Nucleic Acids Research, 1993, Vol. 21, No. 24 5667–5669

A closely related group of RNA-dependent RNA polymerases from double-stranded RNA viruses

Jeremy A.Bruenn
Department of Biological Sciences, State University of New York, Buffalo, NY 14260, USA

Received August 30, 1993; Revised and Accepted October 31, 1993 EMBL accession nos V01059 and V01060

ABSTRACT

Probably one of the first proteinaceous enzymes was an RNA-dependent RNA polymerase (RDRP). Although there are several conserved motifs present in the RDRPs of most positive and double-stranded RNA (dsRNA) viruses, the RDRPs of the dsRNA viruses show no detectable sequence similarity outside the conserved motifs. There is now, however, a group of dsRNA viruses of lower eucaryotes whose RDRPs are detectably similar. The origin of this sequence similarity appears to be common descent from one or more non-infectious viruses of a progenitor cell, an origin that predates the differentiation of protozoans and fungi. The cause of this preservation of sequence appears to be constraints placed on the RDRP by the life-style of these viruses—the maintenance of a stable, persistent, noninfectious state.

INTRODUCTION

Only one gene is common to all RNA viruses: a gene for an RNA-dependent RNA polymerase (RDRP), or, in some cases, an RNA-dependent DNA polymerase (reverse transcriptase). There is enough sequence conservation of several motifs within the RDRP that these have been used to identify this gene in many viruses for which no biochemical evidence is available to define this gene product. A number of attempts to define relationships among the RNA viruses have used sequence comparisons of the RDRP (1, 2, 3, 4, 5). Among the double-stranded RNA (dsRNA) viruses, this has been difficult, since they are a very disparate group without easily detected sequence similarity over the entire length of their RDRPs. For instance, the original publication describing the sequence of the RDRP encoding dsRNA of reovirus failed to identify it as encoding an RDRP (6), and the Phi6 RDRP was reported as having no similarity to known RDRPs (7). However, both of these RDRPs do have several of the conserved motifs characteristic of this enzyme (1, 2, 3, 4, 8).

A new group of dsRNA viruses from lower eucaryotes, however, demonstrates a very conserved RDRP. The peculiar (noninfectious) nature of these viruses leads to a possible explanation for this conservation in a common, ancient origin for RNA-dependent RNA polymerases.

MATERIALS AND METHODS

Programs from the GCG package (9, 10) were used for pairwise sequence alignment and statistical tests (GAP), for multiple sequence alignment (PILEUP), for generation of dendrograms (PILEUP), and for determination of sequence similarity as a function of position in multiple alignments (PLOTSIMILARITY). PHYLIP (11) was used to verify dendrograms. BLAST analyses (12) were done by Michael Hogan at the Genetics Computer Group to evaluate the statistical significance of alignments, and SYSTAT (13) was used for cluster analysis of the BLAST data. The sequences of UmVH1 (GenBank accession number V01059) and ScVLa (GenBank accession number V01060) were determined from cDNA clones in this lab and those of TvV (a partial sequence), GlV, and BcV were obtained as personal communications (see Table 1 for references).

RESULTS

Recently, the sequences of RDRPs from a number of dsRNA viruses of lower eucaryotes have been determined (14, 15, 16, 17, 18, 19). Five of these are spherical, non-infectious viruses, with protein capsids with primarily one capsid polypeptide, and a single essential viral dsRNA. All are recognizably related and their sequences easily aligned by the alignment programs GAP or PILEUP (9, 10). By the GAP evaluation of this set, the most disparate of the five noninfectious viruses of lower eucaryotes are the *Ustilago maydis* virus H1 (UmVH1) and the *Saccharomyces cerevisiae* virus L1 (ScVL1), which have 23.5% sequence identity over the entire length of their RDRP sequences (1020 amino acids with gaps). Their GAP alignment has a 'quality' more than 10 standard deviations above the mean of a random alignment, which is highly significant. The RDRPs of the two protozoan viruses LRV1 and TvV are closely related (28% sequence identity) as are those of the two fungal viruses ScVL1 and ScVLa (also 28% sequence identity). The RDRP of a sixth dsRNA virus of lower eucaryotes, the *Giardia lamblia* virus (GlV) is also possibly similar, with 23.4% sequence identity to the *Leishmania* virus LRV1 (the ordered alignment 3 standard deviations above the mean of random alignments). This sixth virus differs from the other dsRNA viruses of lower eucaryotes in that it is infectious (20, 21).

