Vol. 138, No. 1, 1986 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 16, 1986 Pages 17-23

DETECTION OF A POTENTIAL TRANSCRIPTION CONTROL SEQUENCE ON THE CAULIFLOWER MOSAIC VIRUS GENOME BY DINUCLEOTIDE PRIMED "IN VITRO" TRANSCRIPTION

Richard COOKE and Paul PENON

Laboratoire de Physiologie Végétale, U.A. 565 du CNRS
Université de Perpignan, 66025 PERPIGNAN-Cédex - FRANCE

Received May 12, 1986

SUMMARY. The three sites of selective dinucleotide-primed "in vitro" transcription initiation on a cloned cauliflower mosaic virus DNA fragment have been localised by S1 nuclease mapping. Two of these sites lie within a region which has been shown to be essential for transcription complex formation on the viral sequences, one corresponding to a nuclease S1 hypersensitive site and the other to an imperfect repeat 100bp downstream. These sequences show striking homology with known transcription control sequences. These observations and the effect of the sequences on "in vitro" transcription raise the possibility that they may be involved in control of transcription of the viral genome.

While it has been demonstrated that purified RNA polymerase II is incapable of reproducing specific "in vivo" transcription initiation on homologous templates, several authors have shown that, under certain conditions, the enzyme is capable of a highly selective interaction with particular sequences on these templates (1-6). The similarity of certain of these sequences to known transcription control signals poses the interesting question as to whether the purified polymerase is capable of carrying out certain steps of the transcription process in the absence of other factors. We have previously shown that plant RNA polymerase II shows a highly selective initiation in the presence of a dinucleotide primer within a limited region of a cloned fragment of cauliflower mosaic virus (CaMV) DNA which contains the promoter sequences of the viral inclusion body protein gene (7). This preferential initiation occurs at a small number of sites several hundred base pairs upstream from the "in vivo" initiation site (7). Studies carried out using a deleted plasmid, pCa 4.2, show that this selectivity is dependent on sequences within a 205 bp region (8) located in the reverse transcriptase gene (9, 10). Using S1 nuclease as a probe we have shown the presence within this region of an S1-hypersensitive homopurine-homopyrimidine sequence capable of adopting an alternative, non-B conformation on a supercoiled plasmid and which we tentatively identified as the site of polymerase/DNA interaction (8). We show here that dinucleotide-primed transcription is effectively initiated at the level of these sequences and at similar sequences about 100 bp downstream. These two sites show remarkable similarities to sequences on other viral genomes which are involved in the control of transcription, suggesting that they may play a role in the expression of the CaMV genome.

0006-291X/86 $1.50

MATERIALS AND METHODS.

Restriction enzymes were obtained from Boehringer Mannheim or Amersham (France) and used according to the suppliers. [γ-^{32}P]-ATP was from Amersham. Purification of plasmid pCa 8 and soybean RNA polymerase II was as described (7). RNA was synthesised in the presence of the (ApG+ATP) primer (7), except that reactions were scaled up 50-fold, and purified as described by WASYLYK et al. (11). Restriction fragments were labelled by successive treatment with bacterial alkaline phosphatase and T4 polynucleotide kinase (BRL) as described by the suppliers. Strand separation was carried out on polyacrylamide gels (12) and S1 mapping by the method of Berk and Sharp (13). S1-resistant products were analysed on polyacrylamide-urea gels (14).

RESULTS AND DISCUSSION.

Plasmid pCa 8 (Fig. 1) contains the Bgl II B fragment of the DNA of the cabbage B-JI isolate of CaMV. In the presence of the combination (ApG+ATP), RNA polymerase II preferentially initiates transcription within a region delimited by the Hind II and Hind III restriction sites at 4838 and 5151 base pairs (bp) respectively (7), while the deletion which eliminates selective transcription covers 4915-5119 bp (8). "In vitro" initiation sites were mapped by hybridising RNA to separated strands of the 313 bp fragment delimited by the above Hind sites, 5'-labelled at the Hind III site, to detect transcription of the alpha-strand towards the "in vivo" initiation site, or at the Hind II site, to identify transcripts from the β-strand. Only transcription of the alpha-strand has been detected "in vivo" (15).

Localisation of "in vitro" initiation sites.

