Volume 10 Number 20 1982 Nucleic Acids Research

A tandem repeat gene in a picornavirus

Sonja Forss and Heinz Schaller

Microbiology, University of Heidelberg, Im Neuenheimer Feld 230, 6900 Heidelberg, FRG

Received 1 September 1982

SUMMARY

Three closely related genes for the small genome-linked protein (VPg) of picornaviruses have been identified by sequence analysis as a tandem repeat in the genome of Foot and Mouth Disease Virus (FMDV), strain O_1K. This unusual structure was also found in the genome of strain C_1O, belonging to a different FMDV serotype. Predicted biochemical properties of the three VPg gene products are in excellent agreement with the data from protein analysis of a heterogeneous VPg population from a third FMDV serotype, strain A_{10} (1). Taken together, these data indicate that the VPgs from all three genes function equally well *in vivo*. This is the first report of a tandem repeat gene in a viral genome.

INTRODUCTION

A small (2.4 k) basic protein is covalently bound to the 5'-end of the genomic RNA of all picornaviruses examined. In both poliovirus and FMDV this virus encoded genome-linked protein, termed VPg (2), is bound to the 5' uridyl residue of the RNA via a phosphodiester bond to a tyrosine residue (2-4). Poliovirus VPg has been extensively characterized, its amino acid sequence determined (5, 6), and the exact position and the nucleotide sequence of its gene in the viral RNA are known (7, 8). In FMDV a heterogeneity has been detected in the VPg population isolated from plaque purified virus, and two VPg fractions different in charge, tryptic peptide maps, and in amino acid composition have been separated (1). The reason for this heterogeneity and its possible significance were unclear.

As part of a general analysis of the structural organisation of the FMDV genome we have approached this problem by nucleotide sequence analysis of cloned cDNA from FMDV (9, 10) and a search for sequences that might code for a FMDV VPg peptide. As a re-

© IRL Press Limited, Oxford, England.

0305-1048/82/1020-6441$ 2.00/0

sult we have identified three tandemly arranged VPg genes in the FMDV genome, which code for three closely related but distinctly different VPgs.

MATERIAL AND METHODS

Restriction endonucleases were purchased from New England Biolabs, Boehringer Mannheim, and Bethesda Research Laboratories or purified as described by Roberts et al. (11).

Plasmids. Clone pFMDV 703, containing a cDNA insert from FMDV O_1K, has previously been described (9). Clones containing FMDV C_1O cDNA inserts in plasmid pBR322 were obtained from H. Küpper, Geneva, and were screened for VPg-encoding sequencesusing a 5'-labelled restriction fragment from the VPg region of clone pFMDV 703 (12). Two VPg-related clones were identified, pFMDV C_1-7 and pFMDV C_1-862. Plasmid DNA was amplified and isolated from *E. coli* C600 by standard procedures (13).

Nucleotide sequence analysis was performed as described by Maxam and Gilbert (14). The partial degradation products of the 5'-^{32}P-labelled restriction fragments were separated on thin gels (15), and further resolution of the autoradiograms was achieved by drying the gels prior to printing (16). Nucleotide sequences were stored and processed using the computer programs of Osterburg et al (17).

RESULTS AND DISCUSSION

Strategy for identification of VPg endocing sequences in the genome of FMDV

In previous experiments cDNA segments from FMDV strain O_1K were cloned and amplified in *E. coli*, and an approximate biochemical map of the FMDV gene products was established (9). Nucleotide sequences determined from these cloned cDNA fragments have been fused into a preliminary nucleotide sequence of the FMDV genome (18, 10). This sequence contains an open translational reading frame of about 7000 nucleotides as expected from the mode of protein synthesis employed by picornaviruses, which involves post-translational cleavage of a polyprotein translated from a single initiation site (19).

Amino acid sequencesderived from this preliminary nucleotide

sequence were screened for a putative gene for a VPg. This search was based on known biochemical data for FMDV Vpg (1, 20) and on the assumption that short, functionally related proteins from the same group of viruses, such as poliovirus and FMDV, retain a high degree of structural similarity during evolution. The following criteria were used: FMDV VPg is a small basic protein similar to poliovirus VPg which is 22 amino acids long with a positive charge of four (7). It contains a tyrosine residue (in poliovirus at position 3) for a covalent linkage to the viral RNA (1). In addition, glutamine/glycine cleavage sites were expected for the processing by a virus-specific protease (7). In contrast to poliovirus, at least part of the VPg population contains a methionine residue (1) and being a rare amino acid this is useful for our purpose.

