Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 1470–1473, February 1993
Biochemistry

Identification of C-terminal amino acid residues of cauliflower mosaic virus open reading frame III protein responsible for its DNA binding activity

JEAN-LUC MOUGEOT, THIERRY GUIDASCI, THIERRY WURCH, GENEVIEVE LEBEURIER, AND JEAN-MICHEL MESNARD

Institut de Biologie Moléculaire des Plantes du Centre National de la Recherche Scientifique, Université Louis Pasteur, 12, rue du Général Zimmer, 67084 Strasbourg Cedex, France

Communicated by Theodor O. Diener, November 30, 1992 (received for review May 7, 1992)

ABSTRACT **We cloned in *Escherichia coli* truncated versions of the protein p15 encoded by open reading frame III of cauliflower mosaic virus. We then compared the ability of the wild-type p15 (129 amino acids) and the deleted p15 to bind viral double-stranded DNA genome. Deletions of >11 amino acids in the C-terminal proline-rich region resulted in loss of DNA binding activity of wild-type p15. Moreover, a point mutation of the proline at position 118 sharply reduced the interaction between the viral protein and DNA. These results suggest that cauliflower mosaic virus p15 belongs to the family of DNA binding proteins having a proline-rich motif involved in interaction with double-stranded DNA.**

Cauliflower mosaic virus (CaMV) is a 53.8-nm particle (1) that possesses an 8.0-kb, double-stranded DNA genome (2, 3) having seven open reading frames (ORFs). The genome is enclosed in an icosahedral capsid composed of two types of viral proteins: the major one (420 subunits per virion) has a molecular mass of 37–42 kDa (4) and the minor one has a molecular mass of 11 kDa (5). It has been shown that the minor capsid protein is encoded by the CaMV ORF III (5). This ORF gives rise to two proteins: p15, consisting of 129 amino acids (6), is the primary translation product (7) and is a non-sequence-specific DNA binding protein able to interact strongly with the CaMV genome (8). The other protein, p11, corresponds to a processed form of p15. The amino acids are removed from the C terminus by an as yet unidentified protease (5), but the *in vivo* processing site remains unknown. Only p11 has been found associated with viral particles (5). p11, unlike p15, does not bind to double-stranded DNA (5).

On the basis of these data, we chose to study the influence of the C-terminal region on the interaction between p15 and the CaMV genome. Deletion and point mutation analysis reveal that the C-terminal region of p15 as expected contains a DNA-binding domain and that a proline-rich motif is responsible for the non-sequence-specific recognition between p15 and the viral genome.

MATERIALS AND METHODS

Plasmids and Construction of Mutants. The plasmid pGM301 consisting of pUC8 containing the CaMV ORF III under control of the *lac* promoter–operator has been described (9). It allows synthesis of a fusion protein containing the N terminus of β-galactosidase.

C-terminal deletion mutants were generated from pGM301 by cleavage at the unique *Sal* I site, limited BAL-31 digestion, and blunt-end formation with DNA polymerase I large fragment (Klenow), followed by ligation to an 8-bp linker containing a TAG stop codon (GGCTAGCC; Fig. 1*A*). Plasmids were screened for the linker by *Nhe* I digestion, and the extent of the deletion was determined by sequencing.

For site-directed mutagenesis, pGM301 was digested with *Eco*RI and the 497-bp fragment containing ORF III was inserted into the *Eco*RI site of the pSELECT vector (Promega) and transformed into *Escherichia coli* strain JM109. After infection with the helper phage R408, single-stranded template DNA for mutagenesis was isolated and used for site-directed mutagenesis according to the protocol. Mutation was verified by sequence analysis, the appropriate plasmid DNA was cut with *Eco*RI, and the mutant fragment was reinserted into the *Eco*RI-digested pGM301 to reconstitute plasmid pGM311, which allowed synthesis of the mutated p15.

Analysis of Proteins. Aliquots (1 ml) of *E. coli* strain JM103 cultures grown in LB medium in the presence of ampicillin (100 μg/ml) and 1 mM isopropyl β-D-thiogalactoside were harvested by centrifugation, and the cell pellets were lysed in 200 μl of sample buffer (10).