The sequences of 15 RDRPs of dsRNA viruses in all pairwise combinations were compared by making each a BLAST data set

Table 1. Conserved motifs in RDRPs of dsRNA viruses of lower eucaryotes

RDRP	1	2	3	4	5	6	7	8
LRV1	LLGRG	59 WAANGS.HS	49 GKTRLLL	57 DVDDFNSQHT	46 TLMSGHRATSFINSVLNRAYI	11 HVGDDILM	33 EFLRV	9 YLAR
TvV	LLGRG	58 WSKSGS.HY	45 GKERFIV	50 DYTDFNSQHT	43 TLPSGHRATTFINPVLNWCYT	11 CAGDDVIL	31 EFLRK	9 YPCR
ScVL1	LMNRG	57 WVPGGSVHS	50 GKQRAIV	52 DVDDFNSQHS	52 TLLSGWRLTTFMNTVLNWAYM	15 HNGDDVMI	33 EFLRV	13 YLSR
ScVLa	LENGV	58 IMPGGSVHS	50 GKVRALV	51 DFDDFNSQHS	52 TLFSGWRLTTFFNTALNYCYL	13 HNGDDVFR	33 EFLRV	11 YLTR
UmVH1	LYGRG	66 WLVSGSSAG	61 GKARAIV	55 DVPDFNSMHT	63 GLYSGDRDTTLINTLLNIAYA	20 CHGDDIIT	34 EYLRI	10 CLAR
	*	**	*	* **** *	* ** * * * ** *	***	* **	* *
GIV	LLGKV	65 WGTTGSGYI	41 TKVRAVI	55 DQSNFDRQPD	59 GLPSGWKWTALLGALINTQLL	16 VQGDDIAL	33 EFLRR	13 MMIK
BcV		GPPGGETHM	44 TKVRGVW	52 DWSSFDSSVT	52 GLPSGSVYTSIVGSVVNRLRI	16 TQGDDSLI	35 TFLGR	9 SLDK

The sequences shown are the conserved motifs within the RDRP sequences located in Fig. 1 from the *Leishmania guyanesis* virus LRV1 (16), the *Saccharomyces cerevisiae* virus ScVL1 (15), the *Saccharomyces cerevisiae* virus ScVLa (17), the *Ustilago maydis* virus UmVH1 (18), the *Trichomonas vaginalis* virus TvV (19), and the *Giardia lamblia* virus GlV (14). The five known sequences from noninfectious dsRNA viruses of lower eucaryotes are shown in the first five lines. Residues identical in each of these sequences are indicated by asterisks. The lower two lines show the conserved motifs in GlV and in the most closely related dsRNA virus outside this group, BcV, the beet cryptic virus (22). Motifs 4, 5, 6, and 8 were previously designated motifs IV, V, VI and VII, respectively, by Koonin (1, 2) and motifs 4, 5, and 6 were previously designated motifs 1, 2, and 3, respectively, by Bruenn (4). Motif 7 may be present in some positive strand RNA viruses (5). The number of residues between the conserved motifs is indicated in each case.

and comparing each of the other sequences to it. The only statistically significant matches (except for infectious bursal disease virus, IBDV, and infectious pancreatic necrosis virus, IPNV, which are both birnaviruses) were among the five RDRPs of the non-infectious dsRNA viruses (ScVL1, ScVLa, UmVH1, LRV1, and TvV: see Table 1 for abbreviations). These had MSP (maximal segment pair) scores from 53 to 238, where scores above 55 are considered 'distinguishable from chance similarities' (12). Except for the UmVH1-LRV1 pair (MSP score of 53), all had MSP scores of 69 or above. Clearly, accepted statistical tests place the RDRPs of the non-infectious dsRNA viruses of lower eucaryotes into a monophyletic group.