Sequence analysis indicates the presence of three VPg genes in FMDV

In poliovirus VPg is encoded about two kilobases from the 3'-end of the viral RNA (7, 8), Sequence analysis of approximately corresponding regions in the FMDV genome was carried out with cloned cDNA from plasmid pFMDV 703 which covers this part of the FMDV genome (Fig. 1). On a close examination of the preliminary sequence data some homology to the polio sequence was detected on the level of the encoded amino acid sequences near the *Xho*I cleavage site at position 5500. This region was therefore sequenced carefully in both DNA strands as outlined in Fig. 1. In addition, the nucleotide sequence of a corresponding DNA segment was also determined in cloned cDNA from another serotype of FMDV, strain C_1O. The nucleotide sequences obtained, together with their derived amino acid sequences are shown in Fig. 2. As indicated here this sequence contains in its only open reading frame not only one, but three segments which have characteristics of a VPg as defined above. Some biochemical properties of these putative VPgs, designated VPg-1, VPg-2, and VPg-3, are summarized in Fig. 3 and Table 1 and are compared with the corresponding data from poliovirus. All FMDV sequences are similar in size (23 or 24 amino acids) and net charge (+3 or +4) to the polioprotein (22 amino acids and a charge of +4), and show significant sequence homology among each other and to

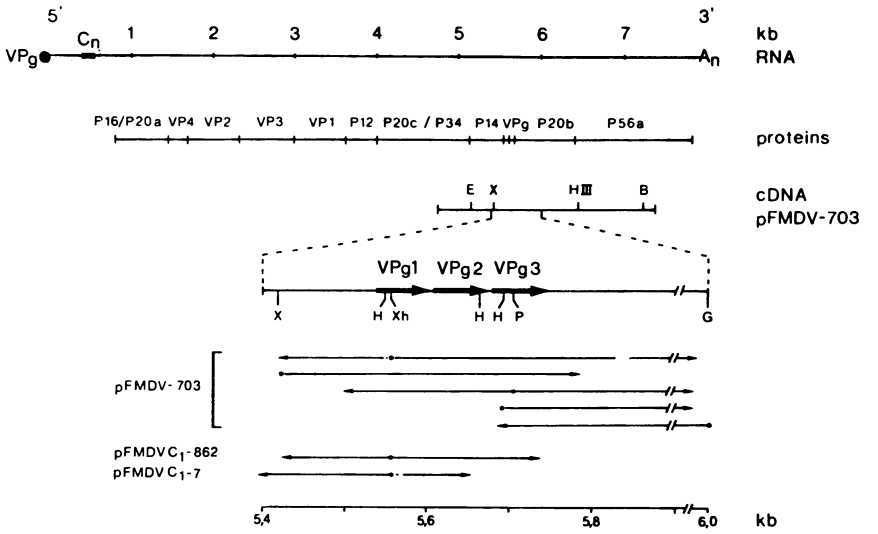

Fig.1: Description of the FMDV genome and localisation of the cDNA clone FMDV 703. Lengths of RNA and cDNA are given in kilobases (kb), using the numbering system of Küpper et al. (9). The protein map of Harris et al. (21) has been modified using data from King et al. (pers. communication) and our unpublished data. C_n and A_n refer to poly(C) and poly(A), respectively. VPg is illustrated as a filled circle at the 5'-end of the RNA. Position of restriction sites and strategy used for nucleotide sequence analysis of the VPg encoding regions of FMDV O_1K and FMDV C_1O are shown enlarged in the lower part of the figure. Thick arrows represent the putative VPgs (termed VPg-1, VPg-2, and VPg-3). Direction and extension of individual sequencing runs are outlined (thin arrows), starting from the ^{32}P-labelled 5'-ends (*). Letters used to designate restriction sites are: B, *Bam*HI; E, *Eco*RI; G, *Hga*I; H, *Hpa*II; HIII, *Hin*dIII; P, *Hph*I; X, *Xba*I; Xh, *Xho*I.