Proteins were fractionated by electrophoresis on SDS/18% polyacrylamide gels, renaturated with washing buffer (11), and then transferred to nitrocellulose (12). After transfer, the filters were used either for immunological detections or for DNA binding assays.

For immunodetection, the filters were incubated with the anti-virion antiserum (13), which, prior to use, was preabsorbed with pellets of bacteria JM103. After reaction with the antiserum, the specific proteins were visualized by using an enhanced chemiluminescence Western blotting detection system (Amersham).

For DNA binding tests, the filters were incubated with ^{32}P-labeled CaMV double-stranded DNA genome as described (11). CaMV DNA was purified from viral particles (14, 15).

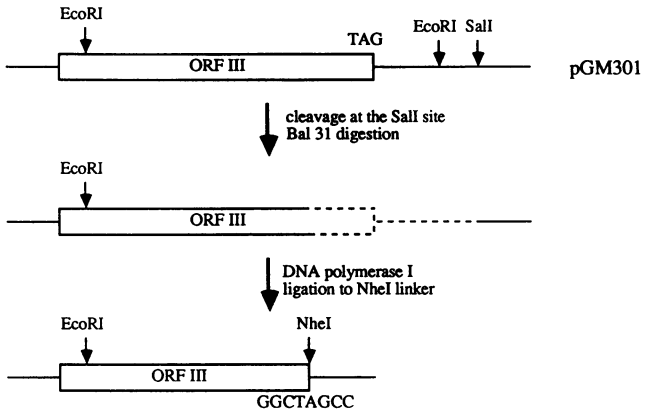
RESULTS

Characterization of Deleted CaMV p15 Clones. The plasmid pGM301 (9) encoding CaMV p15 was linearized at the unique *Sal* I site and sequentially digested with BAL-31 exonuclease. A *Nhe* I linker containing a TAG stop codon was ligated to the truncated DNA (Fig. 1*A*). Ten recombinant clones coding for p15 deleted in the C-terminal region were identified. Due to use of the *Nhe* I linker, deleted p15 had additional amino acids (Fig. 1*B*): one additional glycine for clones pGM126a, pGM122a, pGM118a, pGM117a, pGM116a, and pGM108a (class A mutants); four amino acids for clones pGM117b, pGM112b, and pGM107b (class B mutants); or seven amino acids for clone pGM115c (class C mutant).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: CaMV, cauliflower mosaic virus; ORF, open reading frame.

A. Plasmids



B. Mutants

mutant	Carboxy-terminal amino acid sequence
pGM301	105-elgssgnpkaltwpfkapagwpnqf-129
pGM126a	105-elgssgnpkaltwpfkapagwpG
pGM122a	105-elgssgnpkaltwpfkapG
pGM118a	105-elgssgnpkaltwpG
pGM117a	105-elgssgnpkaltwG
pGM116a	105-elgssgnpkaltG
pGM108a	105-elgsG
pGM117b	105-elgssgnpkaltwRLAG
pGM112b	105-elgssgnpRLAG
pGM107b	105-elgWLAG
pGM115c	105-elgssgnpkalTASRLAG

FIG. 1. Plasmids and mutants used in analysis of CaMV p15 DNA binding domain. (*A*) Construction of C-terminal deletion mutants. Arrows indicate positions of relevant restriction sites. Boxes correspond to coding sequence of CaMV ORF III; thin lines represent bacterial plasmid sequence. (*B*) Mutants. Amino acid sequence of the C terminus of p15 is shown. For each of the p15 mutants depicted below, the sequence is represented and amino acids encoded by the linkers are present, as indicated in capital letters. C-terminal deletion mutants are designated by the number of the farthest C-terminal wild-type amino acid and by a letter corresponding to different classes. Class A mutants encode deleted p15 with an additional glycine, whereas class B and C mutants have, respectively, 4 and 7 additional amino acids.

