could be mutated in the expression of gene II, a gene intact in other respects.

CaMV *Aphid transmissibility* *Gene II nucleotide sequence* *P18 secondary structure* *Regulation mutant*

1. INTRODUCTION

Cauliflower Mosaic Virus (CaMV), the best characterized DNA plant virus (review, [1]), is usually transmitted in nature by aphids [2]. Some strains are not aphid transmissible although they are still mechanically transmissible [3], and aphid transmissibility is known to require the presence of an aphid transmission factor (ATF) present only in plants that have been infected previously with an aphid-transmissible strain of CaMV [4]. The viral genome is a double-stranded circular molecule, 8 kilobase pairs in length, which has been completely sequenced for three different strains [5–7]. Sequence analysis has shown the existence of seven major open reading frames (ORF), first interpreted as putative unspliced genes of CaMV by Franck et al. [5]. Three of these genes have now been precisely identified. The major coat protein (P42) is encoded by part of ORF IV [5,8], the viroplasm protein (P62) by ORF VI [9–11] and ATF (P18) by ORF II [12,13]. This mapping of ATF has been accomplished by correlating the non-transmissible phenotype with absence of P18 in infected leaves using in vitro constructed deletion mutants [12,13]. Also, the precise mutation which in nature gives rise to a defective phenotype is known, in the case of strain CM4-184, to be a deletion of 421 base pairs (bp) in ORF II [14,15].

One very interesting apparent exception to this general picture of ATF is strain CM1841, from which the deleted CM4-184 strain arose. Strain

* Present address: Laboratoire d'Oncologie Moléculaire, Institut Gustave Roussy, 94805 Villejuif Cedex, France

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

CM1841 shows no deletion [16], but nevertheless remains non-transmissible [3]. Nucleotide sequence data [6] show that ORF II of strain CM1841 can encode a P18 polypeptide with several amino acid substitutions, as compared to P18 of strains S [5] and D/H [7], two strains which are transmissible. Since some of these substitutions may play a role in protein conformation, we have attempted to correlate changes in the predicted secondary structure for the gene product of ORF II with the lack of aphid transmissibility observed for strain CM1841. For this study we have chosen strain PV147, from among the cabbage B isolates known to be aphid transmitted [3]; furthermore, it is more closely related to CM1841 than are the other two previously sequenced strains [17]. We first checked PV147 for the presence of P18, and determined the nucleotide sequence of ORF II in this strain. Subsequently, we compared the secondary structure predicted for P18 polypeptides from this and the other transmissible and non-transmissible CaMV strains. Finally, we discuss the possibility that the relevant mutation in strain CM1841 affects the regulation of expression of P18 rather than its structure.

2. MATERIALS AND METHODS

Viruses (cabbage B PV147 from ATCC, B-JI and CM4-184) were propagated in *Brassica rapa* (cv Just Right), and DNAs were prepared as described previously [18,19]. Viroplasms were partially purified as in [20] and proteins were analysed by SDS–PAGE (polyacrylamide gel electrophoresis) as described in [21]. Restriction enzymes from New England Biolabs, Bethesda Research Laboratories or Boehringer were used as recommended by the manufacturer. Restriction fragments were separated by vertical slab gel electrophoresis [22], localized either by their fluorescence in the presence of ethidium bromide under UV illumination or by autoradiography of ^{32}P-labelled material, and eluted electrophoretically [23]. They were 3′-end labelled using [α-^{32}P]dNTPs and the Klenow fraction of *E. coli* DNA polymerase I purchased from New England Nuclear, and sequenced according to Maxam and Gilbert [24]. Nucleotide sequence data were processed using computer programs written by J.P. Dumas (unpublished results) to achieve strain comparisons and to translate nucleotide sequences into proteins. Predictions of the secondary structure of proteins were performed using the algorithm of Garnier et al. [25].

3. RESULTS

3.1. *Analysis of the P18 polypeptide from strain PV147*

Viroplasm preparations from strain PV147 have been analysed for their protein content (fig.1). A polypeptide having a molecular mass of 18 kDa is present in PV147 extracts (lanes a, b and e), as it is in extracts from strain B-JI (lane c) which is aphid transmitted [12], whereas as expected it is absent from extracts of strain CM4-184 (lane d). Previous restriction site mapping [17,26] of strain