poliovirus VPg. All VPgs start with a glycine residue, and tyrosine is the third amino acid. However, unlike in polio, the FMDV VPgs terminate with glutamic acid and not with glutamine as does the polioprotein. This suggests that the FMDV specific protease cleaves Glu/Gly sequences and also a Glu/Ser sequence in the VPg region. This notion is also supported by processing sequences in the structural part of the FMDV genome which indicate that cleavages between coat proteins VP2/VP3 and VP3/VP1 occur at Glu/Gly and Glu/Thr linkages, respectively (10, 18, 22). In addition, Glu/Gly sequences have been found in the FMDV sequence at potential cleavage sites for other nonstructural

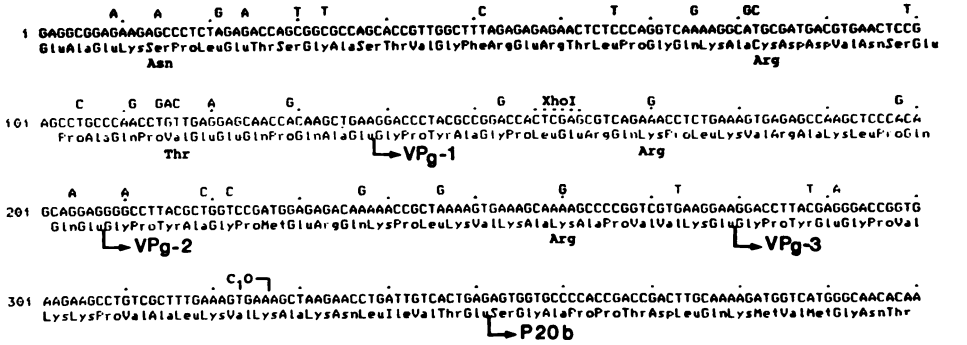

Fig. 2: Nucleotide and predicted amino acid sequences of cloned FMDV cDNA. The nucleotide sequence of the VPg encoding part of pFMDV 703 (O_1K) is presented with the deduced amino acid sequence from the only open reading frame. Analogous data was determined from strain C_1O, from pos. 1 - 326 (the end of the C_1O sequence is indicated), using the clones FMDV C_1-7 and FMDV C_1-862 (see Fig. 1). Nucleotides and amino acids of C_1O which differ from O_1K sequences are given above and below, respectively. The predicted limits for the three VPg genes are indicated by vertical arrows.

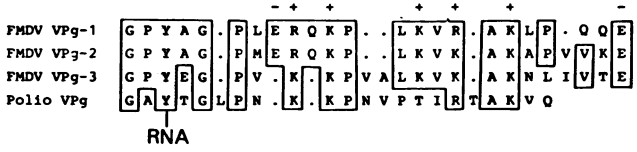

Fig. 3: Alignment of the amino acid sequences of the different FMDV VPgs (as predicted from the nucleotide sequence of FMDV O_1K) and poliovirus VPg (as determined by Kitamura et al. (7)). Homologous amino acids are boxed. Charged blocks are indicated by (+) or (-) signs. Dots represent insertions/deletions used to optimize alignment. Also indicated (shadowed) is the suggested link to the RNA, the tyrosine residue at position three.

Table 1: Biochemical properties of the three FMDV VPgs as derived from the nucleotide sequence of FMDV O_1K, and of VPg protein fractions isolated from FMDV strain A_{10} (1).

gene	net charge	contains Methionine	Leucine containing tryptic peptides	protein fraction
VPg-1 VPg-3	+3	no	GPYAGPLER, PLK, LPQQE PVALK, NLIVTE	VPgA
VPg-2	+4	yes	PLK	VPgB

proteins, whose limits were set by size estimations and by amino acid homologies to poliovirus proteins (10).

Three tandemly repeated VPg genes are also present in another FMDV serotype

To make sure that amplification of the VPg genes is not only confined to FMDV strain O_1K, and to eliminate possible mistakes that might have occured during cloning and in the sequence analysis of FMDV O_1K, we also determined nucleotide sequences in the VPg region of FMDV serotype C_1O. This strain is quite unrelated to strain O_1K as shown by extensive sequence variation in the structural genes which can be as high as 50 % and 30 % for nucleotide and amino acid sequences, respectively (23). For this purpose a cDNA bank from FMDV type C_1O (kindly made available to us by Dr. H. Küpper, Geneva) was screened for clones mapping in the VPg region, as described under METHODS. Two clones carrying VPg sequences were found (C_1-7 and C_1-862) and nucleotide sequences were determined from the *Xho*I site for the VPg encoding segment (Fig. 1). The 3'-terminal end of the third gene could not by analysed because it was not contained in either C_1O clone.