CaMV DNA Binding Studies with the Class A Mutants. In pGM301, CaMV ORF III is under control of the *lac* promoter operator of plasmid pUC8 (9). Upon isopropyl β-D-thiogalactoside induction of the *lac* promoter, pGM301 produced a fusion protein containing the N terminus of β-galactosidase (10 amino acids) and all amino acids of CaMV p15 (Fig. 2, lane 2). Since truncated proteins may have properties different from wild-type p15, we checked their production in *E. coli* by Western blotting. Indeed, the pGM126a deleted p15 (lane 3) was poorly produced, whereas the pGM108a protein (lane 8) was highly expressed. Moreover, the electrophoretic migration of the pGM116a truncated p15 (lane 7) did not correspond precisely to that predicted according to its expected molecular mass. This was especially clear when comparing the mobility of pGM117a and pGM116a polypeptides (lanes 6 and 7).

To test the DNA binding activities of the class A mutant p15, bacterial extracts were separated by SDS/PAGE and electroblotted onto nitrocellulose. The filters were then incubated with ^{32}P-labeled CaMV double-stranded DNA genome. By this approach, Giband *et al.* (5) demonstrated that the DNA binding activity of CaMV p15 expressed in *E. coli* from pGM301 is due solely to the amino acids of p15 and not

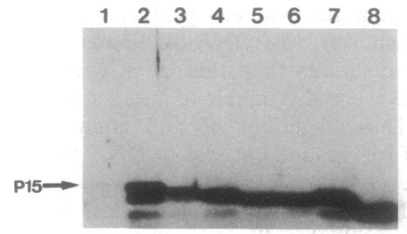


FIG. 2. Immunodetection of β-galactosidase–p15 fusion protein and C-terminal deleted β-galactosidase–p15 fusion polypeptides in *E. coli* JM103 transformed with pUC8 containing complete or truncated CaMV ORF III. After overnight cultures, a 1-ml sample was withdrawn and centrifuged. The cell pellet was lysed in 200 μl of sample buffer, and 5 μl per well was analyzed on SDS/18% polyacrylamide gel. Immunoblotting was done with an anti-virion antiserum diluted 1:1000. Samples correspond to proteins from extracts of bacteria containing pUC8 (lane 1), pGM301 (lane 2), pGM126a (lane 3), pGM122a (lane 4), pGM118a (lane 5), pGM117a (lane 6), pGM116a (lane 7), and pGM108a (lane 8). Arrow, localization of complete β-galactosidase–p15 fusion protein.

to the amino acids of β-galactosidase, this enzyme having no DNA binding activity. Our DNA binding assays confirmed this result (Fig. 3). To compare the DNA binding activity of the different truncated β-galactosidase–p15 fusion proteins with that of the complete β-galactosidase–p15 fusion polypeptide, we determined the rate of expression of each mutated polypeptide by densitography of the immunoblots and we carried out the assays with bacterial extracts containing the same quantity of tested p15 (corresponding to 150 ng of complete p15). Therefore, the concentration of bacterial proteins was not the same for the different clones. For DNA binding assays with clones pGM126a, pGM118a, and pGM117a (lanes 3, 7, and 8, respectively), we detected bacterial proteins with DNA binding activity not detected for the tests with clones pGM301 and pGM122a (lanes 1 and 5, respectively). These bacterial proteins have already been described (5).

To be sure that the quantity of mutant proteins was not altered by the transfer onto nitrocellulose, the amount of p15-derived protein present on the nitrocellulose filter was measured by chemiluminescence assays by using two different antisera: an antiserum directed against the virion (13) and an antiserum raised against the first 19 amino acids of p15 (7). By this approach, we were able to check quantitation of the proteins bound to the nitrocellulose membrane and to confirm that the quantity of tested proteins was the same for all the mutants. Among the class A mutants, the deleted proteins encoded by clones pGM126a, pGM122a, and pGM118a (respectively, lanes 3, 5, and 7 of Fig. 3) were still able to bind

