Volume 236, number 2, 287–290 FEB 06214 August 1988

# Sobemovirus genome appears to encode a serine protease related to cysteine proteases of picornaviruses

Alexander E. Gorbalenya, Eugene V. Koonin, Vladimir M. Blinov and Alexei P. Donchenko

*Institute of Poliomyelitis and Viral Encephalitides, USSR Academy of Medical Sciences, 142782 Moscow Region, USSR*

Received 23 June 1988

A putative serine protease was identified among non-structural proteins of southern bean mosaic virus (SBMV) by sequence comparison with cellular and viral proteases. The predicted SBMV protease displayed a significant similarity to cysteine proteases of picornaviruses, providing a possible evolutionary link between the two enzyme classes. It is suggested that SBMV follows the general expression strategy characteristic of other positive-strand RNA viruses containing 5'-terminal covalently linked proteins (VPg), i.e. generation of functional proteins by polyprotein processing.

Amino acid sequence comparison; Serine protease; Cysteine protease; Positive-strand RNA virus; Enzyme evolution

## 1. INTRODUCTION

Controlled proteolytic processing of polyproteins is one of the major strategies of eukaryotic positive strand RNA virus genome expression. This strategy has been definitely demonstrated for picorna-, como-, poty-, alpha- and flaviviruses, and is suspected for some other groups of viruses [1,2]. In the former 3 groups virus-encoded cysteine proteases have been unequivocally identified, their active centers have been localized by sequence comparison (as well as, in some cases, experimentally), and the cleavage specificities have been determined [3–5]. In alphaviruses, whose capsid protein is known to possess autoproteolytic activity, a domain of this protein was identified as a serine protease [6].

Sequence comparisons have previously proved helpful in analysis of viral proteases, having allowed us to predict that a second protease ($2A^{Pro}$) should be encoded by the poliovirus genome [7]. This prediction has been later confirmed by independent experiments [8]. We have also demonstrated that picornaviral proteases displayed distant but significant sequence similarity to both cellular serine and cysteine proteases [9]. It was suggested that they could provide an evolutionary link between the two classes of proteases.

Here we tentatively identify a serine protease in still another positive strand RNA plant virus, southern bean mosaic virus (SBMV), the type member of the sobemovirus group. Comparison of its sequence with those of other viral and cellular proteases appears to substantiate the above hypothesis and provides clues for understanding SBMV expression strategy.

*Correspondence address:* A.E. Gorbalenya, Institute of Poliomyelitis and Viral Encephalitides, USSR Academy of Medical Sciences, 142782 Moscow Region, USSR

*Abbreviations:* SBMV, southern bean mosaic virus; RV89, rhinovirus type 89, PV1, poliovirus type 1, EMCV, encephalomyocarditis virus (picornaviruses); CPMV, cowpea mosaic virus (comovirus); TVMV, tobacco vein mottling virus (potyvirus). THR, thrombin, CHT, chymotrypsin, TRP, trypsin, PTRPG, protrypsinogen, SGPA, *Streptomyces griseus* protease A, ALP, bacterial $\alpha$-lytic protease (serine proteases); $3C^{Pro}$, picornaviral 3C protease

## 2. METHODS

### 2.1. *Sources of protein sequences*

Protein sequences were extracted from current literature. Cellular serine protease sequences were from [10], those of

*Published by Elsevier Science Publishers B.V. (Biomedical Division)*
00145793/88/$3.50 

picorna- (except HRV89), como- and potyvirus proteases being from [11]. The sequence of RV89 was from [12] and that of SBMV from [13].

2.2. *Protein sequence comparisons*

Protein sequences were compared by a version of the program DIAGON based on the MDM78 amino acid residue scoring matrix [14]. All possible pairs of amino acid residue blocks of varying length ($l$; $11 < l < 45$) from 2 protein sequences were compared, and those matching with the double matching probability [14] below a defined threshold ($p$) were recorded by dots. Each dot corresponded to the center of a segment of length $l$. The program was written in the C programming language and run in the Unix operation system on a WicatS150 computer.