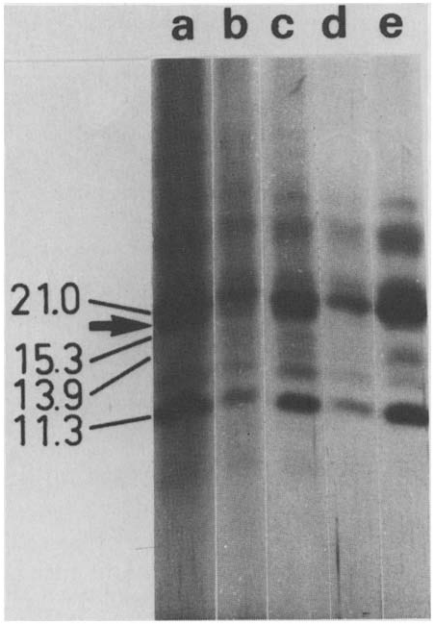

Fig.1. Presence of the P18 polypeptide in CaMV strain PV147 viroplasm preparations. 12.5% SDS polyacrylamide gel-fractionated polypeptides stained with silver nitrate. Viroplasm preparations were isolated from plants infected with different CaMV strains: PV147 (a) 20 μg and (b) 10 μg of protein from the same preparation; PV147 (e) 20 μg of protein from another preparation; B-JI (c) 20 μg of protein; CM4-184 (d) 20 μg of protein. Purified histones were used as molecular mass markers and migrated as indicated in the left margin; molecular masses are given in kDa. The arrow indicates the migration of the P18 polypeptide.

PV147 allowed us to isolate a DNA fragment of 649 bp, extending from an *Alu*I site to a *Bam*HI site (positions 1284 and 1931, respectively, in the CM1841 sequence), which spans the entire ORF II plus its 5′ and 3′ flanking sequences. Fig.2 displays the nucleotide sequence determined for this fragment of PV147 and notes differences observed between this strain and the other three sequenced strains. Only nucleotide substitutions but neither deletions nor insertions are observed within ORF II. The corresponding P18 polypeptide sequences are shown in fig.3. Comparison by pairs, at the nucleotide or at the amino acid levels are presented in table 1. They demonstrate that, among the three transmissible strains (PV147, S and D/H), PV147 is the most closely related to the non-transmissible strain CM1841. If one or more of the observed substitutions are responsible for the defective phenotype of strain CM1841, the corresponding amino acid in this strain should differ from that in all the other three strains. Among the 10 positions where the CM1841 amino acid sequence differs from at least one other strain, only three positions (amino acids 89, 94 and 105) differ from all three. These positions are thus candidates as point mutations responsible for the defective phenotype of strain CM1841.

3.2. *Predicted secondary structure of the different P18 polypeptides*

One way by which the amino acid substitutions described above are likely to affect protein function, and thus phenotype, is by modifying protein conformation. Such modifications can now be conveniently estimated using computer prediction of secondary structure. Fig.4 shows the secondary structure predicted for the P18 polypeptide encoded by ORF II of CaMV strain PV147 using four conformational states: helical (α-helix), extended (β-sheet), coil and turn. The predicted conformations for the P18 polypeptide have been similarly established for the three additional strains. All were compared pairwise. For these comparisons, changes which were merely a one-step shift of a