In contrast, none of the other RDRPs of the dsRNA viruses (except as noted for IBDV and IPNV) are detectably related: all have MSP scores lower than 47. For instance, a GAP alignment of the reovirus and ScVL1 RDRPs over the central conserved region of 600 amino acids (see below) detects only 18% identity and generates an alignment with the same quality as a random alignment (0 standard deviations above the mean by GAP, or a BLAST MSP score of 30). Similarly, an alignment of the reovirus and bluetongue virus RDRPs over the same region detects only 18% identity and generates an alignment with a quality only 1 standard deviation above the mean of random alignments, or an MSP score of 30. This lack of similarity among the dsRNA viruses of higher eucaryotes has been noted previously (2).

Only two RDRPs from this outgroup have possible similarity to the RDRPs of the noninfectious dsRNA viruses of lower eucaryotes: BcV, or beet cryptic virus (22) and GlV (14). For instance, an alignment of the BcV RDRP with that of GlV over their complete sequences detects 23% identity with a quality 2 standard deviations above the mean of random alignments (MSP score of 48), and an alignment of GlV with ScVL1 detects 18% identity with 3 standard deviations above the mean of random alignments (MSP score of 50). Interestingly, BcV, like most of the dsRNA viruses of lower eucaryotes, is also noninfectious.

The automated sequence alignment of the RDRPs of the dsRNA viruses of lower eucaryotes is straightforward. A plot of the similarity along these six sequences aligned by PILEUP is shown as Fig. 1. There are clearly 8 peaks of similarity in the central portion of the RDRP (about 600 amino acids with gaps). These correspond to the 8 conserved motifs of Table 1, of which several have been previously identified in the RDRPs of positive strand and dsRNA viruses (see Table 1). Mutagenesis

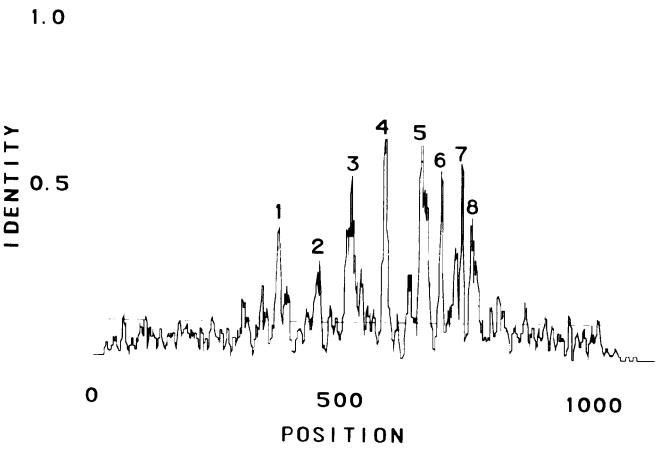

Figure 1. Similarity as a function of length of the RDRPs of lower eucaryotes. This a PLOTSIMILARITY (9, 10) figure generated from the output of PILEUP (10), applied to the entire sequences of the RDRPs of the six known sequences of RDRPs of lower eucaryotes. The origins of the sequences are given in Table 1. On the ordinate, a value of 1 corresponds to identity at each amino acid in the window of 7 amino acids at a given position in each RDRP. The peaks of similarity are the conserved motifs of Table 1.