## 3. RESULTS AND DISCUSSION

### 3.1. *Identification of the putative SBMV protease*

Bearing in mind the results of in vitro translation studies of SBMV RNA demonstrating precursor–product relationships between some of the polypeptides produced [15], we compared the sequence of the 105 kDa protein (P1) encoded by the largest ORF (ORF2) of SBMV with those of picorna-, como- and potyviral proteases. Upon such comparison, a region of P1 has been identified displaying a striking similarity to picornaviral 3C proteases (fig.1A). In the segment of $3C^{Pro}$ around the catalytic Cys residue this similarity reached a highly significant level, higher in fact than that between the proteases of some picornaviruses of different genera which are homologous beyond any doubt (cf. A and B in fig.1). Comparison of the SBMV sequence to those of serine proteases revealed a much lower degree of similarity, although in some cases at least a segment of serine proteases around the catalytic Ser was picked up as the most similar (fig.1C). Strikingly, however, in the SBMV sequence a Ser residue is substituted for the catalytic Cys of picornaviral $3C^{Pro}$. Alignment of sequence stretches centering around the (putative) catalytic residues of picorna- and sobemoviral proteases revealed a number of coincidences and homologous replacements (fig.2A). The similarity between cellular serine proteases and the putative SBMV form was less pronounced in this region, in agreement with the results of DIAGON comparison (fig.2A). Inspection of the SBMV P1 sequence upstream of this site revealed segments closely resembling those around the catalytic His and Asp residues of serine proteases (fig.2B); the distances between these

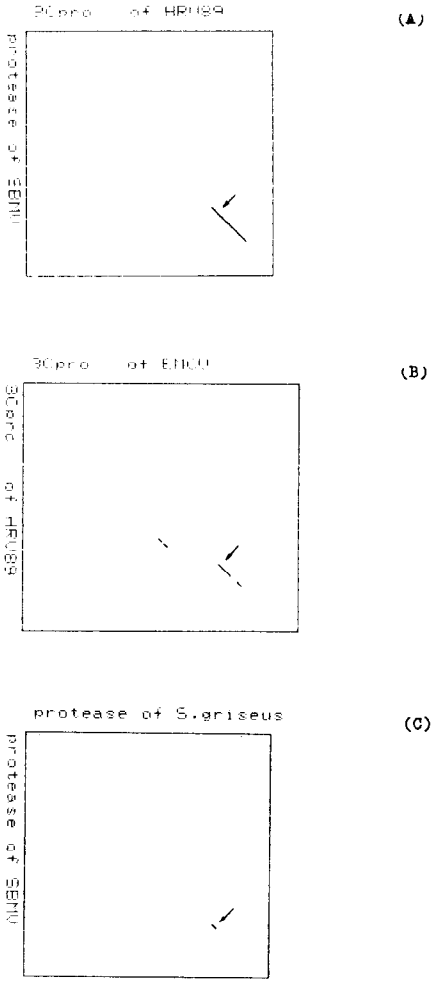


Fig.1. Comparison of amino acid sequences of the putative protease of SBMV with those of cellular and viral proteases. Comparisons obtained at $p = 10^{-4}$ and $l = 27$ (see section 2), which in all cases was close to the optimal $l$ value, are shown: (A) SBMV P1, residues 145–325 vs HRV89 $3C^{Pro}$; (B) HRV89 $3C^{Pro}$ vs EMCV $3C^{Pro}$; (C) SBMV P1, residues 145– 325 vs SGPA. Arrows point to the streaks corresponding to matched segments containing the (putative) catalytic Ser (Cys) residues.

segments and the putative catalytic Ser residue are very similar to those in serine proteases. These observations suggest that the respective portion of the SBMV P1 may be an unusual serine protease related both to 'classical' (chymotrypsin-like) serine proteases and to viral cysteine proteases.

