*Nucleic Acids Research, 2003, Vol. 31, No. 7* ***1821–1829***
*DOI: 10.1093/nar/gkg277*

## SURVEY AND SUMMARY

# A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases

**Jeremy A. Bruenn***

Department of Biological Sciences, State University of New York, Buffalo, NY 14260, USA and Hauptman Woodward Medical Research Institute, 73 High Street, Buffalo, NY 14203, USA

Received October 21, 2002; Revised December 2, 2002; Accepted January 22, 2003

## ABSTRACT

**A systematic bioinformatic approach to identifying the evolutionarily conserved regions of proteins has verified the universality of a newly described conserved motif in RNA-dependent RNA polymerases (motif F). In combination with structural comparisons, this approach has defined two regions that may be involved in unwinding double-stranded RNA (dsRNA) for transcription. One of these is the N-terminal portion of motif F and the second is a large insertion in motif F present in the RNA-dependent RNA polymerases of some dsRNA viruses.**

## INTRODUCTION

The RNA-dependent RNA polymerase (RdRp) probably evolved very early, given that it is the one essential protein encoded by all RNA viruses. Numerous comparisons of the RdRps have been made in attempts to construct a phylogeny of RNA viruses and understand the enzymology of DNA and RNA polymerases (1–11). There are now several structures of RdRps available: the RdRp of Phi6, a double-stranded RNA (dsRNA) bacteriophage (11); the RdRp of hepatitis C virus (HCV), a flavivirus (12–14); the RdRp of poliovirus (polio), a picornavirus (10,15); and the RdRp of rabbit hemorrhagic disease virus (RHDV), a calcivirus (16). Co-crystallization with nucleoside triphosphates or with oligonucleotides has mapped substrate-binding sites, and the binding of $Mg^{2+}$ or $Mn^{2+}$ has mapped the active site of the enzyme.

While all the RNA and DNA polymerases share a basic structure, the RdRps are much more similar to each other than they are to other polymerases. A number of investigators have noted that beyond several conserved motifs, there is no primary sequence conservation among the RdRps of the RNA viruses at large, or among those of the dsRNA viruses (5,11). However, within some families of RNA viruses, there is both enough sequence conservation to permit statistically significant alignments and enough sequence diversity to permit efficient identification of conserved motifs (5). This work has now been extended and supplemented with structural comparisons, confirming the identification of a new conserved motif apparently present in all RdRps which was previously identified in the Totiviridae (5,17).

## MATERIALS AND METHODS

Programs from the GCG package (18,19) were used for pairwise alignment (GAP), for multiple sequence alignment (PILEUP), for determination of sequence similarity as a function of position in multiple alignments (PLOTSIMILARITY), and for creation of datasets for assigning statistical weights to similarities (GCGTOBLAST). Searches for similar sequences used BLAST (20) or PSIBLAST (21). Modeling from X-ray crystal data used MOLSCRIPT (22).

## RESULTS

A systematic way of finding conserved motifs in protein families is multiple sequence alignment. This technique works best when the sequences being compared are limited to those with between 25 and 50% sequence identity to each other. Sequences more closely related than this provide no information on essential regions but simply skew the results towards a subgroup of sequences. Sequences less related than this are usually not successfully aligned by programs such as PILEUP(23). A selection of a dozen RdRps with between 25 and 50% identity (BLAST E values of between ~$10^{-4}$ and $10^{-35}$) to the poliovirus type 3 RdRp (24) was chosen from a BLAST search (20) and aligned by PILEUP (23). The result was analyzed by PLOTSIMILARITYand is labeled Picornaviridae in Figure 1. The highest peaks of similarity include the standard conserved motifs A to E (labeled in the figure) as described previously (3,9,10,25,26). Similar analyses of RdRp sequences of the Totiviridae (BLAST E values between 0.3 and $10^{-35}$) and of sequences related to the Togaviridae (BLAST E values between 0.1 and $10^{-10}$) are also shown in Figure 1. A small selection of negative strand RdRps is shown in the last panel of Figure 1. These RdRps also have the same six conserved motifs, although their sequences are somewhat different than the consensus sequences of the positive strand and dsRNA virus RdRps (Table 1). Clearly, this analysis identifies the same conserved regions in every RdRp, with the exception of some extra peaks of similarity in

*To whom correspondence should be addressed at State University of New York, Buffalo, NY 14260, USA. Tel: +1 716 645 2363; Fax: +1 716 645 2975; Email: cambruen@nsm.buffalo.edu

 

some groups. For instance, there is a prominent peak of similarity between F and A in the Picornaviridae-related sequences. These extra peaks show no obvious sequence similarity from group to group.

The six conserved regions that are so easily visible in this analysis have been identified previously by multiple sequence analysis (3,9,25,26) or by structural comparisons (10,12). The total span of regions F (see below) to E is about the same in each group of positive strand and dsRNA viruses (250–280 amino acids) and slightly more in the negative strand RNA viruses.

This method of defining important regions in proteins is validated by structural comparison. The structures of the Phi6, HCV and RHDV RdRps in the region including conserved motifs A to D are shown in Figure 2. The poliovirus RdRp structure is not shown because it is less complete than that of RHDV, which has essentially the same structure (16) and is 27% identical in sequence with the poliovirus RdRp. The conserved motifs A to D are shown in dark blue in each RdRp. Clearly, the three polymerases are very similar in this region. The positions of the conserved motifs of the Phi6, HCV and RHDV RdRps, identified by primary sequence alignment (Table 1, Figs 1 and 3), correspond exactly to the structurally homologous positions identified by comparison of the three-dimensional structures of the polymerases (Fig. 2).

The identification of conserved motifs in the HCV RdRp is not as straightforward as for the Picornaviridae-related RdRps, since it belongs to a class in which only a few sequences satisfying the criterion of 25–50% identity are available for comparison. This makes the identification of motifs D and E, independent of structural information, more difficult in this case. With HCV, comparison with bovine viral diarrhea virus (BVDV, 23% identity) (27,28) successfully maps all six conserved motifs. Similarly, alignment of HCV (29), HGV (30), HGBV (31) and BVDV (27,28) successfully maps all but motif D (Fig. 3; Flaviviridae). Motif D is EAGK in BVDV. Mutation of the terminal lysine of motif D to an alanine in the BVDV RdRp greatly reduces polymerase activity (28). However, this residue may also be glutamine, asparagine or glutamic acid in the Picornaviridae-related viruses (not shown), implying that any polar amino acid may suffice in