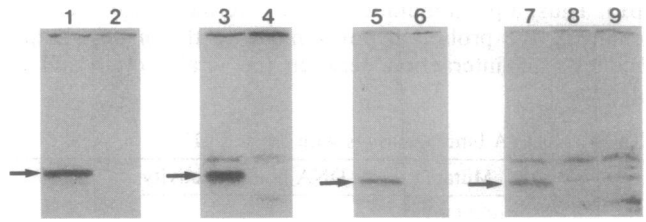


FIG. 3. Autoradiographs of CaMV genome binding profiles. Proteins were separated by SDS/PAGE and electroblotted onto nitrocellulose. Filter was then incubated with CaMV double-stranded DNA genome labeled with ^{32}P. Resulting DNA–protein complexes were visualized by autoradiography. Proteins were from extracts of bacteria containing pGM301 (lane 1), pGM126a (lane 3), pGM122a (lane 5), pGM118a (lane 7), pGM117a (lane 8), and pUC8 (lanes 2, 4, 6, and 9). The quantity of tested p15 corresponded to 150 ng of complete p15 for each DNA binding assay. Likewise, the quantity of each control extract (containing pUC8) depended on the tested clone (containing p15). Arrows, localization of tested p15.

the CaMV genome, whereas pGM117a (Fig. 3, lane 8), 1 amino acid shorter, lost DNA binding activity. For clones with larger deletions than that of pGM117a, the encoded polypeptides were also unable to interact with the DNA (see Table 1). Thus, deletions of >11 amino acids in the C terminus of p15 resulted in loss of DNA binding activity of CaMV p15.

CaMV DNA Binding Studies with Class B and C Mutants. To determine whether the properties of the deleted amino acids were responsible for p15 interaction with the genome, we tested the class B and C mutants for which the deleted amino acids of the C terminus were replaced by amino acids encoded by the added *Nhe* I linkers. Thus, it was interesting to compare the interaction of the CaMV genome with the p15 encoded by clones pGM117a and pGM117b. Both clones made possible the synthesis of viral polypeptides having the same amino acids deleted but with different modifications in the C-terminal region of p15 (see Fig. 1*B*). Before testing DNA binding activities of the class B and C mutants, we checked their expression in *E. coli*. As described for the class A mutants, expression of the different mutated p15 polypeptides was not the same, especially for clone pGM117b, for which the quantity of viral protein was low (Fig. 4). When we carried out DNA binding assays, we found no DNA binding activity in any of the class B and C mutants (see Table 1). This result confirmed that the last 12 amino acids were involved in the interaction between p15 and the CaMV genome.

Mutation of the Proline at Position 118 Alters the DNA Binding Activity of p15. pGM118a (having CaMV DNA binding activity) and pGM117a (without DNA binding activity) differ only by a proline residue at position 118 in pGM118a (Fig. 1*B*). To determine whether this proline was required for the interaction between p15 and the CaMV genome, a point mutation was engineered into wild-type p15. The first cytosine of the proline codon (CCC) was replaced by an adenine to create a threonine codon (ACC). The new plasmid containing the point mutation was called pGM311. Sequence analysis of the C-terminal region of the CaMV ORF III showed that pGM311 contained only the proline to threonine mutation at residue 118. Mutated p15 was still able to bind the CaMV genome but with a lower activity than that of wild-type p15 (Fig. 5, lanes 1 and 2). By densitography, DNA binding activity of the mutant was evaluated at only 20% of the DNA binding activity of wild type. To be sure that the difference of activity was not due to an instability of mutated p15, its rate of expression was checked. As shown in Fig. 5 (lanes 3 and 4), production of both proteins was almost the same: the rate of expression of mutated p15 corresponded to 98% of the rate of expression of wild-type p15. Thus, a point mutation in the C-terminal region of p15, changing the proline at position 118 to threonine, sharply reduces the interaction between the viral protein and its genome.