experiments have shown that sequence conservation within at least two of these conserved motifs (5 and 6) parallels function in the ScVL1 RDRP (23). In contrast, a computer alignment using PILEUP of the 15 RDRPs of dsRNA viruses at large fails to correctly align even the most conserved regions of most of these RDRPs (motifs 4 and 5), even when the central 600 amino acids of the RDRP sequences are used. Such an alignment results in a maximum similarity score (see Fig. 1) of 0.2 (20% identity), while the alignment of the six RDRPs of dsRNA viruses of lower eucaryotes (over the full range of more than 1000 amino acids with gaps) gives a maximum similarity score of 0.65 (65% identity; see Fig. 1). In addition, motifs 1, 2, 3, and 7 are missing from the RDRPs of dsRNA viruses of higher eucaryotes. Note that motif 1 is missing from BcV, but that even the spacing between motifs is well conserved in all seven RDRPs.

Phylogenetic trees derived from sequence alignments of the entire amino acid sequences of the RDRPs of the dsRNA viruses by PILEUP or from 68 amino acids including motifs 5 and 6 by

PILEUP and by PHYLIP (11) or derived from the MSP scores calculated by BLAST and analyzed by the Euclidian distance cluster analysis algorithm of SYSTAT (13) confirm that the five non-infectious dsRNA viruses of lower eucaryotes are more closely related to each other than they are to any other dsRNA viruses, and that GlV and BcV may be related to this group (data not shown).

DISCUSSION

Why are the noninfectious dsRNA viruses of lower eucaryotes much more closely related to each other than are the dsRNA viruses at large? This is probably not the result of RNA recombination, since these viruses replicate in very different hosts and have no infectious cycle. Even ScVLa and ScVL1, which replicate in the same host cells, share only 28% sequence identity in the RDRP region. The most likely explanation for these data is that these five viruses have a common origin. If this origin were recent, the progenitor virus would have had to be infectious to an extremely diverse group of organisms from at least five phyla and subsequently become noninfectious to each host independently. This seems very unlikely, although there is at least one case in which a non-infectious dsRNA virus-like agent appears to have been derived from an infectious single-stranded RNA virus (24). A more likely explanation is that the original virus was a noninfectious virus (or group of viruses) in a single cell type, and that this cell type gave rise to both protozoans and fungi. This model is further supported by data showing that a second *U.maydis* noninfectious dsRNA virus (UmVH2) is more closely related to the ScV viruses than to UmVH1, showing that ScV and UmVH2 probably existed prior to the divergence of *U.maydis* and *S.cerevisiae* (C.M.Park and J.A.Bruenn, unpublished data).

These viruses appear to be of ancient origin, since protozoans and fungi diverged very early in evolution (25), at the time of divergence of animals and fungi, or perhaps even earlier, at the divergence of plants and fungi. This group of viruses should then be widespread in the fungi and protozoans, as current data indicate it is.

Why have these viruses retained detectable sequence similarity over such a long period while none is apparent among most of the other dsRNA viruses? I postulate strong selective pressure to preserve sequences that have been lost in other dsRNA viruses. This could be accounted for by the necessity among the noninfectious dsRNA viruses to preserve interactions with highly conserved cellular proteins. For instance, there are at least three cellular genes required for ScVL1 replication (26). A similar suggestion has been made to explain conservation of sequence among the RDRPs of single-stranded RNA viruses (27).

This model is consistent with the proposed polyphyletic origins of the dsRNA viruses (2): the dsRNA viruses of lower eucaryotes would constitute one large subgroup of monophyletic origin. The possible relationship between the noninfectious dsRNA viruses of lower eucaryotes and GLV would be explained if GLV were derived from a member of this group that had only recently become infectious. The possible relationship to a plant virus, BcV, which is noninfectious, might be explained by an earlier branching of plants from the line that gave rise to the protozoans and fungi in question. The only known noninfectious dsRNA virus of higher animal cells (28) may also have an RDRP detectably similar to those of this group.