Previously, we detected a distant similarity between the catalytic segments of picornaviral $3C^{Pro}$ and those of cellular serine proteases and suggested

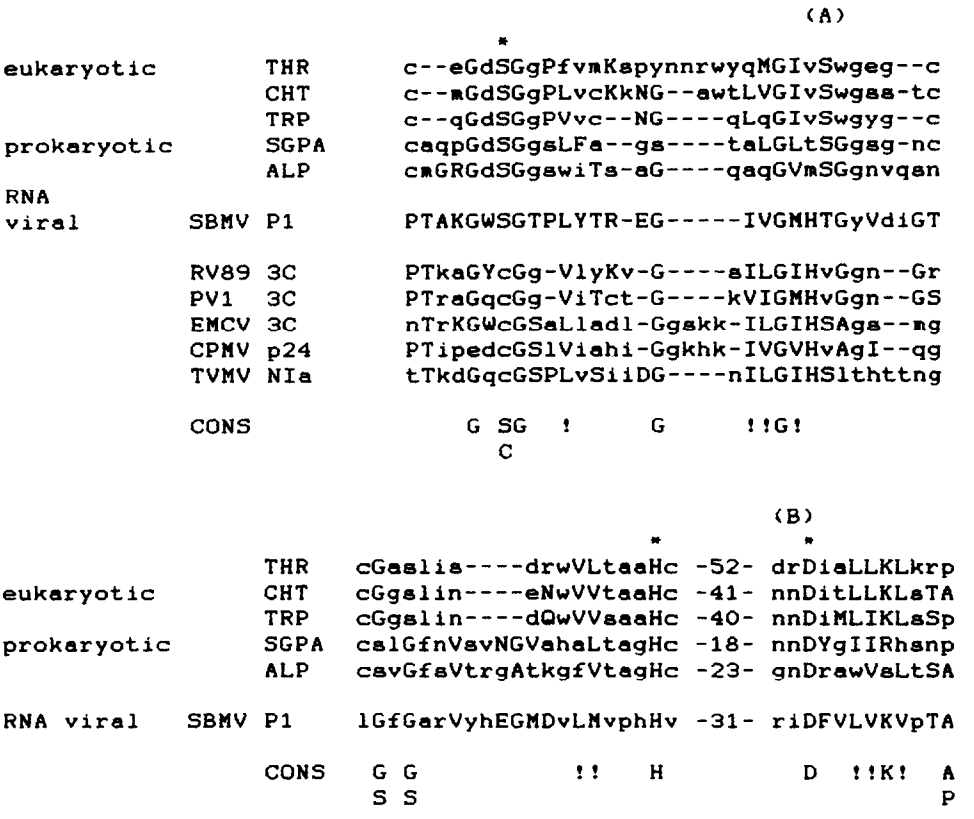


Fig.2. Alignment of conserved segments of viral and cellular proteases. (A) C-terminal fragments (residues 278–306 of SBMV P1; for numbering in other sequences see respective references); (B) N-terminal fragments (residues 161–224 of SBMV P1). Asterisks designate amino acid residues implicated in catalysis. Capital letters denote residues identical or functionally homologous to those in the respective positions of the putative protease of SBMV. Residues belonging to one of the following groups were counted as functionally homologous: G,A; S,T; D,E.N,Q; K,R; I,L,V,M; F,Y,W. Below the aligned stretches the patterns of highly conserved residues are shown, designated CONS for consensus; ! indicates a hydrophobic residue.

that viral enzymes could be evolutionarily related to both cellular serine and cysteine ones [12]. The putative SBMV protease may well be the 'missing link' in this chain, being a '3C-like' serine enzyme.

### 3.2. *Possible cleavage specificity of SBMV protease and implications for SBMV expression strategy*

The resemblance of the putative SBMV protease to picornaviral forms may give some clues as to the cleavage specificity of the former. Picornaviral proteases cleave primarily at Q,E/G,S.A sites [2,3]. Inspection of the SBMV P1 sequence in the vicinity of the catalytic sites of the putative protease reveals 2 E/S sites flanking a polypeptide of 271 amino acid residues (fig.3); both sites are preceded by similar dipeptides, GQ and AQ. Cleavage at these sites would give, in addition to the putative protease, a polypeptide of about 60 kDa, comprising the C-terminal part of P1 and closely resembling in size the P2 protein observed upon in vitro translation of SBMV RNA and shown to be derived from P1 [15]. Alternatively, the protease could be generated by cleavage at 2 E/T sites, resulting in a protein of 181 amino acid residues (fig.3). Curiously, this coincides with the characteristic size of picornaviral proteases [11].