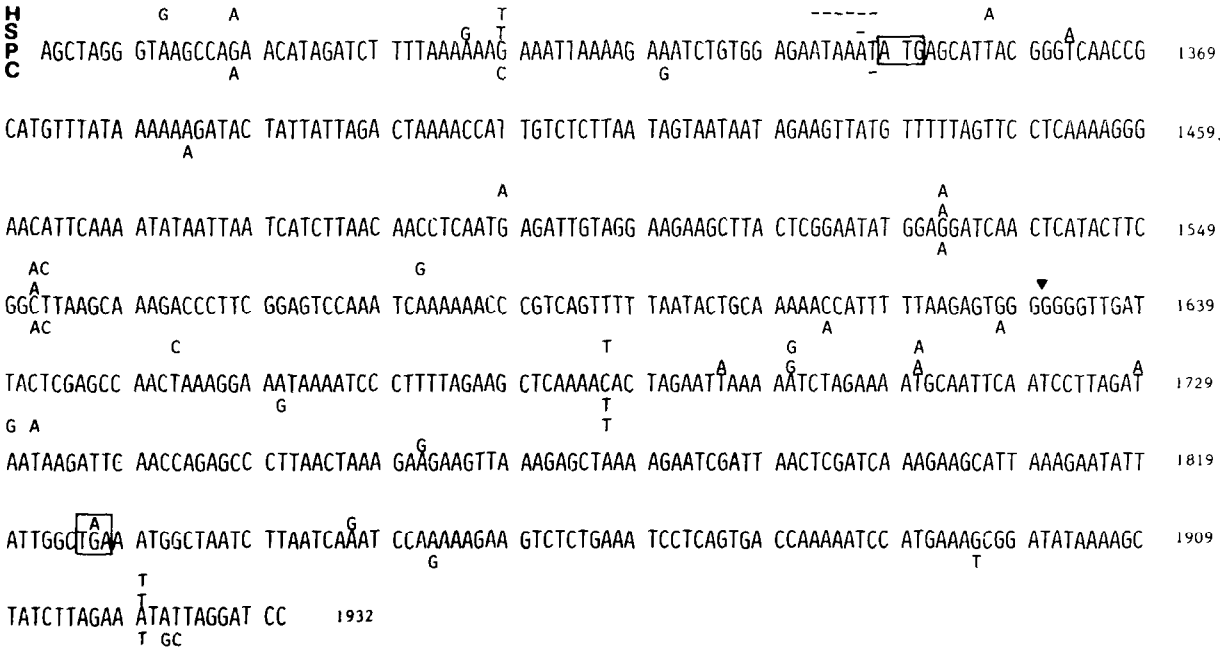

Fig.2. Nucleotide sequence of CaMV ORF II and its flanking regions. Nucleotide sequence of PV147 (P) DNA strand β is presented in the 5′ to 3′ direction. Differences with strain CM1841 (C) are indicated below this sequence, and differences with the other two strains S and D/H (H) are indicated above it. A, G, C or T are substitutions, and a dash indicates a deletion. Initiation and termination codons in ORF II are boxed. ▼ indicates the position of the discontinuity Δ3. Numbering corresponds to the position in strain CM1841 [6].

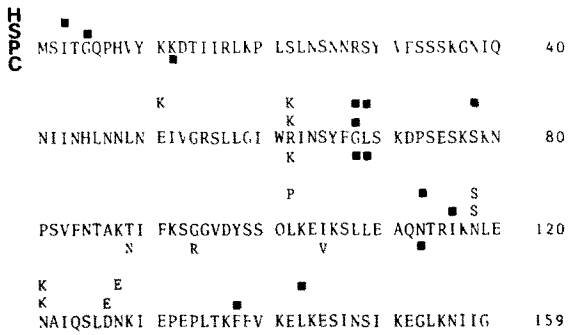

Fig.3. Primary structure of the P18 polypeptide of CaMV strain PV147. Amino acid sequence of PV147 (P) protein P18 is presented from the N- to the C-terminal end. Nomenclature is from IUPAC-IUB. Amino acids which differ in strain CM1841 (C) are indicated below the sequence, and differences with the other two strains S and D/H (H) are indicated above it. Amino acids which are identical in two strains but correspond to different codons are indicated by ■.

given conformation segment have been neglected. In order to assess the degree of consistency of any change in conformational state predicted at a given amino acid position, the two highest conformation parameters, not only the maximal one, have been taken into account. The only noticeable change between the CM1841 peptide and those of two other strains occurs between positions 115 and 120, where a coil region replaces part of the long α-helix (positions 100–125) present in P18 of both strains S and D/H. Since, however, this region is identical in strains CM1841 and PV147, this change cannot be related to the loss of aphid transmissibility. A similar conclusion can also be inferred from the amino acid sequence comparisons described previously. None of the three amino acid substitutions (positions 89, 94 or 105) which are unique to strain CM1841 appear to have a significant effect on the predicted conformation of P18 proteins:

Table 1

ORF II of CaMV transmissible and non-transmissible strains

	Pairs compared[a]	Number of differences	Positions[b]									
(A)	C/P	8	1384	1533	1552	1553	1614	1628	1661	1687		
	C/S	13	1363	1384	1553	1614	1628	1661	1687	1696	1701	1711
			1729	1762	1827							
	C/H	13	1357	1384	1499	1582	1614	1628	1653	1661	1701	1711
			1730	1732	1777							
	P/S	9	1363	1533	1552	1696	1701	1711	1729	1762	1827	
	P/H	13	1357	1499	1533	1552	1553	1582	1653	1687	1701	1711
			1730	1732	1777							
	S/H	14	1357	1363	1499	1553	1582	1653	1687	1696	1729	1730
			1732	1762	1777	1827						
(B)	C/P	4		62	89	94		105				
	C/S	6			89	94		105	118	121	127	
	C/H	8	51		89	94	102	105	118	121		128
	P/S	4		62					118	121	127	
	P/H	6	51	62			102		118	121		128
	S/H	4	51				102				127	128