ACKNOWLEDGEMENTS

I thank Michael Hogan at the Genetics Computer Group for performing the BLAST analyses, John Antoniw, Alice Wang, Jung-Hsiang Tai, and Eugene Koonin for communicating results prior to publication, and Ian Baldwin and Patricia Wainright for reading the manuscript. Support was from NSF grants DMB-9106818 and KOSEF INT-9020780, USDA grant 92-37303-8310 and BARD grant IS-1928-91R.

REFERENCES

1. Koonin, E.V. (1991). J. Gen. Virol. **72**, 2197–3206.
2. Koonin, E.V. (1992). Seminars in Virol. **3**, 327–339.
3. Poch, O., Sauvaget, I., Delarue, M. and Tordo, N. (1989). EMBO J. **8**, 3867–3874.
4. Bruenn, J.A. (1991). Nucleic Acids Res. **19**, 217–226.
5. Koonin, E.V. and Dolja, V.V. (1993). Crit. Rev. Biochem. Molec. Biol. in press,.
6. Wiener, J.R. and Joklik, W.K. (1989). Virology **169**, 194–203.
7. Mindich, L., Nemhauser, I., Gottlieb, P., Romantschuk, M., Carton, J., Frucht, S., Strassman, J., Bamford, D.H. and Kalkkinen, N. (1988). J. Virol. **62**, 1180–1185.
8. Morozov, S.Y. (1989). Nucl Acids Res. **17**, 5394.
9. Devereux, J., Haeberli, P. and Smithies, O. (1984). Nucleic Acids Res. **12**, 387–395.
10. Genetics Computer Group (1991). Program Manual for the GCG Package, Version 7. Madison, WI. GCG.
11. Felsenstein, J. (1993). PHYLIP. U. Washington. U.Washington.
12. Altschul, S.F., Gish, W., Miller, W. and Myers, E.W. (1990). J. Molec. Biol. **215**, 403–410.
13. Wilkinson, L. (1989). SYSTAT: The System for Statistics. SYSTAT, Inc. Evanston, IL.
14. Wang, A.L., Yang, H.-M., Shen, K.A. and Wang, C.C. (1993). Proc. Natl. Acad. Sci. USA in press.
15. Diamond, M.E., Dowhanick, J.J., Nemeroff, M.E., Pietras, D.F., Tu, C.-L. and Bruenn, J.A. (1989). J. Virol. **63**, 3983–3990.
16. Stuart, K.D., Weeks, R., Guilbride, L. and Myler, P.J. (1992). Proc. Natl. Acad. Sci. USA **89**, 8596–8600.
17. Park, C.-M., Lopinski, J., Tzeng, T.-H. and Bruenn, J.A. (unpublished). Genbank accession no. V01060.
18. Park, C.-M., Han, J.-O. and Bruenn, J.A. (unpublished). Genbank accession no. V01059.
19. Tai, J.-H. (1993). Personal communication.
20. White, T.C. and Wang, C.C. (1990). Nucl. Acids Res. **18**, 553–9.
21. Furfine, E.S. and Wang, C.C. (1990). Mol. Cell. Biol. **10**, 3659–3662.
22. Xie, W.S., Antoniw, J.F. and White, R.F. (1993). J. Gen. Virol. in press.
23. Ribas, J.C. and Wickner, R.B. (1992). Proc. Natl. Acad. Sci. USA **89**, 2185–2189.
24. Koonin, E.V., Choi, G.H., Nuss, D.L., Shapira, R. and Carrington, J.C. (1991). Proc. Natl. Acad. Sci. USA **88**, 10647–10651.
25. Wainright, P.O., Hinkle, G., Sogin, M.L. and Stickel, S.K. (1993). Science **260**, 340–342.
26. Wickner, R.B. (1989). Faseb J **3**, 2257–65.
27. Koonin, E.V. and Gorbalenya, A.E. (1989). J. Molec. Evol. **28**, 524–527.
28. Scott, M.P., Fostel, J.M. and Pardue, M.L. (1980). Cell **22**, 929–941.