Volume 252, number 1,2, 42–46 FEB 07419 July 1989

Tentative identification of RNA-dependent RNA polymerases of dsRNA viruses and their relationship to positive strand RNA viral polymerases

E.V. Koonin, A.E. Gorbalenya* and K.M. Chumakov

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

Received 29 May 1989

Amino acid sequence stretches similar to the four most conserved segments of positive strand RNA viral RNA-dependent RNA polymerases have been identified in proteins of four dsRNA viruses belonging to three families, i.e. P2 protein of bacteriophage $\phi 6$ (Cystoviridae), RNA 2 product of infectious bursa disease virus (Birnaviridae), $\lambda 3$ protein of reovirus, and VP1 of bluetongue virus (Reoviridae). High statistical significance of the observed similarity was demonstrated, allowing identification of these proteins as likely candidates for RNA-dependent RNA polymerases. Based on these observations, and on the previously reported sequence similarity between the RNA polymerases of a yeast dsRNA virus and those of positive strand RNA viruses, a possible evolutionary relationship between the two virus classes is discussed.

Sequence comparison; dsRNA-containing virus; RNA virus, positive-strand; RNA-dependent RNA polymerase; Virus evolution

1. INTRODUCTION

The relationships between different classes of RNA viruses are far from clear. In particular, the origin of dsRNA-containing viruses is obscure. Such a crucial aspect of genomic replication and expression strategy as the presence of a virion transcription/replication apparatus associated with the non-infectivity of RNA genomes might relate them to negative strand RNA viruses [1,2]. Extensive sequencing of viral genomes has led to the possibility of addressing this problem directly. Very recently, sequence similarity has been revealed between certain regions of (putative) RNA-dependent RNA polymerases of yeast dsRNA virus and IBDV, and highly conserved segments of positive strand RNA viral polymerases [3,4]. Here, based on comparison with the same enzyme class, we tentatively identify the RNA-dependent RNA polymerase of three other dsRNA viruses, $\phi 6$ bacteriophage, reovirus, and bluetongue virus, extend our observations on the (putative) IBDV polymerase, and briefly discuss the relationships between the polymerases of different virus classes, and between viruses themselves.

Correspondence address: E.V. Koonin, Institute of Microbiology, USSR Academy of Sciences, 7 Prospekt 60-letiya Oktyabrya, 117819 Moscow, USSR

Abbreviations: CV, coxsackie virus type B4; EMCV, encephalomyocarditis virus; FMDV, foot-and-mouth disease virus; TEV, tobacco etch virus; SNBV, Sindbis virus; WNV, West Nile virus; TRV, tobacco rattle virus; CMV, cucumber mosaic virus; CarMV, carnation mottle virus; BNYVV, beet necrotic yellow vein virus; IBV, infectious bronchitis virus; GA, GA bacteriophage (positive strand RNA viruses); IBDV, infectious bursa disease virus; $\phi 6$, $\phi 6$ bacteriophage; REO, reovirus; BTV, bluetongue virus; ScV, *Saccharomyces cerevisiae* virus (dsRNA viruses)

2. METHODS

2.1. *Amino acid sequences and alignments*

The amino acid sequences of bacteriophage $\phi 6$ proteins were from [5–7], and those of IBDV proteins from [8,9]. The sequence of reovirus (type 3) $\lambda 3$ protein was from [10] where source references for other reovirus proteins can be found, and the sequence of VP1 of BTV was from [11]. The alignment of

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

conserved segments of 41 (putative) positive strand RNA viral RNA-dependent RNA polymerases was an updated version (in preparation) of that published [12]. The alignments of RNA-dependent RNA polymerases of negative strand RNA viruses were from [13–15].

2.2. *Amino acid sequence comparisons*

Amino acid sequences of proteins of dsRNA viruses were searched for patterns conserved in RNA-dependent RNA polymerases of positive and negative strand RNA viruses using the pattern-searching program SRCH and by visual inspection. Sequence stretches containing these patterns were manually fitted into the respective alignments. The statistical significance of the alignments was assessed by the program SCORE. The average alignment score per residue was calculated for a comparison of a query sequence with an alignment using the MDM78 matrix [16]. This calculation was simulated with 300 randomly scrambled versions of the query sequence, and two values of adjusted alignment score were calculated in standard deviation (SD) units: AS1 = $S^a - S^r/\sigma$, and AS2 = $S^a - S^{mr}/\sigma$ where S^a denotes the score actually observed, S^r and S^{mr} represent the mean and maximal random scores, respectively, and σ is the SD. The aligned sequences were also compared using the program COMP generating a distance matrix for an alignment using the formula [17]:

$$D_{ij} = -\ln(S - S^r/S^m - S^r)$$

where D_{ij} is the distance between sequences i and j, S represents the comparison score for these sequences, S^r is the expected score for two random sequences of the same composition, and S^m is the average of the scores obtained upon comparison of each sequence with itself. Scores were calculated using the MDM78 matrix. Initial pairwise sequence comparison was by the program DOTHELIX generating a full map of local similarity (A.E.G. et al., in preparation), and alignment was by the program OPTAL [18,19].