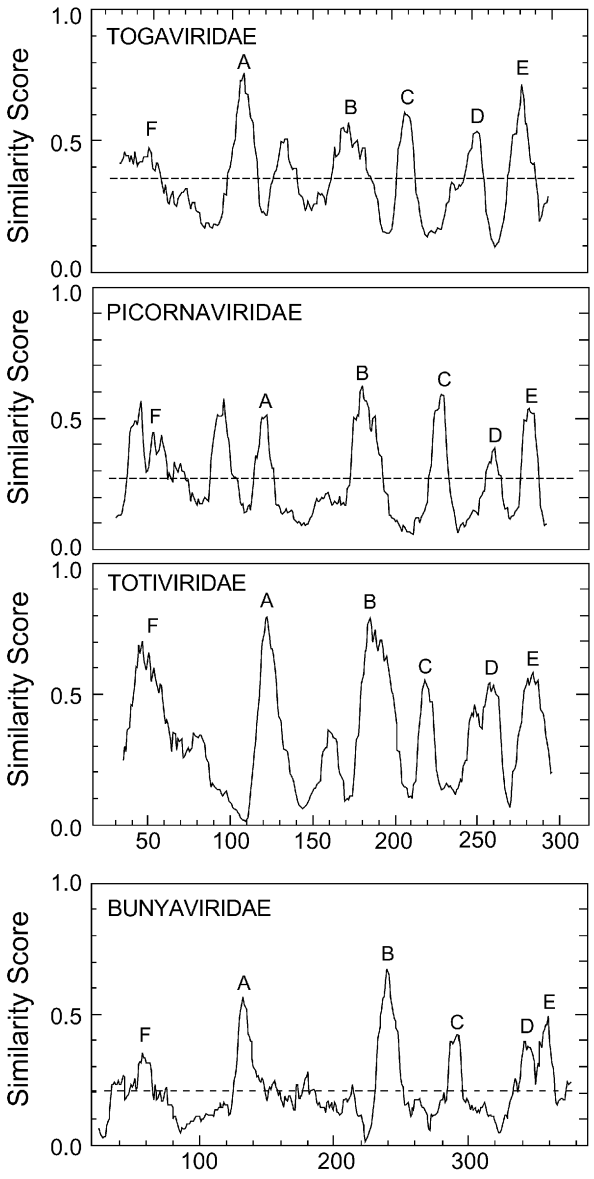


**Figure 1.** Similarity peaks in four groups of RdRps. These are PLOTSIMILARITY outputs of four PILEUP (23) alignments of the RdRps of Picornaviridae-related RdRps, Totiviridae RdRps, Togaviridae-related RdRps and Bunyaviridae-related RdRps. Picornaviridae and Togaviridae are positive strand RNA viruses; Bunyaviridae are negative strand RNA viruses; and Totiviridae are dsRNA viruses. The Picornaviridae-related RdRps include broad bean wilt virus (Comoviridae, AF144234) (U.Lee, unpublished); porcine enterovirus (Picornaviridae, AJ011380) (55); cowpea mosaic virus, (ComoviridaeX00206) (56); echovirus 23 (Picornaviridae, AJ005695) (57); grapevine fanleaf virus (Comoviridae, D00915) (58); hepatitis A virus, (Picornaviridae, M59809) (59); maize chlorotic dwarf virus (Sequiviridae, U67839) (60); poliovirus type 3, (Picornaviridae, A03900) (24); rice tungro spherical virus, (Sequiviridae, X98396) (61); sacbrood virus, (Picornaviridae, AF092924) (62); Satsuma dwarf virus, (Comoviridae, D45026) (63); and tomato ringspot virus, (Comoviridae, AF135410) (A.Wang, unpublished). The Togaviridae-related RdRps are of tobacco mosaic virus (Tobamovirus, Z29370) (64), sindbis virus (Togaviridae, M69205) (65), Sagiyama virus (Togaviridae, AB032553) (66), rape mosaic virus (Tobamovirus, U309204) (67), grapevine leafroll virus 2 (Closterovirus, AF039204) (68), grapevine leafroll virus 1 (Closterovirus, Y14131) (N.Abou-Ghanem, unpublished), Citrus tristeza virus (Closterovirus, AF260651) (69), beet yellows closterovirus (Closterovirus, AF056585) (70), and alfalfa mosaic virus (Alfalfa mosaic virus group, K2702) (71). The Totiviridae RdRps are those of ScVL1, ScVLa, TvV1, TvV2, TvV5, LrV1-1, LrV2-1, GaVl1, SsV1, SsV2, Hv190sv, EbV1, UmVP1H1, GlV and Abv1L1. This set was chosen to exclude any RdRps with more than ~50% identity with any other Totiviridae RdRp. The CcV sequence is excluded because it is only a partial sequence. The accession numbers and literature references for all of these sequences are in Table 2. The Bunyaviridae-related RdRps were from rice stripe virus (Tenuivirus, D31879) (72), Uukuniemi virus (Bunyaviridae, D10759) (73), Bunyamwera virus (Bunyaviridae, X14383) (74), Tomato spotted wilt virus (Bunyaviridae, D10066) (75), Hantaan virus (Bunyaviridae, X55901) (76) and La Crosse virus (Bunyaviridae, U12396) (77). PLOTSIMILARITY was run with a window of 10 amino acids. PILEUP was run with a similarity matrix of Blosum30 for the Totiviridae and the Bunyaviridae and Blosum62 for both the Picornaviridae- and Togaviridae-related RdRps. Gap weight was set at 5 and gap length weight at 1 for the Bunyaviridae, gap weight at 8 and gap length weight at 1 for the Totiviridae, and at the default values otherwise (gap weight at 8 and gap length weight at 2 for Blosum62 and at 15 and 5 for Blosum30). The horizontal dashed lines indicate levels of similarity that could arise by chance.

**Table 1.** Conserved motifs in the viral RdRps and HIV-1 reverse transcriptase (for comparison)

| Motif (8,12,39), (this work) | | F1 | | F2 | | F3 |
|---|---|---|---|---|---|---|
| Totiviridae consensus | | RRxF.9-30. | | KLEH | | GKT.RAIY |
| Picornaviridae consensus | | KDELLxx...KIxEx | | | | GKT.RLFS |
| Toqaviridae consensus | | KRDV.13-22 | | KxIN | 9-18 | Lxx.RIML |
| Bunyaviridae consensus | | KEVx.4-10 | | KxQR | | TGxD.REIY |
| Phi6 | 204 | RAES | 15 | KDRM | 40 | ERR.RTAM |
| polio | 160 | KDELRSKA.. | | KVEQ | | GKS.RLIV |
| HCV | 140 | KNEVFCADVS | | KGGR | | .KPARLIE |
| RHDV | 173 | KDELRPLD. | | KVKE | | GKK.RLLW |
| HIV-1 RT | 65 | | | KKDS | | TKW RKLV |

| Motif (10) | | A | | B |
|---|---|---|---|---|
| Totiviridae consensus | | DYDDFN.SQH | | SGxxx..TTFx.NS |
| Picornaviridae consensus | | DYSxFD.Gxx | | SGFxx..TVIL.NS |
| Toqaviridae consensus | | DIASFDKSQD | | SGxxx..TWIG.NS |
| Bunyaviridae consensus | | DxxKWSxQxH | | QGxxxYxSSLL.HS |
| Phi6 | 324 | DVSDHD.TFW | 393 | SGQGA..TDLM.GT |
| HCV | 220 | DTRCFD.STV | 282 | SGVLT..TSLG.NT |
| Polio | 233 | DYTGYDASLS | 288 | SGCSG..TSIF.NS |
| RHDV | 250 | DYSKWDSTMS | 308 | SGMPF..TSVI.NS |
| HIV-1 RT | 110 | DVGDAYFSVP | 151 | QGWKG..SPAIFQS |

| Motif (10) | | C | | D |
|---|---|---|---|---|
| Totiviridae consensus | | GDD | | EFL.R |
| Picornaviridae consensus | | GDD | | EFLSK |
| Toqaviridae consensus | | GDD | | YFCSK |
| Bunyaviridae consensus | | SDD | | YxxxK |
| Phi6 | 452 | SDD | 473 | EML.K |
| HCV | 317 | GDD | 341 | EAMTR |
| Polio | 327 | GDD | 345 | Q.SGK |
| RHDV | 353 | GDD | 373 | ENL R |
| HIV-1 RT | 184 | MDD | 207 | QHLLR |

| Motif (10) | | E |
|---|---|---|
| Totiviridae consensus | | GYLAR |
| Picornaviridae consensus | | xFLKR |
| Toqaviridae consensus | | KLLVK |
| Bunyaviridae consensus | | EFxSE |
| Phi6 | 495 | AFLGD |
| HCV | 362 | ELITS |
| Polio | 371 | TFLKR |
| RHDV | 396 | SFLKR |
| HIV-1 RT | 229 | WMGYE |

The first amino acid of each motif and/or region is indicated to the left of the sequence. Sequences and numbers are according to the PDB files when GenBank and PDB files differ. Distances between motifs are similar in all the RdRps, as shown in Figure 1 (with the exception of distances within motif F, see text). Similar or identical residues are in pink.