Fig. 2a shows that, at a low nuclease S1 concentration, five major bands are detected which correspond to transcription of the alpha-strand (lane 1), two doublets at 192 and 198 nucleotides (nt) and 93 and 98 nt and one band at 117 nt. At higher S1 concentrations (lanes 2

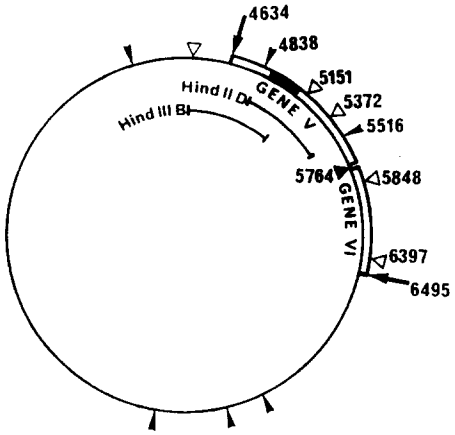

Fig. 1.- Physical map of plasmid pCa 8. The structural features of the CaMV DNA Bgl II insert are shown, with the positions of the coding regions and restriction enzyme sites for Bgl II (↓), Hind II (Y) and Hind III (▽). Fragments Hind II D and Hind III B, to which "in vitro" synthesised RNAs hybridise are shown, as is the extent of the deletion in pCa 4.2 (▬) which is described in the text. The arrow at 5764 bp shows the transcription initiation site for gene VI.

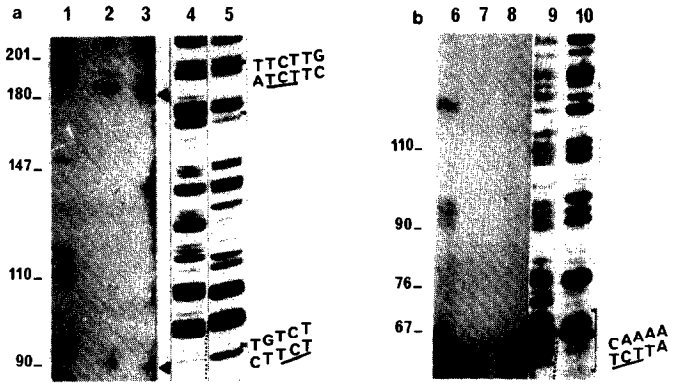

Fig. 2.- S1 nuclease mapping of (ApG+ATP)-primed initiation sites. RNA was synthesised as described (7) in reactions scaled up 50-fold and were hybridised to separated strands of the Hind II-Hind III fragment between 4838 and 5151 bp on the CaMV genome (see fig. 1), labelled at the 5'-extremity as described. Hybrids were treated with 5,000 (lanes 1,6), 10.000 (lanes 2,7) or 20.000 (lanes 3,8) units of nuclease S1 for 5h at 23°C, resistant fragments recovered by phenol extraction and ethanol precipitation in the presence of 8µg carrier tRNA and fragments analysed on 8% polyacrylamide-urea denaturing gels with A+G (lanes 4,9) and G (LANES 5, 10) sequencing lanes of the fragment used for hybridisation. Figures show the sizes of marker fragments. (a) Hybridisation to the alpha-strand, labelled at the Hind III site at 5151 bp. Radioactive material to the right of lane 3 is drying artefact. (b) Hybridisation to the ß-strand, labelled at the Hind II site at 4838 bp. Resistant fragments are indicated by arrows and the sequence of the transcribed strand at the initiation sites is shown. Underlined TCT sequences correspond to the (ApG+ATP) primer.

and 3) the upper band of the doublets disappears progressively, suggesting that it may be due to incomplete digestion of the last few nucleotides of the single-stranded regions of the DNA probe (16). The band at 117 nt is no longer visible at higher enzyme concentrations and may be artefactual. Hybridisation to the ß-strand (Fig. 2b) shows only one protected fragment of 62 nt at higher S1 concentrations (Fig. 2b).

A comparison of the protected fragments with the sequence ladders of the strands used for hybridisation (Fig. 2, lanes 4, 5, 9, 10) allows us to identify the three AGA sequences, corresponding to the (ApG+ATP) primer, on the viral genome (Fig. 3). All three lie within a 175 bp region which we have shown to contain a remarkable concentration of symmetrical or repeating elements (8). More particularly, the 192 nt fragment corresponds to an initiation site (alpha-1) within the S1 nuclease hypersensitive sequence previously described (8), while alpha-2 lies within an imperfect direct repeat of these sequences situated some 100 bp downstream. The beta-1 site corresponds to the first AGA sequence upstream from alpha-1. Fig. 3b shows the localisation of the "in vitro" initiation sites on the CaMV genome, within the reverse transcriptase gene (9, 10), some 800 bp upstream from one of the two known "in vivo" promoters.