[a] CaMV strains have been abbreviated as follows: C = CM1841; P = PV147; S = S; H = D/H

[b] Nucleotide positions are from the sequence of strain CM1841 [6]; amino acid positions are numbered from the first ATG codon in ORF II

(A) Comparisons at the nucleotide sequence level. (B) Comparisons at the amino acid sequence level

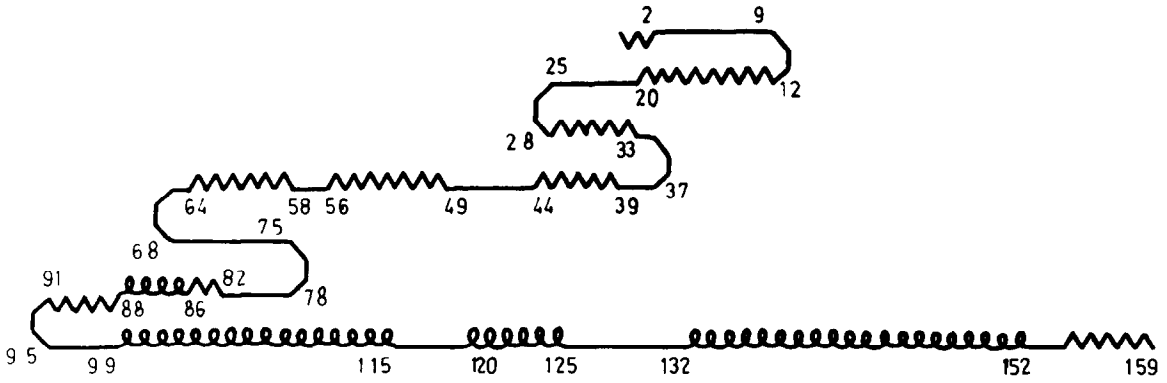

Fig.4. Secondary structure of the P18 polypeptide of CaMV strain PV147. ⌇⌇⌇⌇⌇, α-helix; ∧∧∧∧, β-sheet; ——, coil; β turns are indicated by a reversal of polypeptide chain. Numbers correspond to the positions of amino acids in the sequence presented in fig.3.

between positions 89 and 99 a β-sheet is followed by a turn and a coil segment in all three strains. Only the length of these three segments may vary slightly from one strain to another. The predicted secondary structures obtained with polypeptides substituted at each one of these positions singly show that only substitutions at position 89 (Thr ⟶ Asn) or 94 (Gly ⟶ Arg) could be responsible for such very limited changes in length.

4. DISCUSSION

The main argument for the involvement of CaMV ORF II in the transmission of the virus by aphids is the concomitant loss of aphid transmissibility and P18 polypeptide when this part of the genome is deleted. This conclusion has been reached by analyzing deletions introduced in vitro in the DNA of transmissible strains B-JI [12] or NY8153 [13] and more recently B-D [27] and S [28,29], and by the fact that the naturally occurring deletion in ORF II of CM4-184 is also associated with a defective phenotype and absence of P18 [12]. However, strain CM1841, the progenitor of CM4-184, though likewise defective is not deleted.

In order to determine whether a point mutation inside ORF II could be responsible for the loss of aphid transmissibility, we have determined the nucleotide sequence for the ORF II of another strain PV147, whose genome sequence is very close to that of CM1841 whereas its phenotype is the opposite. Comparison of primary and secondary structures of the P18 polypeptides encoded by non-transmissible and transmissible strains have been achieved. The knowledge of the PV147 sequence turned out to be useful because it restricts the number of point mutations which could be involved in the defective phenotype of CM1841. The only candidates are amino acid substitutions at position 89 and 94 in P18 polypeptide, corresponding to nucleotides 1614 and 1628 in the CM1841 sequence, since the Asn substitution at position 121, which had been suspected by Armour et al. [13] to play a role in the defective phenotype, is excluded by our results. The same authors also noticed the substitution Gly ⟶ Arg at position 94, because of its possible conformational consequences, but our analyses indicate that neither this substitution nor the substitution at position 89 have a significant effect on the predicted conformations. Thus, the involvement of any one of these substitutions is not very likely in terms of overall conformation, and could only be explained by the specific destruction of an active site.