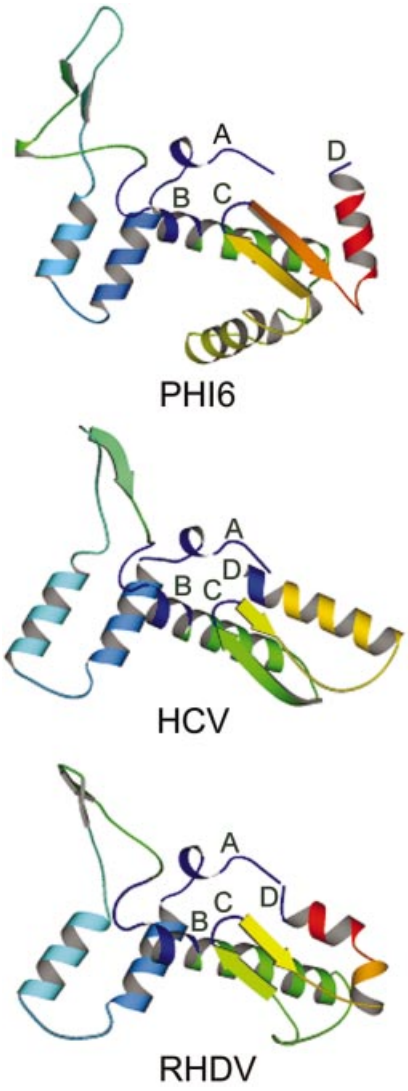


**Figure 2.** Structures of the Phi6, HCV and RHDV RdRps in the region including conserved motifs A to D. These are MOLSCRIPT(22) figures derived from the PDB files for the structures of the Phi6 (1HHS) (11), HCV (1QUV) (13) and RHDV (1KHV) (16) RdRps in the designated region. The conserved motifs are all shown in dark blue and labeled.

this position. Mutation of the terminal arginine in motif D of HCV to a lysine increases polymerase activity (32), which is not surprising, since many of the polymerases have a lysine in this position (Table 1). Motif E in HCV was approximately mapped by sequence comparison to BVDV, and a structural comparison between the poliovirus RdRp and HCV (not shown) unambiguously locates motif E ending one amino acid N-terminal to the region identified as region cc of Lai *et al.* (28). Motif E plays a role in binding the priming nucleotide (not the incoming nucleotide) in HCV (33).

In Phi6, all six conserved regions were mapped by alignment of the Phi6, 8 and 13 sequences (34–36) (Fig 3; Cystoviridae). Because the Cystoviridae and HCV-related RdRp alignments (Fig. 3) include only a few sequences, the six conserved motifs, although still prominent, are not the only visible peaks of similarity. The position of motif D in the Cystoviridae, at the C-terminal end of an alpha-helical section, in the same relative position as in poliovirus and HCV polymerases, was confirmed by structural comparison (Fig. 2). Motif E is a prominent peak in the PLOTSIMILARITY diagram generated from a PILEUP alignment of the Cystoviridae RdRps (Fig. 3), and its position is confirmed by structural comparison with poliovirus and HCV RdRps (not shown).

Some of the conserved motifs tend to fall in regions between conserved beta sheets or alpha helices, whose positions and extent are conserved even though the primary amino acid sequences within them are not. The position and sequences of the conserved motifs of Phi6, HCV and poliovirus are shown in Table 1.

In addition to the well-documented motifs A to E, there is another prominent peak of similarity (F) in all the RdRps of Figure 1. Motif F was originally mapped in the HCV RdRp solely by sequence comparison with the BVDV RdRp [motif nc of Lai *et al.* (28)], by alignment of the ChV1 RdRp with several positive strand RNA plant virus RdRps [motifs I and II of Koonin *et al.* (8)] and in the Totiviridae by multiple sequence alignment [motif 3 of Bruenn (5)]. Subsequently, the generality of motif F and its definition as an NTP-binding site was established by structural analysis (11,12). In the negative strand RNA viruses, motif F is within the 'premotif A' noted

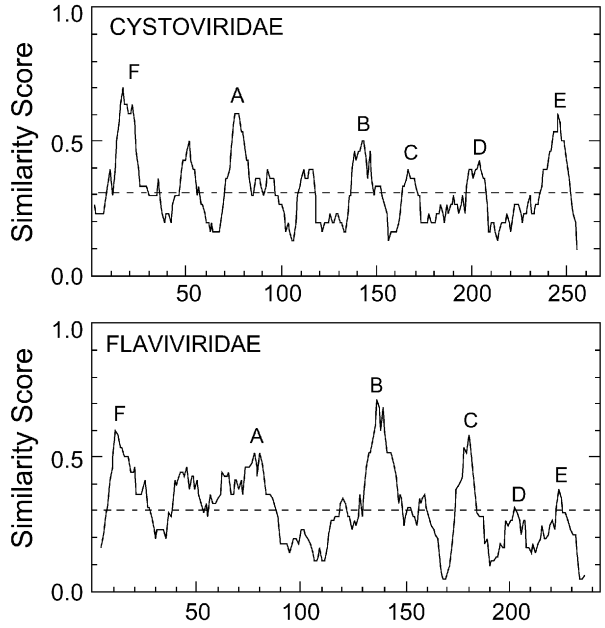


**Figure 3.** Similarity peaks in the Cystoviridae and Flaviviridae RdRps. The Cystoviridae sequences were of Phi6 (M17461) (34), Phi8 (AF226851) (35) and Phi13 (AF261668) (36); the Flaviviridae sequences were of HCV (D84264) (78), hepatitis G virus (AB021287) (30), hepatitis GB virus (U94421) (31), and bovine viral diarrhea virus (M31182) (79). PLOTSIMILARITY was run with a window of 10 amino acids. PILEUP was run with a similarity matrix of Blosum30 for both groups of viruses. Gap weight was set at 5 and gap length weight at 1 for the Flaviviridae and at the default values for the Cystoviridae.

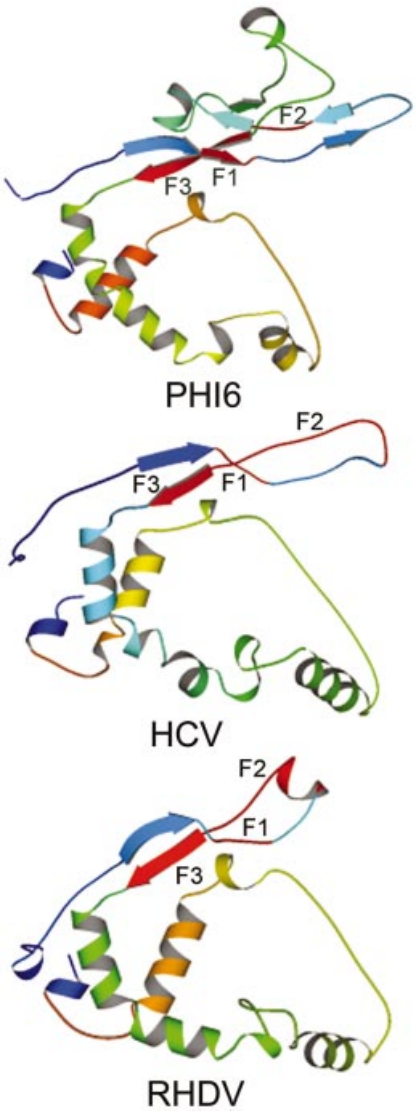


**Figure 4.** Structures of the Phi6, HCV and RHDV RdRps in the N-terminal region, including conserved motif F. As in Figure 2, but regions F1, F2 and F3 are shown in red and labeled. The residues in the Phi6 RdRp that may function as a helicase are shown (K247, D248 and R251). For comparison with Figure 2, the regions shown include conserved motif A but end just prior to conserved motif B.

as a sixth region of sequence conservation (26). This site is apparently universal (see following) and has been mapped (Table 1) by structural and/or sequence comparison in all four RdRps with known three-dimensional structures (10–13,16). A structural comparison of the three RdRps with a completely known structure in the N-terminal region is shown as Figure 4. Again, it is obvious that all three polymerases share secondary and tertiary structure throughout this region. There are three conserved motifs within region F: F1, F2 and F3, which are essentially contiguous in the Picornaviridae (Table 1). F2 and F3 are contiguous in the Totiviridae, Picornaviridae and Bunyaviridae but not in the Togaviridae. F1 and F2 are separated by only four amino acids in the Picornaviridae but usually by many more in the other groups. F2 and F3 are separated by a region of 40 amino acids in Phi6 (Table 1). This region may serve a function in strand separation (see below).