The sequence determined shows only a few base changes from the O_1K sequence in the VPg encoding segment (13/188 = 7 %). As shown in Fig. 2 most of these (eleven) are silent third base changes in the VPg related reading frame whereas the remaining two only cause a change between the functionally similar amino acids lysine and arginine (Fig. 2). This and the absence of stop codons indicate that the reading frame was chosen correctly, and that the deduced amino acid sequence is most likely of functional significance.

Biological significance of three VPg genes in FMDV

Since the degree of sequence conservation is high between the serotypes C_1O and O_1K we assumed that the amino acid sequence derived from O_1K could also be related to type A_{10} for which a heterogeneity in VPg had been reported by King at al. (1). According to them, two fractions of VPg in FMDV, VPgA and VPgB can be separated by electrofocusing. These fractions occur reproducibly in a 2:1 molar ratio in all FMDV strains analysed, and they differ in a number of biochemical criteria, such as charge, methionine content, and tryptic peptide maps. As outli-

ned in Table 1 all of these data can be explained on the basis of our O_1K sequence if we assume that VPg-1 and VPg-3 together constitute the more acid fraction VPgA, and VPg-2 the more basic minor fraction VPgB of King et al. Thus, VPg-2 (the VPgB fraction) has the same net charge as poliovirus VPg (+4), and VPg-1 plus VPg-2 (the VPgA fraction) carry one positive charge less. Only VPg-2 (VPgB) contains a methionine, while all three VPgs contain a tyrosine residue as required for genome linkage. Finally and most convincingly, there is a perfect agreement between the [^{3}H]leucine-labelled tryptic peptides obtained from the two VPg fractions by King et al. and the cleavage products predicted from our amino acid sequence. Thus, VPg-1 and VPg-3 together contain five such peptides which by charge and size distribution fit very well to the 2:1:1:1 pattern observed for the four peptides from fraction VPgA, whereas VPgB contains only a single leucine-containing tryptic peptide as expected for VPg-2. Again, as observed by King et al., this peptide (PLK, according to the predicted sequence) is identical to one of the VPgA peptides.

Several conclusions can be drawn from the perfect coincidence of the biochemical properties predicted for the FMDV VPgs from the nucleotide sequences and those observed for the VPg population from strain A_{10} by biochemical analysis (1).

- Firstly, FMDV strain A_{10} must also contain three VPg genes with a structure very similar to that described here for the strains from serotypes O and C. This prediction has very recently been confirmed by sequence analysis of the VPg region in cDNA from the A_{10} strain (D. Rowlands, pers. communication). Thus, the unique triplet gene structure for VPg seems to be a general feature of all aphtoviruses.
- Secondly, the limits for the individual genes were most likely set corectly which, as discussed before, suggests that Glu/Gly linkages, and not Gln/Gly linkages as in poliovirus, are the targets for the virus encoded protease in FMDV.
- Thirdly, all three VPgs appear to be utilized and to function equally well *in vivo*, since the RNA encapsidated in the virion carries the different VPgs in an equimolar ratio (1). This result further implies that neither variations in the amino acid

sequence between the individual VPgs nor their position in a common precursor protein (25, A. King, pers. communication) are of any importance for an efficient transfer of each VPg to the 5'-end of the viral RNA.

The problem how different forms of VPg could be generated in FMDV (1) has been unequivocally explained by the presence of three closely related, but distinctly different VPg genes. A similar amplification of a gene has not yet been reported for any viral genome, and it seems particularly difficult to explain how such an event might have occured in an RNA virus. However, it should be noted that marker rescue by recombinational events has been reported for FMDV (24), that the VPg genes are short and flanked by more or less exact repeat sequences coding for the processing sites, and that there are also direct repeat sequences within the VPg genes in both FMDV and poliovirus. Such sequences could have been used in rare recombination events to create a triplet repeat structure, which was stabilized in the following by further base changes without affecting VPg function in any of the genes. In the absence of any known specialized function for an individual VPg the latter finding indicates a selective advantage for a high VPg gene copy number in picornaviruses.

VPg gene amplification must have occured early in the evolution of picornaviruses, long before FMDV serotypes arose under the selective pressure of the immune system. This is indicated by the facts that the number of the VPg genes is constant between different FMDV serotypes, and that there is much more sequence variation between the individual VPg genes than between serotypes. In addition, there are many silent base changes (mostly transitions) in the codons for conserved amino acids which reach a similar level as observed between FMDV and poliovirus. From the structural homology VPg-3 is the closest to poliovirus VPg, and therefore, this may have been the ancestral FMDV VPg. However, sequences from other picornaviruses will be needed to establish a more reliable scheme for VPg evolution.