3. RESULTS AND DISCUSSION

3.1. *Identification of putative RNA-dependent RNA polymerases of dsRNA viruses and of their functionally important sites*

Inspection of amino acid sequences of proteins of four dsRNA viruses, ϕ6 bacteriophage, IBDV, reovirus, and bluetongue virus revealed stretches similar to all four highly conserved segments of positive strand RNA viral RNA-dependent RNA polymerases ([12] and unpublished) in protein P2, RNA 2 product, λ3 protein, and VP1, respectively (fig.1). In the reovirus protein, segment II was apparently duplicated, leading to two versions of the alignment in this region (fig.1). No comparable similarity to the conserved segments of RNA polymerases of negative strand RNA viruses was detectable in dsRNA viral proteins. All the amino acid residues invariant in positive strand RNA viral polymerases could be identified in dsRNA viral proteins as well as a number of partially conserved residues. Also, the lengths of the spacers separating the conserved segments were almost fully within the range determined by the positive strand RNA viral polymerases (with the exception of the distance between segments I and II in the BTV protein). This suggested that the respective proteins might be the RNA polymerases involved in dsRNA replication, the regions similar to the conserved segments of positive strand RNA viral polymerases being important for polymerase functions. As reovirus and BTV both belong to the Reoviridae family and have virtually identical genome organizations and expression strategies [20], it was reasonable to compare the sequences of their putative polymerases in more detail. Comparison by the program DOTHELIX indeed revealed a moderate but convincing similarity which was most prominent in the central parts of both proteins encompassing the segments shared with positive strand viral polymerases (not shown). Alignment of these regions by the program OPTAL (fig.2) showed significance at the approx. 7 SD level. There was, however, a complication in that the BTV sequence corresponding to region II of positive strand polymerases was aligned with a reovirus sequence different from (though overlapping with) those initially suggested (fig.1). Despite deviating significantly from the positive strand RNA viral consensus, this sequence could still be another candidate for this region in reovirus. To address this problem more adequately, sequences from other members of the Reoviridae are necessary. Roy and co-workers [11] claimed that the sequence of VP1 protein of BTV was related to those of the eukaryotic and vaccinia virus DNA-dependent RNA polymerase largest subunits. However, neither the sequences shared by BTV and reovirus (see above), nor those typical of DNA-dependent RNA polymerases [21,22] were conserved in their alignment. In contrast, we revealed here the conservation in BTV of exactly the most conserved segments of positive strand RNA virus polymerases. Moreover, of the four (putative) dsRNA viral polymerases shown in fig.1, that of BTV most closely conformed to the positive strand virus consensus. In this respect, our identification seemed more convincing than that