Several basic residues of motif F have been shown by co-crystallization with nucleoside triphosphates to participate in nucleotide binding in both the Phi6 and the HCV RdRps (11–13). F2 appears on one side of a long loop extending out around the mouth of the tunnel into which the template RNA enters. One side of this tunnel is partially formed by motif B, and the active site of the enzyme includes a magnesium ion bound to the conserved aspartate residues of motifs A and C (for instance, residues D324 and D453 of the Phi6 RdRp). The three sites for nucleotide binding, the I (interrogation), P (priming) and C (catalytic) sites share overlapping basic residues in this region. The I site includes K151, K155 and R158 of the HCV RdRp regions F2 and F3 (33) and K223, R225, R268 and R270 of the Phi6 RdRp regions F2 and F3 (11). This is remarkable because F2 and F3 are contiguous in the HCV RdRp (as well as in polio and RHDV RdRps and in HIV-1 RT) but separated by 40 amino acids in Phi6 (Table 1). Examination of the two structures shows that F2 and F3 in Phi6 are contiguous in space when viewed along the 'NTP tunnel', as in figure 2 of Bressanelli *et al.* (33) and as shown in Figure 4. F2 and F3 were located in HIV-1 RT by a comparison of the unliganded structure (37) with that of HCV, RHDV and Phi6. A similar comparison mapped regions A to D of HIV-1 RT (10). The sequences of the conserved motifs of HIV-1 RT are shown in Table 1. As predicted by the sequence similarity, residues K65 and R72 take part in nucleotide binding (38). HIV-1 RT has an abbreviated loop between the two beta strands and appears to be missing region F1 (not shown).

**Table 2.** Conserved region F (motif 3) in the dsRNA virus RdRp

| Virus | F1 | | F2 | F3 | pos. | Acc. No. | Reference |
|---|---|---|---|---|---|---|---|
| scvl1 | RNKF | 27 | KYEW..GKQ..RAIY | | 385-425 | M28353 | (80) |
| scvla | RSKK | 27 | KYEW..GKV..RALY | | 395-435 | AR052310 | (81) |
| tvv1 | RKEF | 21 | KLEH..GKE..RFIY | | 302-338 | U08999 | (82) |
| tvv5 | RKEF | 21 | KLEH..GKE..RFIY | | 302-338 | U57898(83) | |
| tvv2 | RLEF | 21 | KLEH..GKT..RFIY | | 328-364 | AF127178 | (84) |
| lrv1-1 | RKEG | 29 | KLEH..GKT..RLLL | | 362-405 | M92355 | (85) |
| lrv2-1 | RKEV | 29 | KLEH..GKS..RLLL | | 258-301 | U32108 | (86) |
| lrv1-4 | RKEG | 29 | KLEH..GKT..RLLL | | 366-409 | U01899 | (87) |
| qavl1 | RRAW | 22 | KLEA..GKT..RAIF | | 320-356 | AAK11656(Tuomivirta, unpub) | |
| ssv2 | RAVM | 22 | KLEH..GKT..RAIF | | 320-356 | AF039080 | (88) |
| hv190sv | RPMA | 22 | KLEN..GKD..RAIF | | 333-369 | U41345 | (89) |
| ssv1 | RRMA | 23 | KLEH..GKT..RAIF | | 337-373 | AF038665 | (88) |
| ebv1 | RRAA | 22 | KVEQ..GKG..RAIY | | 365-401 | NP_108651 | (Fraga, unpub.) |
| umvp1h1 | KRSV | 26 | KLNETGGKA..RAIY | | 400-442 | U01059 | (90) |
| abv111 | RNYD | 9 | KHEV..GKNASRSLW | | 126-151 | X94361 | (91) |
| qlv | KRSF | 30 | KPEL..TKV..RAVI | | 417-462 | L13218 | (92) |
| ccv | KRVW | 25 | KYEN..GKP..RALY | | 41-80* | X92203 | (93) |
| ahv | RSHISK....RDNL...KV..RPVY | | | | 260-275 | L39125 | (94) |
| fsv | RGHLSP....RDDP...KT..RLAW | | | | 190-205 | D55668 | (95) |
| ppv | RTQLTD....LTEK..TKI..RNVW | | | | 183-199 | AB012616 | (96) |
| zbv | RPKF 12 RQSVG.....RSIW | | | | 128-152 | AF224490 | (97) |
| cpv | RCQLR.....KRGE..NKP..RLIW | | | | 154-169 | U95995 | (98) |
| rsv | RNHISK....RDGT..LKV..RPVY | | | | 326-342 | AF133290(99) | |
| fpv1 | RNHISD....RDGI..LKQ..RPVY | | | | 395-311 | AF047013(100) | |
| chv1 | KSQAVP..GQPLLAPRMKDL.RTVV | | | | 1832-1853 | M57938(101) | |
| chv2 | KSQAVD..GTALLPPKLKDL.RTVV | | | | 1959-1977 | L29010(102) | |
| chv3 | KDEVLP....KSYVEREKI..RTII | | | | 1826-1841 | (40) | |
| ibdv | KAERYD....KSTWL.TKT..RNIW | | | | 321-338 | U30819 | (103) |
| ipnv | KGERYT....QVDWD.KKT..RNIW | | | | 305-322 | M58756 | (104) |
| reo | KIFQ 24 RNQVQ.RRP..RSIM | | | | 490-529 | M24734 | (105) |
| rota | RYTP 16 RDVP..GRRT.RIIF | | | | 432-463 | X16830 | (106) |
| btv | RVMDAADC.FRNSAD.....RNIF | | | | 573-590 | L20447 | (107) |
| Phi13 | REQM 14 KVRQ39TRR..RTAM | | | | 253-320 | AF261668(36) | |
| Phi8 | RDQLEKPG..KKRD29LRR..RTAY | | | | 225-272 | AF(35)226851 | |
| Phi6 | RAEST 15 KDRM40ERR..RTAM | | | | 204-273 | M17461 | (34) |

Similar or identical residues are colored pink. Included are sequences of Totiviridae, Partitiviridae, Hypoviridae, Birnaviridae, Reoviridae and Cystoviridae RdRps. The cryptic dsRNA viral elements of plants (108–110) are not included because they are more similar to the alfalfa mosaic virus-like group of positive strand RNA viruses (111). Similarly, the T and W dsRNAs of *Saccharomyces cerevisiae* (and related elements) are excluded because they are really replicative intermediates of positive strand RNA viruses (112).

*Only a partial sequence of the CcV RdRp is available. This is the sequence of the 'Cucurbit yellows-associated virus mRNA', which appears to code for a portion of the RdRp of a contaminating dsRNA fungal virus.

Motif F is identified for all the dsRNA viral RdRps in Table 2. Region F was mapped by choosing the appropriate peak of similarity N-terminal to region A in plots like those of Figures 1 and 3. GAp or PILEUP did alignments of the Phi6, Phi8 and Phi13 sequences; the IBDV and IPNV sequences; the ChV1, ChV2 and ChV3 sequences; the FpV1, RsV AhV, FsV, PpV, CpV and ZbV sequences; and the Totiviridae sequences. The assignment of motif F in the birnaviruses by alignment of IBDV and IPNV agrees with a multiple sequence alignment using two new, additional, viral RdRps (39). There were no pairs or groups adequate for such instructive alignments of REO, BTV and rotavirus RdRps, which were aligned by hand. The ChV1, 2 and 3 sequence alignment agrees (in this region) with previous alignments (40). These groups of aligned sequences were then compared with the Totiviridae alignment for the completion of Table 2. Regions F1 and F2 of the Cystoviridae RdRp were mapped as peaks in the PLOTSIMILARITY output and subsequently verified by structural analysis (Fig. 4).

Among the dsRNA viruses, the Totiviridae are an interesting group, in which a single viral dsRNA encodes all viral functions. A previous comparison of the RdRps from this group of viruses defined eight conserved motifs (5). This comparison also identified motifs A (4), B (5), C (6), D (7) and E (8). Of the remaining three conserved motifs, one was motif F (3). The final conserved motifs, motifs 1 and 2, which are N-terminal to motif F, may be unique to the Totiviridae. The identification of motifs A to F of the Totiviridae is shown in Figure 1. Substitution mutagenesis experiments have shown that region F in the ScVL1 polymerase is essential to RdRp activity (17,41).