Thus, it is likely that organization of functional domains in P1 resembles that in C-terminal portions of polyproteins of other VPg-containing viruses [11] and that the strategy of SBMV genome expression is generally similar to those utilized by

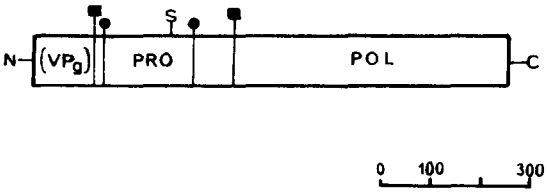


Fig.3. Scheme demonstrating proposed domain organization in the product of the long open reading frame (P1) of SBMV. PRO, protease; POL, RNA-dependent RNA polymerase. E/S (squares) and E/T (circles) sites flanking the putative protease domain with a catalytic Ser residue (S*) are shown. The position of VPg is proposed by analogy to other VPg-containing viruses [11].

these viruses, based on polyprotein processing mediated by a virus-encoded protease(s).

*Acknowledgements:* The authors are grateful to Professor V.I. Agol for critical reading of the manuscript and helpful discussions and to Dr K.M. Chumakov for providing the program DIAGON.

## REFERENCES

[1] Strauss, E.G. and Strauss, J.H. (1983) Curr. Top. Microbiol. Immunol. 105, 1-98.
[2] Wellink, J.E. and Van Kammen, A. (1988) Arch. Virol. 98, 1-26.
[3] Nicklin, M.J.H., Toyoda, H., Murray, M.G. and Wimmer, E. (1986) Bio/Technology 4, 33-42.
[4] Garcia, J.A., Schrijvers, L., Tan, A., Vos, P., Wellink, J. and Goldbach, R. (1987) Virology 159, 67-75.
[5] Carrington, J.C. and Dougherty, W.G. (1987) J. Virol. 61, 2540-2548.
[6] Strauss, J.H., Strauss, E.G., Hahn, C.S., Hahn, Y.S., Galler, R., Handy, W.R. and Rice, C.M. (1987) in: Positive Strand RNA Viruses (Brinton, M.A. and Rueckert, R.R. eds) pp. 209-225, A.R. Liss, New York.
[7] Blinov, V.M., Donchenko, A.P. and Gorbalenya, A.E. (1985) Dokl. Akad. Nauk SSSR 281, 984-987.
[8] Toyoda, H., Nicklin, M.J.H., Murray, M.G., Anderson, C.W., Dunn, J.J., Studier, F.W. and Wimmer, E. (1986) Cell 45, 761-770.
[9] Gorbalenya, A.E., Blinov, V.M. and Donchenko, A.P. (1986) FEBS Lett. 194, 253-257.
[10] Craik, C.S., Rutter, W.J. and Fletterick, R. (1983) Science 220, 1125-1129.
[11] Zimmern, D. (1988) in: RNA Genetics (Holland, J.J. et al. eds) CRC Press, Boca Raton, FL, in press.
[12] Duechler, M., Skern, T., Sommergruber, W., Neubauer, C., Gruendler, P., Fogy, I., Blaas, D. and Kuechler, E. (1987) Proc. Natl. Acad. Sci. USA 84, 2605-2609.
[13] Wu, S., Rinehart, C.A. and Kaesberg, P. (1987) Virology 161, 73-80.
[14] Staden, R. (1982) Nucleic Acids Res. 10, 2951-2961.
[15] Mang, K., Ghosh, A. and Kaesberg, P. (1982) Virology 116, 264-274.