Table 1. DNA binding assays with deleted p15

Mutant	DNA binding activity
pGM301	+
pGM126a	+
pGM122a	+
pGM118a	+
pGM117a	−
pGM116a	−
pGM108a	−
pGM117b	−
pGM112b	−
pGM107b	−
pGM115c	−

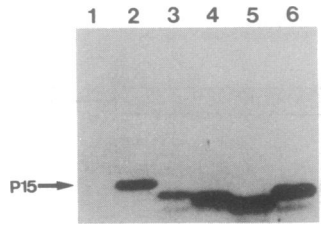


FIG. 4. Immunodetection of modified β-galactosidase–p15 fusion proteins expressed from class B and C mutants. Proteins were prepared and fractionated as described in the legend to Fig. 2. Immunoblotting was done with an anti-virion antiserum 1:1000. Bacterial extracts were with pUC8 (lane 1), pGM301 (lane 2), pGM117b (lane 3), pGM112b (lane 4), pGM107b (lane 5), and pGM115C (lane 6). Arrow, localization of modified β-galactosidase–p15 fusion protein.

DISCUSSION

The CaMV ORF III product p15 is able to interact with the CaMV genome (8) but the exact location of its DNA binding domain remains unknown. However, the mature form of p15, processed at the C terminus, is unable to bind double-stranded DNA (5). We have produced a series of deleted mutants in the C-terminal domain of p15 and demonstrated that deletions of >11 amino acids resulted in the loss of DNA binding activity.

The ability of p15 to bind the CaMV genome is due to a proline-rich motif. Its C-terminal domain contains serine, proline, and basic residues, characteristic of the proline-rich DNA binding proteins (16). Deletion of three of four prolines in the C-terminal domain abolishes completely the DNA binding activity of p15. Moreover, a mutation of the proline at position 118 reduces the binding of p15 with the viral genome.

CaMV p15 is the only known case of a plant viral protein for which it has been demonstrated that a proline-rich motif is involved in the interaction between a structural protein and its genome. Other viral proteins can be classified as proline-rich DNA binding proteins: the hepatitis B virus coat protein (16) and protein M of bacteriophage P2 (17). The proline-rich motif of hepatitis B virus coat protein consisting of contiguous SPRRR motifs (18) belongs to the SPKK group (16), whereas the bacteriophage P2 protein M having a GRP motif (17) is classified as belonging to the GRP group (16). A search for homology between CaMV p15 and bacteriophage protein M revealed close structural resemblance between the C-terminal domains, corresponding to their DNA binding domain (Fig. 6), but this alignment reveals the presence of a tryp-

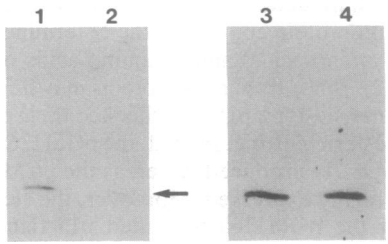


FIG. 5. DNA binding assay with point-mutated p15. Single colonies containing either pGM301 or pGM311 were used to inoculate cultures in LB medium in the presence of ampicillin. At A_{600} = 0.7, isopropyl β-D-thiogalactoside was added. Two hours later, a 1-ml sample was withdrawn and centrifuged. Bacterial proteins were extracted, separated by SDS/PAGE, and analyzed as described. Extracts of bacteria containing either pGM301 (lanes 1 and 3) or pGM311 (lanes 2 and 4) were used either for DNA binding tests (lanes 1 and 2) or for immunological detections with an anti-virion antiserum diluted 1:1000 (lanes 3 and 4). Arrow, localization of mutated p15.

FIG. 6. Alignment of CaMV p15 with bacteriophage P2 protein M. C-terminal regions of both viral polypeptides have been aligned, from position 111 to position 126 for p15 (upper line) and from position 220 to position 242 for protein M (lower line). Identical or similar residues are boxed.

tophan residue for p15 instead of an arginine residue for protein M. However, involvement of a tryptophan within a proline-rich motif in the interaction with DNA has already been described for the archaebacterial MC1 protein (19). Protein MC1 is the major chromosomal protein in methanosarcinaceae, protecting DNA against thermal denaturation and inducing DNA bending and supercoiling. CaMV p15 has two motifs with a tryptophan adjacent to a proline in its C-terminal domain. The presence of such repeated motifs in a DNA binding domain has also been described for the mammalian MeCP2 protein, a chromosomal protein that possesses repeated SPKK motifs contiguous to repeated GRP motifs (20).