A possible role for the sequences in CaMV expression?

We were intrigued by the presence within a coding region of the CaMV genome of sequences showing such unusual structural characteristics and which serve as sites of preferential

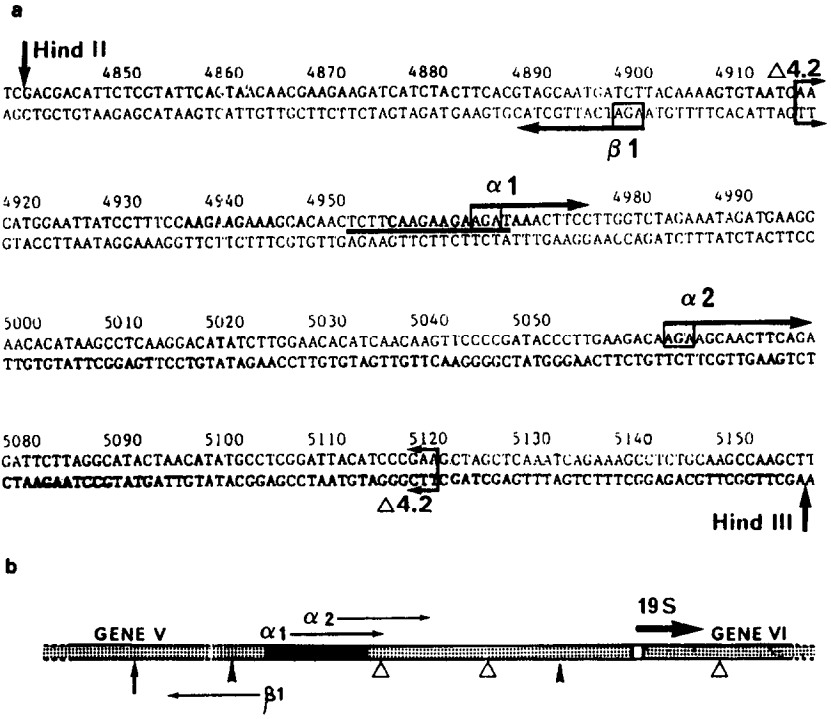

Fig. 3.- Localisation of "in vitro" initiation sites on the CaMV genome. (a) sequence of the region between the Hind sites at 4838 and 5151 bp. The alpha-1 and alpha-2 transcripts correspond respectively to the 192 and 93 nt protected fragments of fig. 2a, the beta-1 transcript to the unique fragment of fig. 2b. AGA sequences corresponding to the primer are boxed and the S1-hypersensitive sequences indicated by a solid line (▬). Restriction enzyme sites are shown, as are the limits of the deletion in plasmid pCa 4.2 (see text). (b) Localisation of the initiation sites on the viral genome. Bgl II (ϟ), Hind II (Ψ) and Hind III (▽) restriction sites are shown, as is the "in vivo" initiation site of 19 S RNA (transcript of gene VI) and the sequences deleted in pCa 4.2 (▬).

interaction with RNA polymerase II in the presence of a dinucleotide/purine nucleoside triphosphate combination (7). We do not observe selective interaction in the absence of the dinucleotide (our unpublished results), suggesting that the primer may allow the polymerase to initiate transcription at sites which do not normally serve as initiation sites but for which the enzyme has a certain affinity. In this respect, it is interesting to note that WILKINSON et al. (16) have observed that dinucleotide primers appear to circumvent the need for certain factors present in cell-free transcription systems. Further, ZHU et al. (19) have shown that sequences in the 5'-flanking region of a globin gene and which show similar nuclease hypersensitivity may serve as minor transcriptional starts "in vivo". The observation by these authors (19) and others (20) that similar patterns of hypersensitivity may be observed in chromatin and on supercoiled plasmids indicates that the torsional strain in the latter induces structures similar to those existing in chromatin and suggests that the hypersensitive structures adopted by the CaMV sequences on pCa 8 may also be present on viral minichromosomes (21).