ACKNOWLEDGEMENTS

We thank A. King for a detailed discussion of both published and unpublished data on VPg, J. Wolters and O. Steinau for help

with the computer analysis. This work was supported by a research grant from BIOGEN S.A.

Preliminary reports of this study were presented during a meeting of the "European Study Group on the Molecular Biology of Picornaviruses", August 9 - 12, 1981, Hamburg; and the Spring Meeting of the "Deutsche Gesellschaft for Hygiene und Mikrobiologie, Sektion Virologie", March 25 - 27, 1982, Heidelberg.

REFERENCES

1. King, A.M.Q., Sangar, D.V., Harris, T.J.R., and Brown, F. (1980) J. Virol. 34, 627 - 634
2. Wimmer, E. (1982) Cell 28, 199 - 201
3. Rothberg, P.G., Harris, T.J.R., Nomoto, A., and Wimmer, E. (1978) Proc. Natl. Acad. Sci. USA 75, 4868 - 4872
4. Ambros, V. and Baltimore, D. (1978) J. Biol. Chem. 253, 5263 - 5266
5. Kitamura, N., Adler, C.J., Rothberg, P.G., Martinko, J., Nathenson, S.G., and Wimmer, E. (1980) Cell 21, 295 - 302
6. Semler, B.L., Anderson, C.W., Kitamura, N., Rothberg, P.G., Wishart, N.L., and Wimmer, E. (1981) Proc. Natl. Acad. Sci. USA 78, 3464 - 3468
7. Kitamura, N., Semler, B.L., Rothberg, P.G., Larsen, G.R., Adler, C.J., Dorner, A.J., Emini, E.A., Hanecak, R., Lee, J.J., van der Werf, A., Anderson, C.W., and Wimmer, E. (1981) Nature 291, 547 - 553
8. Racaniello, V.R. and Baltimore, D. (1981) Proc. Natl. Acad. Sci. USA 78, 4887 - 4891
9. Küpper, H., Keller, W., Kurz, Ch., Forss, S., Schaller, H., Franze, R., Strohmaier, K., Marquardt, O., Zaslavsky, V., and Hofschneider, P.H. (1981) Nature 289, 555 - 559
10. Forss, S. (1982) Thesis, University of Heidelberg
11. Roberts, R.J., Breitmeier, J.B., Tabacknik, N.F., and Myers, P.A. (1975) J. Mol. Biol. 91, 121 - 123
12. Grundstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961 - 3965
13. Clewell, D.B. and Helinski, D.R. (1969) Proc. Natl. Acad. Sci. USA 62, 1159 - 1165
14. Maxam, A.M. and Gilbert, W. (1980) Methods in Enzymol. 65, 499 - 560
15. Sanger, F. and Coulson, A.R. (1978) FEBS Lett. 87, 107 - 110
16. Garoff, H. and Ansorge, W. (1981) Analyt. Biochem. 115, 454 - 457
17. Osterburg, G., Glatting, K.H., and Sommer, R. (1982) Nucleic Acids Res. 10, 207 - 216
18. Kurz, Ch., Forss, S., Küpper, H., Strohmaier, K., and Schaller, H. (1981) Nucleic Acids Res. 9, 1919 - 1931
19. Sangar, D.V. (1979) J. gen. Virol. 45, 1 - 13
20. Sangar, D.V., Rowlands, D.J., Harris, T.J.R., and Brown, F. (1977) Nature 268, 648 - 650
21. Harris, T.J.R., Brown, F., and Sangar, D.V. (1981) Virology, 112, 91 - 98
22. Boothroyd, J.C., Harris, T.J.R., Rowlands, D.J., and Lowe, P.A. (1982) Gene 17, 153 - 161

23. Beck, E., Feil, G., and Strohmaier, K. (1982) submitted for publication
24. King, A.M.Q., McCahon, D., Slade, W.R., and Newman, J.W.I. (1982) Cell, 29, 921 - 928
25. Semler, B.L., Anderson, C.W., Hanecak, R., Dorner, L.F., and Wimmer, E. (1982) Cell, 28, 405 - 412.