The structural similarity of the RdRps extends some 100 amino acids N-terminal to motif F (Fig. 4). This region is missing in the poliovirus RdRp structure, but a comparison of the HCV, RHDV and Phi6 structures shows that F1, F2 and F3 are all close to each other in space, although F1 and F2 are separated by a minimum of four amino acids (RHDV) and F2 and F3 by a maximum of 40 amino acids (Phi6). F1 and F2 are primarily in coil structures, although a portion of F1 is the tail end of a beta strand in Phi6. A portion of F2 is in alpha helix in RHDV, but this region is coil in HCV, and only one of the RHDV monomers shows this structure in the crystals analyzed; the second monomer has no defined structure precisely where this helix exists in the other monomer (16). This region also has no defined structure in the crystals of poliovirus RdRp analyzed (10). F2 is therefore in a region that probably adopts more than one conformation. F3, which contains the conserved arginine residue participating in nucleotide binding (e.g. R270 in Phi6), is in beta strand in each of the three structures complete in this region (and in HIV-1 RT as well).

## DISCUSSION

Mapping peaks of similarity in groups of proteins provides a bioinformatic method of examining protein structure as useful as mapping RNA secondary structure by looking for conservation of base-paired regions (42). In combination with direct structural comparison, it can make significant contributions to understanding function.

Consider the combination of structural and sequence comparison of the HCV, RHDV and Phi6 polymerases that maps a 40 amino acid insertion within region F of the Phi6 RdRp. This insertion in the Phi6 polymerase is situated outside the mouth of the tunnel leading to the active site of the enzyme. This region has been described as a 'plough-like protuberance' that may function to separate the strands of the dsRNA, so that one strand can enter the template tunnel (11). There are three prominent charged residues at the tip of the 40 amino acid insert in the Phi6 polymerase (K247, D248 and R251) that may play a role in this helicase activity. Generally, the dsRNA viruses have a very AU-rich region at the 5′ end of the plus strand (43), which is where this separation takes place, as the template strand enters the polymerase.

If this region in the Phi6 RdRp is a helicase necessary for strand separation, it should be present in all the dsRNA viral RdRps (model 1). However, of the dsRNA viral RdRps, only the Cystoviridae RdRps have insertions between F2 and F3 (Table 2). Since no other dsRNA viral RdRp structure is known, this data is hard to interpret. Interpretation is also complicated by the fact that there are two possible modes of transcription in dsRNA viruses. It is possible that the semiconservative (strand displacement) mechanism of transcription in Phi6 (44) puts different requirements on its polymerase in this region. Of the dsRNA viruses whose RdRp sequences

are known, only IPNV(45) and Phi6 (44) are known to have a semi-conservative mechanism. The Reoviridae (46) and at least some of the Totiviridae (47–50) have conservative synthesis. Clearly, the conservative mode of synthesis is not a function of how many viral dsRNA segments are present. The Birnaviridae (IPNV and IBDV) RdRps do not have an insertion between F2 and F3 [Table 2; (39)]. However, since they do have a circular permutation of motifs A, B and C, it is possible that they have a region like the F2–F3 insertion of the Cystoviridae elsewhere.

In the Totiviridae, F1, F2 and F3 are part of a region in the ScV RdRp that can serve as an RNA-binding domain upon deletion of region A and the sequence between A and B (41). This is consistent with some portion of region F serving a strand separation function. One candidate for such a function is F1, a region conserved in all the RdRps examined (Tables 1 and 2). Since F1 is in the same place in all the structures determined (Fig. 4), just at the mouth of the template tunnel, it is possible that it serves a role in strand separation, ensuring that the template strand enters the tunnel. In the RdRps that function conservatively, there would be no requirement for a second strand separation region, pushing the plus strand out of the polymerase (and out of the viral particle). The original plus strand would re-anneal with the template strand after passing through the active site of the polymerase (11) and the new plus strand would exit the polymerase (and the viral particle). Consistent with this model (model 2), at least some of the RdRps of single-strand RNA viruses (i.e. poliovirus RdRp) also have an intrinsic helicase activity coupled to elongation (51). Since the substrate used in this experiment was only partially double-stranded, it is not clear that poliovirus RdRp can unwind an entirely duplex template; nor is it clear that displacement synthesis would be the result, as it was for this substrate. HIV-1 RT, which is apparently missing region F1, uses a primer–template complex, and would not have to (and does not) separate strands prior to synthesis (38). At least some reverse transcriptases do not unwind dsRNA (51).

In summary, the combination of structural comparisons and systematic primary sequence alignments has defined two regions of the viral RdRp that may be involved in unwinding dsRNA for transcription. One of these is the N-terminal portion of motif F and the second is a large insertion in motif F present in the RdRps of the cystoviridae. Defining the functional homologies of the RdRps in this region will require determination of the structures of several more polymerases from additional groups of both ssRNA and dsRNA viruses. However, testing the proposed helicase activity in the Phi6 polymerase should be possible simply by deleting the 40 amino acid insertion; the resultant polymerase should be able to synthesize dsRNA from ssRNA templates but should fail to synthesize viral plus strands from genomic dsRNAs (model 1) or should synthesize viral plus strands conservatively from dsRNA templates (model 2). Both replicase and transcriptase activities can be assayed *in vitro* using the purified recombinant enzyme (52–54).

A similar bioinformatic approach to finding conserved motifs has shown that motif C in the birnaviruses (IBDV and IPNV) is in a circularly permuted arrangement with motifs A and B; that is, in the order CAB, rather than ABC (39). This discovery was made possible by the sequencing of two new viral genomes, the predicted RdRp of one of which, TaV, showed significant similarity to IPNV. This further validates the approach outlined here: mapping conserved motifs in groups of proteins all of which are significantly, but distantly, related.

## ACKNOWLEDGEMENTS

I thank Craig Cameron for communication of results prior to publication, Bill Duax and Bruce Nicholson for reading the manuscript, and the National Science Fouccation (MCB-9727630) and the Margaret L. Wendt Foundation for support.

## REFERENCES

1. Kamer,G. and Argos,P. (1984) Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Res.*, **12**, 7269–7282.
2. Argos,P. (1988) A sequence motif in many polymerases. *Nucleic Acids Res.*, **16**, 9909–9916.
3. Delarue,M., Poch,O., Tordo,N., Moras,D. and Argos,P. (1990) An attempt to unify the structure of polymerases. *Protein Eng.*, **3**, 461–467.
4. Bruenn,J.A. (1991) Relationships among the positive strand and double-strand RNA viruses as viewed through their RNA-dependent RNA polymerases. *Nucleic Acids Res.*, **19**, 217–226.
5. Bruenn,J.A. (1993) A closely related group of RNA-dependent RNA polymerases from double-stranded RNA viruses. *Nucleic Acids Res.*, **21**, 5667–5669.
6. Ghabrial,S.A. (1998) Origin, adaptation and evolutionary pathways of fungal viruses. *Virus Genes*, **16**, 119–131.
7. Koonin,E.V., Gorbalenya,A.E. and Chumakov,K.M. (1989) Tentative identification of RNA-dependent RNA polymerases of dsRNA viruses and their relationship to positive strand RNA viral polymerases. *FEBS Lett.*, **252**, 42–46.
8. Koonin,E.V., Choi,G.H., Nuss,D.L., Shapira,R. and Carrington,J.C. (1991) Evidence for common ancestry of a chestnut blight hypovirulence-associated double-stranded RNA and a group of positive-strand RNA plant viruses. *Proc. Natl Acad. Sci. USA*, **88**, 10647–10651.
9. Poch,O., Sauvaget,I., Delarue,M. and Tordo,N. (1989) Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.*, **8**, 3867–3874.
10. Hansen,J.L., Long,A.M. and Schultz,S.C. (1997) Structure of the RNA-dependent RNA polymerase of poliovirus. *Structure*, **5**, 1109–1122.
11. Butcher,S.J., Grimes,J.M., Makeyev,E.V., Bamford,D.H. and Stuart,D.I. (2001) A mechanism for initiating RNA-dependent RNA polymerization. *Nature*, **410**, 235–240.
12. Lesburg,C.A., Cable,M.B., Ferrari,E., Hong,Z., Mannarino,A.F. and Weber,P.C. (1999) Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nature Struct. Biol.*, **6**, 937–943.
13. Ago,H., Adachi,T., Yoshida,A., Yamamoto,M., Habuka,N., Yatsunami,K. and Miyano,M. (1999) Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Struct. Fold Des.*, **7**, 1417–1426.
14. Bressanelli,S., Tomei,L., Roussel,A., Incitti,I., Vitale,R.L., Mathieu,M., De Francesco,R. and Rey,F.A. (1999) Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Proc. Natl Acad. Sci. USA*, **96**, 13034–13039.
15. Hobson,S.D., Rosenblum,E.S., Richards,O.C., Richmond,K., Kirkegaard,K. and Schultz,S.C. (2001) Oligomeric structures of poliovirus polymerase are important for function. *EMBO J.*, **20**, 1153–1163.
16. Ng,K.K., Cherney,M.M., Vazquez,A.L., Machin,A., Alonso,J.M., Parra,F. and James,M.N. (2002) Crystal structures of active and inactive conformations of a caliciviral RNA-dependent RNA polymerase. *J. Biol. Chem.*, **277**, 1381–1387.
17. Routhier,E. and Bruenn,J.A. (1998) Functions of conserved motifs in the RNA-dependent RNA polymerase of a yeast double-stranded RNA virus. *J. Virol.*, **72**, 4427–4429.
18. Devereux,J., Haeberli,P. and Smithies,O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.*, **12**, 387–395.