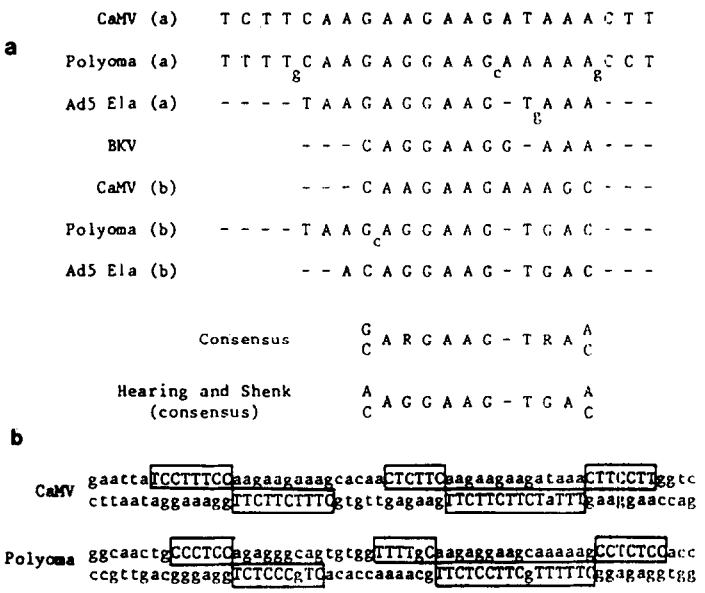

Fig. 4.- Homology between (ApG+ATP) primed "in vitro" transcription initiation sites on CaMV DNA and known viral and cellular enhancer sequences. (a) Under the CaMV sequences, containing the S1-hypersensitive region and the alpha-1 initiation, site are shown homologous sequences contained in enhancer regions of the adenovirus type 5 E1a gene (25) and the polyoma (23) and BKV (29) genomes. Dashes indicate missing nucleotides, supplementary nucleotides are indicated as subscripts. The consensus sequence derived from this alignment is compared with that of HEARING and SHENK (25). (b) Structural homology between the alpha-1 region of CaMV DNA and a polyoma enhancer region. Pyrimidine-rich sequences are shown in capital letters. CaMV coordinates from 4921, polyoma from 5166.

Nuclease hypersensitive sites, often within homopurine-homopyrimidine sequences, associated with an increasing number of viral and cellular genes have been described (22, and references therein) and have been proposed to play a role in the control of gene expression. We thus carried out a computer-assisted search for homologous sequences to those which we have detected in the CaMV genome on known viral and cellular gene sequences. Fig. 4a shows that the hypersensitive region shows a striking homology with a 25 bp region of the polyoma virus genome which shows the same structure of a purine-rich region between two blocks of pyrimidines and in which 19 bases are identical to the CaMV sequence and at three other positions the purine or pyrimidine nature of the base is conserved. Interestingly, the polyoma sequences lie within one of the two enhancer elements described by HERBOMEL et al. (23), overlapping the "SV40 enhancer" homologous region, and correspond to one of the nuclease-hypersensitive sites within this region of polyoma chromatin (24). Moreover, in addition to the strict sequence homology, the corresponding regions of the two viral genomes show a remarkable structure of alternating pyrimide-rich regions on the two DNA strands over a 50 bp stretch of the genome (Fig. 4b).

In addition to the extended region of homology between the CaMV and polyoma sequences, we detect a more limited homology between the two alpha-strand "in vitro" initiation sites and

known enhancer elements on other viral genomes (fig. 4a). These homologies cover an 11-12 base pair purine-rich sequence and correspond closely to the canonical adenovirus E1a enhancer consensus sequence first cited by HEARING and SHENK (25). It is interesting to note that the spatial separation (about 100 bp) between the repeated motifs of the polyma and adenovirus genomes is also observed between the initiation sites on CaMV DNA.

The above observations raise the interesting question as to whether the particular sequences which we have detected on the CaMV genome play a role in the expression of the viral genome. The possibility that a protein-coding sequence may play such a double role appears plausible in the light of the observation that a transcriptional enhancer sequence is located within the reverse transcriptase gene of the hepatitis B virus (26), a virus which shows certain structural and functional similarities to CaMV (27). While the lack of a suitable experimental system has so far prevented a direct test of the effect of the CaMV sequences on transcription, certain of our observations tend to suggest that they may influence gene expression. Firstly, we hare previously observed that, in the absence of the CaMV sequences, no transcription complexes are detected on the vector plasmid, pKC 7, under conditions in which both vector and insertion sequences of pCa 8 are efficiently transcribed (7) while, under the same conditions, transcription of a plasmid in which these sequences have been deleted is drastically reduced (8). These observations suggest the presence within the deleted region of sequences capable of increasing the level of transcription of the plasmids. Secondly, our results show a direct and selective interaction of RNA polymerase II with the CaMV sequences, an observation in agreement with the hypothesis that similar sequences on other genomes may play the role of RNA polymerase "entry sites" (28). The current availability of suitable vectors should allow us to test directly the effect of these sequences on gene expression.