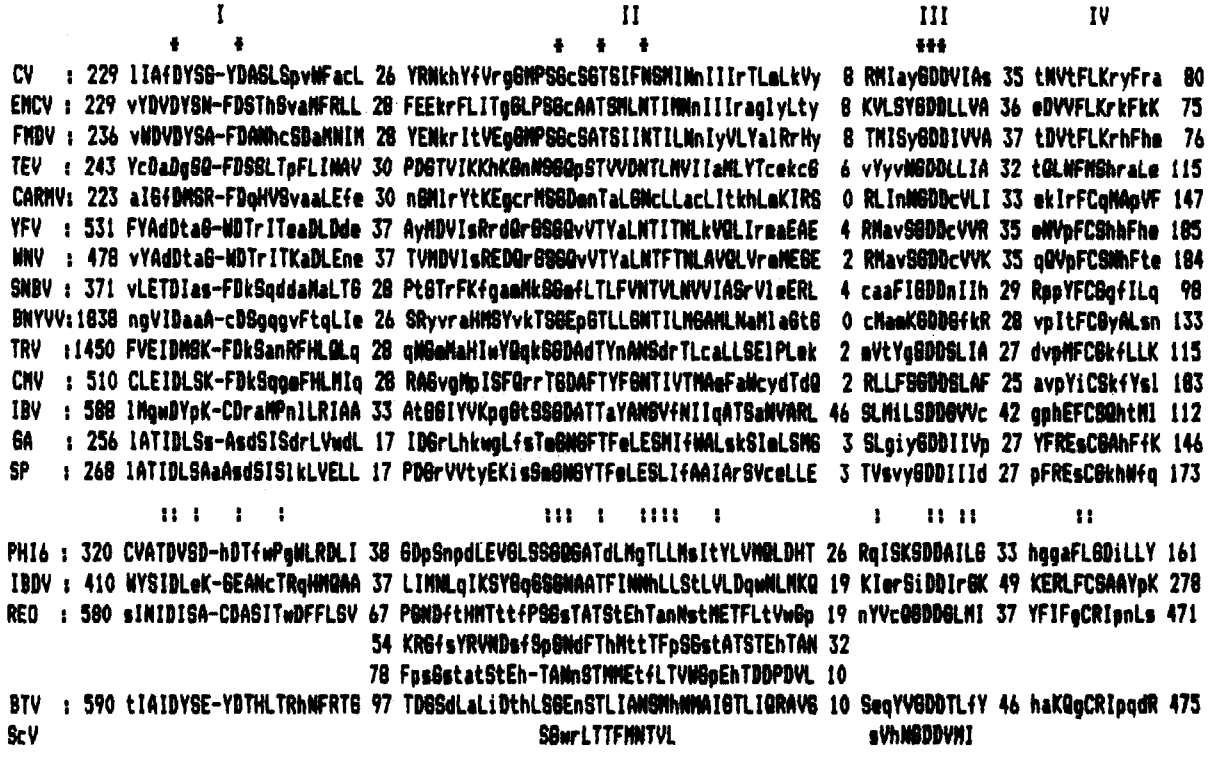

Fig.1. Alignment of fragments of putative RNA-dependent RNA polymerases of dsRNA viruses with the four conserved segments (I–IV) of selected polymerases of positive strand RNA viruses. Capitals highlight amino acid residues identical or similar (i.e. belonging to one of the following groups: V,L,I,M; F,Y,W; G,A; S,T; D,E,N,Q; R,K) in proteins of the two virus classes. Colons indicate the positions where identical or similar residues occurred in half or more of the presented sequences of each class. Asterisks designate the consensus residues of positive strand RNA virus polymerases [12]. Numbers of amino acid residues between protein termini and conserved segments, and between the latter are indicated. In BNYVV, precise boundaries of the polymerase are not known, and distances to the polyprotein termini are indicated. In IBV, polymerase boundaries have been tentatively identified by comparison of putative cleavage sites [23]. For reovirus, three possible versions of the alignment of segment II are shown (see text).

suggested by Roy et al., and the latter might be of doubtful significance.

A quantitative evaluation of the alignment in fig.1 showed that, for the putative polymerases of the four dsRNA viruses, the average scores obtained upon comparison with the aligned conserved segments of positive strand RNA viral polymerases fell within the range typical of the latter (with the exception of the last version of the reovirus alignment). Moreover, these scores exceeded not only the mean but also the maximal values obtained with randomly scrambled sequences at a statistically significant level (table 1). Interestingly, the most convincing values were obtained with BTV where our identification was in conflict with that in the original paper [11].

Inspection of the distance matrix generated for the alignment of the conserved segments of positive strand RNA viral and dsRNA viral (putative) polymerases showed that, for the latter, affinities existed with certain groups of positive strand RNA viral polymerases that were even closer than those between most of the groups within the positive class (table 2). The present identification of dsRNA viral polymerases agreed with the available experimental evidence in that, for all the viruses, the products of the respective genomic segments have been implicated in RNA synthesis [23–26]. Moreover, for reovirus, it had been shown that L1 genomic segment determined the pH optimum of RNA synthesis, indicating that λ3 protein might be the polymerase proper [25].