19. Genetics Computer Group (2002) Version 10.3 Edn. Accelrys Inc., San Diego, CA.
20. Altschul,S.F., Gish,W., Miller,W. and Myers,E.W. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
21. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
22. Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
23. Genetics Computer Group (1991) 7 Edn. GCG, Madison, WI.
24. Stanway,G., Hughes,P.J., Mountford,R.C., Reeve,P., Minor,P.D., Schild,G.C. and Almond,J.W. (1984) Comparison of the complete nucleotide sequences of the genomes of the neurovirulent poliovirus P3/Leon/37 and its attenuated Sabin vaccine derivative P3/Leon 12a1b. *Proc. Natl Acad. Sci. USA*, **81**, 1539–1543.
25. Poch,O., Blumberg,B.M., Bougueleret,L. and Tordo,N. (1990) Sequence comprison of five polymerases (L proteins) or unsegmented negative-strand viruses: theoretical assignment of functional domains. *J. Gen. Virol.*, **71**, 1153–1162.
26. Lukashevich,I.S., Djavani,M., Shapiro,K., Sanchez,A., Ravkov,E., Nichol,S.T. and Salvato,M.S. (1997) The Lassa fever virus L gene: nucleotide sequence, comparison and precipitation of a predicted 250 kDa protein with monospecific antiserum. *J. Gen. Virol.*, **78**, 547–551.
27. Colett,M.S., Larson,R., Gold,C., Strick,D., Anderson,D.K. and Purchio,A.F. (1988) Molecular cloning and nucleotide sequence of the pestivirus bovine viral diarrhea virus. *Virology*, **165**, 191–199.
28. Lai,V.C., Kao,C.C., Ferrari,E., Park,J., Uss,A.S., Wright-Minogue,J., Hong,Z. and Lau,J.Y. (1999) Mutational analysis of bovine viral diarrhea virus RNA-dependent RNA polymerase. *J. Virol.*, **73**, 10129–10136.
29. Tokita,H., Okamoto,H., Iizuka,H., Kishimoto,J., Tsuda,F., Miyakawa,Y. and Mayumi,M. (1998) The entire nucleotide sequences of three hepatitis C virus isolates in genetic groups 7–9 and comparison with those in the other eight genetic groups. *J. Gen. Virol.*, **79**, 1847–1857.
30. Naito,H., Hayashi,S. and Abe,K. (2000) The entire nucleotide sequence of two hepatitis G virus isolates belonging to a novel genotype: isolation in Myanmar and Vietnam. *J. Gen. Virol.*, **81**, 189–194.
31. Leary,T.P., Desai,S.M., Erker,J.C. and Mushahwar,I.K. (1997) The sequence and genomic organization of a GB virus A variant isolated from captive tamarins. *J. Gen. Virol.*, **78**, 2307–2313.
32. Lohmann,V., Korner,F., Herian,U. and Bartenschlager,R. (1997) Biochemical properties of hepatitis C virus NS5B RNA-dependent RNA polymerase and identification of amino acid sequence motifs essential for enzymatic activity. *J. Virol.*, **71**, 8416–8448.
33. Bressanelli,S., Tomei,L., Rey,F.A. and De Francesco,R. (2002) Structural analysis of the hepatitis C virus RNA polymerase in complex with ribonucleotides. *J. Virol.*, **76**, 3482–3492.
34. Mindich,L., Nemhauser,I., Gottlieb,P., Romantschuk,M., Carton,J., Frucht,S., Strassman,J., Bamford,D.H. and Kalkkinen,N. (1988) Nucleotide sequence of the large double-stranded RNA segment of bacteriophage phi 6: genes specifying the viral replicase and transcriptase. *J. Virol.*, **62**, 1180–1185.
35. Hoogstraten,D., Qiao,X., Sun,Y., Hu,A., Onodera,S. and Mindich,L. (2000) Characterization of phi8, a bacteriophage containing three double-stranded RNA genomic segments and distantly related to Phi6. *Virology*, **272**, 218–224.
36. Qiao,X., Qiao,J., Onodera,S. and Mindich,L. (2000) Characterization of phi 13, a bacteriophage related to phi 6 and containing three dsRNA genomic segments. *Virology*, **275**, 218–224.
37. Hsiou,Y., Ding,J., Das,K., Clark,A.D.,Jr, Hughes,S.H. and Arnold,E. (1996) Structure of unliganded HIV-1 reverse transcriptase at 2.7 Å resolution: implications of conformational changes for polymerization and inhibition mechanisms. *Structure*, **4**, 853–860.
38. Huang,H., Chopra,R., Verdine,G.L. and Harrison,S.C. (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science*, **282**, 1669–1675.
39. Gorbalenya,A.E., Pringle,F.M., Zeddam,J.-L., Luke,B.T., Cameron,C.E., Kalmakoff,J., Hanzlik,T.N., Gordon,K.H.J. and Ward,V.K. (2002) The palm subdomain-based active site is internally permuted in viral RNA-dependent RNA polymerases of an ancient lineage. *J. Mol. Biol.*, **324**, 47–62.
40. Smart,C.D., Yuan,W., Foglia,R., Nuss,D.L., Fulbright,D.W. and Hillman,B.I. (1999) Cryphonectria hypovirus 3, a virus species in the family hypoviridae with a single open reading frame. *Virology*, **265**, 66–73.
41. Ribas,J.C., Fujimura,T. and Wickner,R.B. (1994) A cryptic RNA-binding domain in the Pol region of the L-A double-stranded RNA virus Gag-Pol fusion protein. *J. Virol.*, **68**, 6014–6020.
42. Neefs,J.M. and De Wachter,R. (1990) A proposal for the secondary structure of a variable area of eukaryotic small ribosomal subunit RNA involving the existence of a pseudoknot. *Nucleic Acids Res.*, **18**, 5695–5704.
43. Bruenn,J. (2002) In Tavantzis,S.M. (ed.), *dsRNA Genetic Elements. Concepts and Applications in Agriculture, Forestry and Medicine*. CRC Press, Boca Raton, FL, pp. 109–124.
44. VanEtten,J.L., Burbank,D.E., Cuppels,D.A., Lane,L.C. and Vidaver,A.K. (1980) Semiconservative synthesis of single-stranded RNA by bacteriophage phi6 RNA polymerase. *J. Virol.*, **33**, 769–773.
45. Dobos,P. (1995) Protein-primed RNA synthesis *in vitro* by the virion-associated RNA polymerase of infectious pancreatic necrosis virus. *Virology*, **208**, 19–25.
46. Silverstein,S., Christman,J.K. and Acs,G. (1976) The reovirus replicative cycle. *Annu. Rev. Biochem.*, **45**, 375–408.
47. Fujimura,T. and Wickner,R.B. (1988) Replicase of L-A virus-like particles of *Saccharomyces cerevisiae*. *In vitro* conversion of exogenous L-A and M1 single-stranded RNAs to double-stranded form. *J. Biol. Chem.*, **263**, 454–460.
48. Nemeroff,M. and Bruenn,J. (1986) Conservative replication and transcription of yeast viral double-stranded RNAs *in vitro*. *J. Virol.*, **57**, 754–758.
49. Sclafani,R.A. and Fangman,W.L. (1984) Conservative replication of yeast double-stranded RNA by displacement of progeny single strands. *Mol. Cell. Biol.*, **4**, 1618–1626.
50. Williams,T.L. and Leibowitz,M.J. (1987) Conservative mechanism of the *in vitro* transcription of killer virus of yeast. *Virology*, **158**, 231–234.
51. Cho,M.W., Richards,O.C., Dmitrieva,T.M., Agol,V. and Ehrenfeld,E. (1993) RNA duplex unwinding activity of poliovirus RNA-dependent RNA polymerase 3Dpol. *J. Virol.*, **67**, 3010–3018.
52. Laurila,M.R., Makeyev,E.V. and Bamford,D.H. (2002) Bacteriophage phi 6 RNA-dependent RNA polymerase: molecular details of initiating nucleic acid synthesis without primer. *J. Biol. Chem.*, **277**, 17117–17124.
53. Yang,H., Makeyev,E.V. and Bamford,D.H. (2001) Comparison of polymerase subunits from double-stranded RNA bacteriophages. *J. Virol.*, **75**, 11088–11095.
54. Makeyev,E.V. and Bamford,D.H. (2000) The polymerase subunit of a dsRNA virus plays a central role in the regulation of viral RNA metabolism. *EMBO J.*, **19**, 6275–6284.
55. Doherty,M., Todd,D., McFerran,N. and Hoey,E.M. (1999) Sequence analysis of a porcine enterovirus serotype 1 isolate: relationships with other picornaviruses. *J. Gen. Virol.*, **80**, 1929–1941.
56. Lomonossoff,G.P. and Shanks,M. (1983) The nucleotide sequence of cowpea mosaic virus B RNA. *EMBO J.*, **2**, 2253–2258.
57. Ghazi,F., Hughes,P.J., Hyypia,T. and Stanway,G. (1998) Molecular analysis of human parechovirus type 2 (formerly echovirus 23). *J. Gen. Virol.*, **79**, 2641–2650.
58. Ritzenthaler,C., Viry,M., Pinck,M., Margis,R., Fuchs,M. and Pinck,L. (1991) Complete nucleotide sequence and genetic organization of grapevine fanleaf nepovirus RNA1. *J. Gen. Virol.*, **72**, 2357–2365.
59. Najarian,R., Caput,D., Gee,W., Potter,S.J., Renard,A., Merryweather,J., Van Nest,G. and Dina,D. (1985) Primary structure and gene organization of human hepatitis A virus. *Proc. Natl Acad. Sci. USA*, **82**, 2627–2631.
60. Reddick,B.B., Habera,L.F. and Law,M.D. (1997) Nucleotide sequence and taxonomy of maize chlorotic dwarf virus within the family Sequiviridae. *J. Gen. Virol.*, **78**, 1165–1174.
61. Thole,V. and Hull,R. (1998) Rice tungro spherical virus polyprotein processing: identification of a virus-encoded protease and mutational analysis of putative cleavage sites. *Virology*, **247**, 106–114.
62. Ghosh,R.C., Ball,B.V., Willcocks,M.M. and Carter,M.J. (1999) The nucleotide sequence of sacbrood virus of the honey bee: an insect picorna-like virus. *J. Gen. Virol.*, **80**, 1541–1549.