ACKNOWLEDGEMENTS

This work was supported by grants from the Centre National de la Recherche Scientifique (CNRS). We thank Georges VILLELONGUE and Alain GOT for excellent technical assistance and Marie-Thérèse LACOSTE for typing the manuscript.

REFERENCES

1 LESCURE, B., WILLIAMSON, V. and SENTENAC, A. (1981). Nucleic Acids Res. 9, 31-45.
2 LESCURE, B., BENNETZEN, J. and SENTENAC, A. (1981). J. Biol. Chem. 256, 11018-11024.
3 LESCURE, B. (1983). J. Biol. Chem. 258, 946-952.
4 BALLARIO, P., BUONGIORNO-NARDELLI, M., CARNEVALI, F., di MAURO, E. and PEDONE, F. (1981). Nucleic Acides Res. 9, 3959-3978.
5 CARNEVALI, F., CASERTA, M. and di MAURO, E. (1983). J. Mol. Biol. 165, 59-77.
6 TSUDA, M. and SUZUKI, Y. (1982). J. Biol. Chem. 257, 12367-12372.
7 COOKE, R., PENON, P., GOT, C. and MIASSOD, R. (1983). Eur. J. Biochem. 137, 365-371.
8 COOKE, R. and PENON, P. (1986). Eur. J. Biochem. (in press).
9 TOH, H., HAYASHIDA, H. and MIYATA, T. (1983). Nature (Lond;) 305, 827-829.
10 VOLOVITCH, M., MODJTAHEDI, N., YOT, P. and BRUN, G. (1984). EMBO J. 3, 309-314.

11 WASYSLYK, B., KEDINGER, C., CORDEN, J., BRISON, O. and CHAMBON, P. (1980). Nature (Lond.) 285, 367-373.
12 MANIATIS, T., FRITISCH, E. F. and SAMBROOK, J. Eds; Molecular Cloning: A Laboratory Manual. Cold Spring Ha Laboratory, N.Y.
13 BERK, A.J. and SHARP, P.A. (1977). Cell, 12, 45-55.
14 MAXAM, A.M. and GILBERT, W. (1980). Methods Enzymol. 65, 499-560.
15 HOWELL, S.H. and HULL, R. (1978). Virology, 86, 468-481.
16 WILKINSON, J.A.K., MILLER, K.G. and SOLLNER-WEBB, B. (1983). J. Biol. Chem. 258, 13919-13928.
17 LEE, J.S., WOODWORTH, M.L., LATIMER, L.J.P. and MORGAN, A.R. (1984). Nucleic Acids Res. 12, 6603-6614.
18 SCHON, E., EVANS, J., WELSH, J. and EFSTRATIADIS, A. (1983). Cell 35, 837-848.
19 ZHU, J., ALLAN, M. and PAUL, J. (1984). Nucleic Acids Res. 12, 9191-9204.
20 LARSEN, A. and WEINTRAUB, H. (1982). Cell 29, 609-622.
21 OLSZEWSKI, N., HAGEN, G. and GUILFOYLE, T.J. (1982). Cell 29, 395-402.
22 REEVES, R. (1984). Biochim. Biophys. Acta 782, 343-393.
23 HERBOMEL, P., BOURACHOT, B. and YANIV, M. (1984). Cell 39, 653-662.
24 HERBOMEL, P., SARAGOSTI, S., BLANGY, D. and YANIV, M. (1981). Cell 25, 651-658.
25 HEARING, P. and SHENK, T. (1983). Cell 33, 695-703.
26 TOGNONI, A. CATTANEO, R., SERFLING, E. and SCHAFFNER, W. (1985). Nucleic Acids Res. 13, 7457-7472.
27 PFEIFFER, P. and HOHN, T. (1983). Cell 33, 781-789.
28 MOREAU, P., HEN, R., WASYLYK, B., EVERETT, R.D., GAUB, M.P. and CHAMBON, P. (1981). Nucleic Acids Res. 9, 6047-6068.
29 ROSENTHAL, N., KRESS, M., GRUSS, P. and KHOURY, G. (1983). Science 222, 749-755.
30 FRIED, M., GRIFFITHS, M., DAVIES, B., BJURSELL, G., LA MANTIA, G., and LANIA, L. (1983). Proc. Natl. Acad. Sci. USA, 80, 2117-2121.