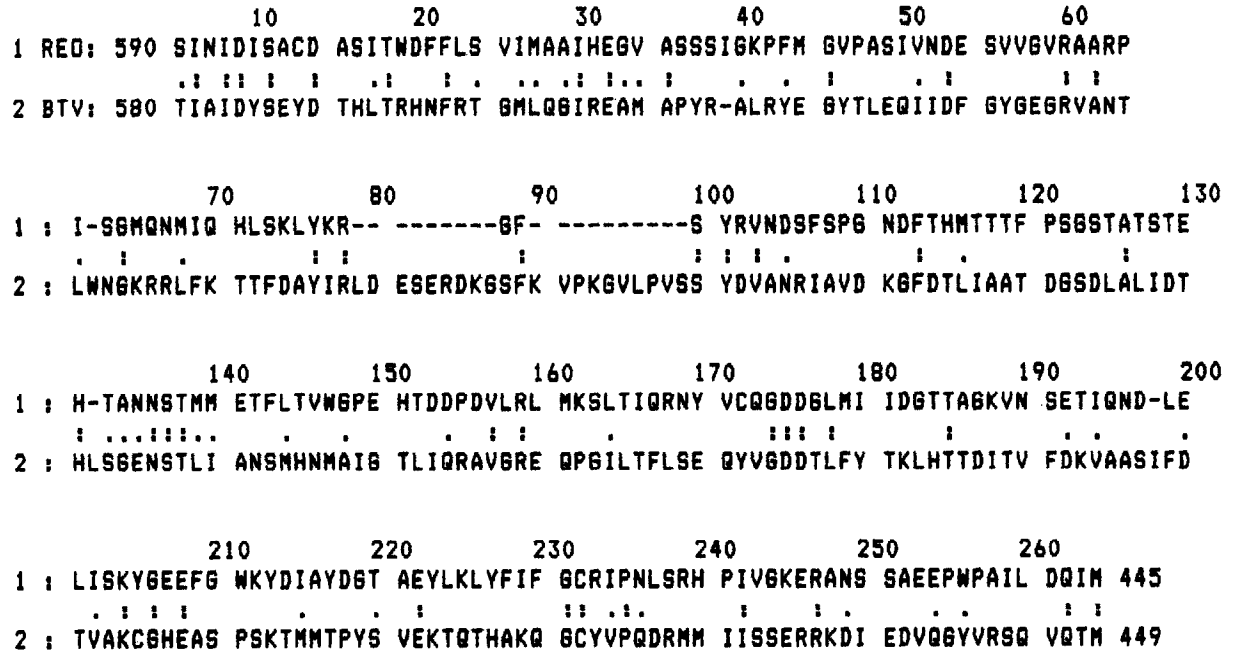

Fig.2. Alignment of the central portions of the putative polymerases of reovirus and BTV. Colons highlight identical residues, dots designating similar residues defined as in fig.1. The distances to the protein termini are indicated.

Table 1

Alignment scores for comparison of the putative polymerases of dsRNA viruses with positive strand RNA viral polymerases

	S^a	AS1 (SD)	AS2 (SD)
Positive strand RNA viruses	9.83–10.92	NC	NC
ϕ6	10.07	7.69	4.34
IBDV	10.01	7.33	4.46
REO	9.94	6.74	3.62
	9.97	7.10	4.21
	9.62	5.72	2.92
BTV	10.13	8.74	5.55

Scores were calculated as described in section 2 for the 4 segments shown in fig.1 (85 residues total). For positive strand RNA viruses, each sequence was extracted from the alignment of 41 sequences of (putative) polymerases and compared to the remaining 40 sequences. This procedure is subject to a bias toward higher values because of the presence of several groups of closely related sequences (e.g. those of picornaviruses) within the reference set. The dsRNA viral sequences were compared to the whole reference set of 41 sequences. For reovirus, the calculations were performed separately for the three versions of the alignment shown in fig.2. NC, not calculated

3.2. *Implications for the origin of dsRNA viruses*

We have tentatively identified, by sequence comparison, RNA polymerases of dsRNA viruses of three very different families and showed that they are related to the polymerases of positive strand RNA viruses, and not to those of negative strand RNA viruses. Recently, similar observations have been reported for the polymerase of yeast dsRNA virus ([3], and fig.1). These observations are somewhat unexpected in view of the similarities in genomic strategies between dsRNA viruses and negative strand RNA viruses (see section 1). The question as to whether all dsRNA viruses, some of which differ strikingly in structure and properties, share a common ancestry is intriguing. The putative polymerases of the three groups analyzed in this work are less similar to each other than to some of the positive strand RNA viral polymerases (cf. table 2). This raises the possibility that different groups of dsRNA viruses could originate from different groups of positive strand RNA viruses. On the complete sequences of other viral

Table 2

Distances between the sequences of conserved segments of putative polymerases of dsRNA viruses and selected polymerases of positive strand RNA viruses