63. Iwanami,T., Yamao,F., Seno,T. and Ieki,H. (1996) Nucleotide sequence of the 3′-terminal region of RNA1 of satsuma dwarf virus. *Ann. Phytopathol. Soc. Jpn.*, **62**, 4–10.
64. Dorokhov, Yu,L., Ivanov,P.A., Novikov,V.K., Agranovsky,A.A., Morozov,S., Efimov,V.A., Casper,R. and Atabekov,J.G. (1994) Complete nucleotide sequence and genome organization of a tobamovirus infecting cruciferae plants. *FEBS Lett.*, **350**, 5–8.
65. Shirako,Y., Niklasson,B., Dalrymple,J.M., Strauss,E.G. and Strauss,J.H. (1991) Structure of the Ockelbo virus genome and its relationship to other Sindbis viruses. *Virology*, **182**, 753–764.
66. Shirako,Y. and Yamaguchi,Y. (2000) Genome structure of Sagiyama virus and its relatedness to other alphaviruses. *J. Gen. Virol.*, **81**, 1353–1360.
67. Aguilar,I., Sanchez,F., Martin Martin,A., Martinez-Herrera,D. and Ponz,F. (1996) Nucleotide sequence of Chinese rape mosaic virus (oilseed rape mosaic virus), a crucifer tobamovirus infectious on *Arabidopsis thaliana*. *Plant Mol. Biol.*, **30**, 191–197.
68. Zhu,H.Y., Ling,K.S., Goszczynski,D.E., McFerson,J.R. and Gonsalves,D. (1998) Nucleotide sequence and genome organization of grapevine leafroll-associated virus-2 are similar to beet yellows virus, the closterovirus type member. *J. Gen. Virol.*, **79**, 1289–1298.
69. Albiach-Marti,M.R., Mawassi,M., Gowda,S., Satyanarayana,T., Hilf,M.E., Shanker,S., Almira,E.C., Vives,M.C., Lopez,C., Guerri,J., Flores,R., Moreno,P., Garnsey,S.M. and Dawson,W.O. (2000) Sequences of *Citrus tristeza* virus separated in time and space are essentially identical. *J. Virol.*, **74**, 6856–6865.
70. Peremyslov,V.V., Hagiwara,Y. and Dolja,V.V. (1998) Genes required for replication of the 15.5-kilobase RNA genome of a plant closterovirus. *J. Virol.*, **72**, 5870–5876.
71. Cornelissen,B.J., Brederode,F.T., Veeneman,G.H., van Boom,J.H. and Bol,J.F. (1983) Complete nucleotide sequence of alfalfa mosaic virus RNA 2. *Nucleic Acids Res.*, **11**, 3019–3025.
72. Toriyama,S., Takahashi,M., Sano,Y., Shimizu,T. and Ishihama,A. (1994) Nucleotide sequence of RNA 1, the largest genomic segment of rice stripe virus, the prototype of the tenuiviruses. *J. Gen. Virol.*, **75**, 3569–3579.
73. Elliott,R.M., Dunn,E., Simons,J.F. and Pettersson,R.F. (1992) Nucleotide sequence and coding strategy of the Uukuniemi virus L RNA segment. *J. Gen. Virol.*, **73**, 1745–1752.
74. Elliott,R.M. (1989) Nucleotide sequence analysis of the large (L) genomic RNA segment of Bunyamwera virus, the prototype of the family Bunyaviridae. *Virology*, **173**, 426–436.
75. de Haan,P., Kormelink,R., de Oliveira Resende,R., van Poelwijk,F., Peters,D. and Goldbach,R. (1991) Tomato spotted wilt virus L RNA encodes a putative RNA polymerase. *J. Gen. Virol.*, **72**, 2207–2216.
76. Schmaljohn,C.S. (1990) Nucleotide sequence of the L genome segment of Hantaan virus. *Nucleic Acids Res.*, **18**, 6728.
77. Roberts,A., Rossier,C., Kolakofsky,D., Nathanson,N. and Gonzalez-Scarano,F. (1995) Completion of the La Crosse virus genome sequence and genetic comparisons of the L proteins of the Bunyaviridae. *Virology*, **206**, 742–745.
78. Tokita,H., Okamoto,H., Iizuka,H., Kishimoto,J., Tsuda,F., Miyakawa,Y. and Mayumi,M. (1998) The entire nucleotide sequences of three hepatitis C virus isolates. *J. Gen. Virol.*, **79**, 1847–1857.
79. Collett,M.S., Larson,R., Gold,C., Strick,D., Anderson,D.K. and Purchio,A.F. (1988) Molecular cloning and nucleotide sequence of the pestivirus bovine viral diarrhea virus. *Virology*, **165**, 191–199.
80. Diamond,M.E., Dowhanick,J.J., Nemeroff,M.E., Pietras,D.F., Tu,C.-L. and Bruenn,J.A. (1989) Overlapping genes in a yeast dsRNA virus. *J. Virol.*, **63**, 3983–3990.
81. Park,C.-M., Lopinski,J., Masuda,J., Tzeng,T.-H. and Bruenn,J.A. (1996) A second double-stranded RNA virus from yeast. *Virology*, **216**, 451–454.
82. Tai,J.H. and Ip,C.F. (1995) The cDNA sequence of *Trichomonas vaginalis* virus-T1 double stranded RNA. *Virology*, **206**, 773–776.
83. Su,H.M. and Tai,J.H. (1996) Genomic organization and sequence conservation in type I *Trichomonas vaginalis* viruses. *Virology*, **222**, 470–473.
84. Bessarab,I.N., Liu,H.W., Ip,C.F. and Tai,J.H. (2000) The complete cDNA sequence of a type II *Trichomonas vaginalis* virus. *Virology*, **267**, 350–359.
85. Stuart,K.D., Weeks,R., Guilbride,L. and Myler,P.J. (1992) Molecular organization of *Leishmania* RNA virus 1. *Proc. Natl Acad. Sci. USA*, **89**, 8596–8600.
86. Scheffter,S.M., Ro,Y.T., Chung,I.K. and Patterson,J.L. (1995) The complete sequence of Leishmania RNA virus LRV2-1, a virus of an Old World parasite strain. *Virology*, **212**, 84–90.
87. Scheffter,S., Widmer,G. and Patterson,J.L. (1994) Complete sequence of Leishmania RNA virus 1–4 and identification of conserved sequences. *Virology*, **199**, 479–483.