The nucleic acid binding activity of CaMV p15 may reflect that a tryptophan adjacent to a proline intercalates between the base plates within the minor groove in a way similar to the binding drug Hoechst 33258 (16). Moreover, this motif could constitute the structural basis of p15 for nucleic acid condensation. The hepatitis B virus coat protein (18), the archaebacterial MC1 protein (19), and the mammalian MeCP2 protein (20) are all involved in nucleic acid condensation. For CaMV p15, the proline-rich motif DNA binding domain is limited to its C-terminal domain and is well structured. The characteristics of CaMV p15 define it rather as a foldase, participating in the formation of a stable nucleocapsid capable of transmission by aphids, long distance movement, and decapsidation in a favorable cellular environment.

The results of DNA binding experiments of different polypeptides having a proline-rich motif are consistent with the notion that this simple motif preferentially recognizes a particular DNA conformation rather than a certain DNA sequence (16). Indeed, CaMV p15 is a non-sequence-specific DNA binding protein (8). However, the characteristics of the DNA sequences recognized by p15 remain unknown. It will be of interest to carry out a precise mapping of the binding interactions with the viral genome by footprinting experiments.

1. Cheng, R. H., Olson, N. H. & Baker, T. S. (1992) *Virology* **186**, 655–668.
2. Mesnard, J. M., Lebeurier, G. & Hirth, L. (1990) in *Chromosomes: Eukaryotic, Prokaryotic and Viral*, ed. Adolph, K. W. (CRC, Boca Raton, FL), Vol. 3, pp. 143–162.
3. Covey, S. N. (1991) *Semi. Virol.* **2**, 151–159.
4. Al Ani, R., Pfeiffer, P. & Lebeurier, G. (1979) *Virology* **93**, 188–197.
5. Giband, M., Mesnard, J. M. & Lebeurier, G. (1986) *EMBO J.* **5**, 2433–2438.
6. Franck, A., Guilley, H., Jonard, G., Richards, K. E. & Hirth, L. (1980) *Cell* **21**, 285–294.
7. Xiong, C., Lebeurier, G. & Hirth, L. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6608–6612.
8. Mesnard, J. M., Kirchherr, D., Wurch, T. & Lebeurier, G. (1990) *Virology* **176**, 622–624.
9. Mesnard, J. M., Geldreich, A., Xiong, C., Lebeurier, G. & Hirth, L. (1984) *Gene* **32**, 39–47.
10. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
11. Mellor, J., Fulton, A. M., Dobson, M. J., Roberts, N. A., Wilson, W., Kingsman, A. J. & Kingsman, S. M. (1985) *Nucleic Acids Res.* **13**, 6249–6263.
12. Towbin, H., Staehlin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4350–4354.
13. Givord, L., Xiong, C., Giband, M., Koenig, I., Hohn, T., Lebeurier, G. & Hirth, L. (1984) *EMBO J.* **3**, 1423–1427.
14. Hull, R., Shepherd, R. J. & Harvey, J. D. (1976) *J. Gen. Virol.* **31**, 93–100.
15. Hohn, T., Hohn, B., Lesot, A. & Lebeurier, G. (1980) *Gene* **11**, 21–31.
16. Churchill, M. E. A. & Travers, A. A. (1991) *Trends Biochem. Sci.* **16**, 92–97.
17. Linderoth, N. A., Ziermann, R., Haggard-Ljungquist, E., Christie, G. E. & Calendar, R. (1991) *Nucleic Acids Res.* **19**, 7207–7214.
18. Nassal, M. (1992) *J. Virol.* **66**, 4107–4116.
19. Katouzian-Safadi, M., Laine, B., Chartier, F., Cremet, J. Y., Belaiche, D., Sautiere, P. & Charlier, M. (1991) *Nucleic Acids Res.* **19**, 4937–4941.
20. Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F. & Bird, A. (1992) *Cell* **69**, 905–914.