	WNV	SNBV	TRV	CV	TEV	IBV	$\phi 6$	IBDV	REO	BTV
GA	1.02	0.97	1.09	1.11	1.48	1.17	1.59	1.25	1.15	1.18
WNV		0.89	0.82	0.99	1.02	1.13	1.34	0.94	1.19	1.11
SNBV			0.69	1.08	1.04	0.82	1.22	1.38	1.13	0.94
TRV				1.12	0.89	1.04	1.33	1.04	1.06	1.00
CV					0.89	0.95	0.91	1.36	1.09	1.21
TEV						1.01	1.01	1.35	1.18	1.24
IBV							1.21	1.11	1.11	1.11
$\phi 6$								1.58	1.73	1.82
IBDV									1.39	1.52
REO										0.78

D_{ij} values calculated as indicated in the text are shown. For reovirus, the calculations were for the first (shown in the upper line in fig.1) version of the alignment

genomes becoming available, it will be possible to scrutinize this hypothesis further.

Acknowledgements: The authors thank Drs L.I. Brodsky, A.P. Donchenko and A.M. Leontovich for providing some of the computer programs used in this work.

REFERENCES

[1] Baltimore, D. (1971) Bacteriol. Rev. 35, 235–245.
[2] Ishihama, A. and Nagata, K. (1988) CRC Crit. Rev. Biochem. 23, 27–76.
[3] Pietras, D.F., Diamond, D.F. and Bruenn, J.A. (1988) Nucleic Acids Res. 16, 6225.
[4] Gorbalenya, A.E. and Koonin, E.V. (1988) Nucleic Acids Res. 16, 7735.
[5] McGraw, T., Mindich, L. and Frangione, B. (1986) J. Virol. 58, 142–151.
[6] Gottlieb, P., Metzger, S., Romantschuk, M., Carton, J., Strassman, J., Bamford, D., Kalkkinen, N. and Mindich, L. (1988) Virology 163, 183–190.
[7] Mindich, L., Nemhauser, I., Gottlieb, P., Romantschuk, M., Carton, J., Frucht, S., Strassman, J., Bamford, D. and Kalkkinen, N. (1988) J. Virol. 62, 1180–1185.
[8] Hudson, P.J., McKern, N.M., Power, B.E. and Azad, A.A. (1986) Nucleic Acids Res. 14, 5001–5012.
[9] Morgan, M.M., Macreadie, I.G., Harley, V.R., Hudson, P.J. and Azad, A.A. (1988) Virology 163, 240–242.
[10] Wiener, J. and Joklik, W.K. (1989) Virology 169, 194–203.
[11] Roy, P., Fukusho, A., Ritter, G.D. and Lyon, D. (1988) Nucleic Acids Res. 16, 11759–11767.
[12] Koonin, E.V., Gorbalenya, A.E., Chumakov, K.M., Donchenko, A.P. and Blinov, V.M. (1987) Mol. Genet. (Moscow) 7, 27–39.
[13] Shioda, T., Iwasaki, K. and Shibuta, H. (1986) Nucleic Acids Res. 14, 1545–1563.
[14] Blumberg, B.J., Crowley, J.C., Silverman, J.I., Menonna, J., Cook, S.D. and Dowling, P.C. (1988) Virology 164, 487–497.
[15] Tordo, N., Poch, O., Ermine, A., Keith, G. and Rougeon, F. (1988) Virology 165, 565–576.
[16] Dayhoff, M.O., Barker, W.C. and Hunt, L.T. (1983) Methods Enzymol. 91, 519–545.
[17] Feng, D.F., Johnson, M.S. and Doolittle, R.F. (1985) J. Mol. Evol. 21, 112–125.
[18] Pozdnyakov, V.I. and Pankov, Y.A. (1981) Int. J. Peptide Protein Res. 17, 284–291.
[19] Gorbalenya, A.E., Blinov, V.M., Donchenko, A.P. and Koonin, E.V. (1989) J. Mol. Evol. 28, 256–268.
[20] Joklik, W.K. (1983) The Reoviridae, Plenum, New York.
[21] Broyles, S.S. and Moss, B. (1986) Proc. Natl. Acad. Sci. USA 83, 3141–3145.
[22] Kock, J., Evers, R. and Cornelissen, A.W.C.A. (1988) Nucleic Acids Res. 16, 8753–8772.
[23] Muller, H. and Nitschke, R. (1987) Virology 159, 174–177.
[24] Sinclair, J.F. and Mindich, L. (1976) Virology 75, 209–217.
[25] Drayna, D. and Fields, B. (1982) J. Virol. 41, 110–118.
[26] Mertens, P.P.C., Brown, F. and Sangar, D.V. (1984) Virology 135, 207–217.
[27] Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P. and Blinov, V.M. (1989) Nucleic Acids Res., in press.