Therefore, we came to another hypothesis: mutation in CM1841 could affect the control of expression of the P18 gene rather than the gene itself. This could result in the absence of the P18 polypeptide in leaves infected with strain CM1841. It is already known that another isolate, the Campbell isolate, which is non-transmissible and lacks P18, is not deleted [12]. In this case, however, the absence of P18 can alternatively be due to the existence of a stop codon inside ORF II. So far, the structural elements responsible for the expression

of ORF II are not known. Recently, Woolston et al. [12] suggested a role for upstream control sequences. The main difference between strains around ORF II is found just upstream of the first ATG codon. Compared to the sequences of strain S and CM1841, there is a 5 nucleotide deletion in strain D/H and a 1 nucleotide insertion in strain PV147. Since these features modify neither the end of ORF I nor the first ATG in ORF II, and since they are not specific to strain CM1841, they cannot be related to the mutant phenotype. The possible control elements could be located in a more distant position, affecting either the transcription or the processing of the mRNA which encodes the P18 polypeptide.

ACKNOWLEDGEMENT

We are greatly indebted to Dr Suzan, Elsevier-Saint-Just, for critical reading of the manuscript.

REFERENCES

[1] Hohn, T., Richards, K. and Lebeurier, G. (1982) Curr. Topics Microbiol. Immunol. 96, 193–236.
[2] Day, M.F. and Venables, D.G. (1961) Aust. J. Biol. Sci. 14, 187–197.
[3] Lung, M.C.Y. and Pirone, T.P. (1973) Phytopathology 63, 910–914.
[4] Lung, M.C.Y. and Pirone, T.P. (1974) Virology 60, 260–264.
[5] Franck, A., Guilley, H., Jonard, G., Richards, K. and Hirth, L. (1980) Cell 21, 285–294.
[6] Gardner, R.C., Howarth, A.J., Hahn, P., Brown-Luedi, M., Shepherd, R.J. and Messing, J. (1981) Nucl. Acids Res. 9, 2871–2888.
[7] Balazs, E., Guilley, H., Jonard, G. and Richards, K. (1982) Gene 19, 239–249.
[8] Daubert, S., Richins, R., Shepherd, R.J. and Gardner, R.C. (1982) Virology 122, 444–449.
[9] Odell, J.T. and Howell, S.H. (1980) Virology 102, 349–359.
[10] Covey, S.N. and Hull, R. (1981) Virology 111, 463–474.
[11] Xiong, C., Muller, S., Lebeurier, G. and Hirth, L. (1982) EMBO J. 1, 971–976.
[12] Woolston, C.J., Covey, S.N., Penswick, J.R. and Davies, J.W. (1983) Gene 23, 15–23.
[13] Armour, S.L., Melcher, U., Pirone, T.P., Lyttle, D.J. and Essenberg, R.C. (1983) Virology 129, 25–30.
[14] Hull, R. and Howell, S.H. (1978) Virology 86, 482–493.
[15] Howarth, A.J., Gardner, R.C., Messing, J. and Shepherd, R.J. (1981) Virology 112, 678–685.
[16] Hull, R. (1980) Virology 100, 76–90.
[17] Volovitch, M., Modjtahedi, N., Chouikh, Y. and Yot, P. (1981) Biochem. Int. 3, 225–232.
[18] Shepherd, R.J., Bruening, G.E. and Wakeman, R.J. (1970) Virology 41, 339–347.
[19] Hull, R., Shepherd, R.J. and Harvey, J.D. (1976) J. Gen. Virol. 31, 93–100.
[20] Shockey, M.W., Gardner, C.O. jr, Melcher, U. and Essenberg, R.C. (1980) Virology 105, 575–581.
[21] Laemmli, U.K. (1970) Nature 227, 680–685.
[22] Danna, K.J., Sack, G.H. jr and Nathans, D. (1973) J. Mol. Biol. 78, 363–376.
[23] Chouikh, Y., Volovitch, M. and Yot, P. (1979) Mol. Biol. Reports 5, 237–239.
[24] Maxam, A.L. and Gilbert, W. (1980) Methods Enzymol. 65, 490–560.
[25] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[26] Volovitch, M., Drugeon, G., Dumas, J.P., Haenni, A.L. and Yot, P. (1979) Eur. J. Biochem. 100, 245–255.
[27] Daubert, S., Shepherd, R.J. and Gardner, R.C. (1983) Gene 25, 201–208.
[28] Dixon, L.K., Koenig, I. and Hohn, T. (1983) Gene 25, 189–199.
[29] Givord, L., Xiong, C., Giband, M., Koenig, I., Hohn, T., Lebeurier, G. and Hirth, L. (1984) EMBO J. 3, 1423–1427.