88. Preisig,O., Wingfield,B.D. and Wingfield,M.J. (1998) Coinfection of a fungal pathogen by two distinct double-stranded RNA viruses. *Virology*, **252**, 399–406.
89. Huang,S. and Ghabrial,S.A. (1996) Organization and expression of the double-stranded RNA genome of *Helminthosporium victoriae* 190S virus, a totivirus infecting a plant pathogenic filamentous fungus. *Proc. Natl Acad. Sci. USA*, **93**, 12541–12546.
90. Kang,J., Wu,J., Bruenn,J.A. and Park,C. (2001) The H1 double-stranded RNA genome of *Ustilago maydis* virus-H1 encodes a polyprotein that contains structural motifs for capsid polypeptide, papain-like protease and RNA-dependent RNA polymerase. *Virus Res.*, **76**, 183–189.
91. Vander Lende,T.R., Duitman,E.H., Gunnewijk,M.G., Yu,L. and Wessels,J.G. (1996) Functional analysis of dsRNAs (L1, L3, L5 and M2) associated with isometric 34-nm virions of *Agaricus bisporus* (white button mushroom). *Virology*, **217**, 88–96.
92. Wang,A.L., Yang,H.-M., Shen,K.A. and Wang,C.C. (1993) Giardiavirus double-stranded RNA genome encodes a capsid polypeptide and a gag-pol-like fusion protein by a translational frameshift. *Proc. Natl Acad. Sci. USA*, **90**, 8595–8599.
93. Coffin,R.S. and Coutts,R.H.A. (1995) Relationships among *Trialeurodes vaporariorum*-transmitted yellowing viruses from Europe and North America. *J. Phytopathol.*, **143**, 375–380.
94. Oh,C.S. and Hillman,B.I. (1995) Genome organization of a partitivirus from the filamentous ascomycete *Atkinsonella hypoxylon*. *J. Gen. Virol.*, **76**, 1461–1470.
95. Nogawa,M., Kageyama,T., Nakatani,A., Taguchi,G., Shimosaka,M. and Okazaki,M. (1996) Cloning and characterization of mycovirus double-stranded RNA from the plant pathogenic fungus, *Fusarium solani* f. sp. robiniae. *Biosci. Biotechnol. Biochem.*, **60**, 784–788.
96. Osaki,H., Kudo,A. and Ohtsu,Y. (1998) Nucleotide sequence of seed- and pollen-transmitted double-stranded *RNA*, which encodes a putative RNA-dependent RNA polymerase, detected from Japanese pear. *Biosci. Biotechnol. Biochem.*, **62**, 2101–2106.
97. Schmitt,M.J., Poravou,O., Trenz,K. and Rehfeldt,K. (1997) Unique double-stranded RNAs responsible for the anti-Candida activity of the yeast *Hanseniaspora uvarum*. *J. Virol.*, **71**, 8852–8855.
98. Khramtsov,N.V., Woods,K.M., Nesterenko,M.V., Dykstra,C.C. and Upton,S.J. (1997) Virus-like, double-stranded RNAs in the parasitic protozoan *Cryptosporidium parvum*. *Mol. Microbiol.*, **26**, 289–300.
99. Strauss,E.E., Lakshman,D.K. and Tavantzis,S.M. (2000) Molecular characterization of the genome of a partitivirus from the basidiomycete *Rhizoctonia solani*. *J. Gen. Virol.*, **81**, 549–555.
100. Compel,P., Papp,I., Bibo,M., Fekete,C. and Hornok,L. (1999) Genetic interrelationships and genome organization of double-stranded RNA elements of *Fusarium poae*. *Virus Genes*, **18**, 49–56.
101. Shapira,R., Choi,G.H. and Nuss,D.L. (1991) Virus-like genetic organization and expression strategy for a double-stranded RNA genetic element associated with biological control of chestnut blight. *EMBO J.*, **10**, 731–739.
102. Hillman,B.I., Halpern,B.T. and Brown,M.P. (1994) A viral dsRNA element of the chestnut blight fungus with a distinct genetic organization. *Virology*, **201**, 241–250.
103. Kibenge,F.S., Nagarajan,M.M. and Qian,B. (1996) Determination of the 5′ and 3′ terminal noncoding sequences of the bi-segmented genome of the avibirnavirus infectious bursal disease virus. *Arch. Virol.*, **141**, 1133–1141.
104. Duncan,R., Mason,C.L., Nagy,E., Leong,J.A. and Dobos,P. (1991) Sequence analysis of infectious pancreatic necrosis virus genome segment-B and its encoded Vp1 protein—a putative RNA-dependent RNA polymerase lacking the Gly Asp Asp motif. *Virology*, **181**, 541–552.
105. Wiener,J.R. and Joklik,W.K. (1989) The sequences of the reovirus serotype 1, 2 and 3 L1 genome segments and analysis of the mode of divergence of the reovirus serotypes. *Virology*, **169**, 194–203.
106. Mitchell,D.B. and Both,G.W. (1990) Completion of the genomic sequence of the simian rotavirus SA11: nucleotide sequences of segments 1, 2 and 3. *Virology*, **177**, 324–331.

107. Huang,I.J., Hwang,G.Y., Yang,Y.Y., Hayama,E. and Li,J.K. (1995) Sequence analyses and antigenic epitope mapping of the putative RNA-directed RNA polymerase of five U.S. bluetongue viruses. *Virology*, **214**, 280–288.
108. Pfeiffer,P. (1998) Nucleotide sequence, genetic organization and expression strategy of the double-stranded RNA associated with the '447' cytoplasmic male sterility trait in *Vicia faba*. *J. Gen. Virol.*, **79**, 2349–2358.
109. Moriyama,H., Horiuchi,H., Koga,R. and Fukuhara,T. (1999) Molecular characterization of two endogenous double-stranded RNAs in rice and their inheritance by interspecific hybrids. *J. Biol. Chem.*, **274**, 6882–6888.
110. Moriyama,H., Nitta,T. and Fukuhara,T. (1995) Double-stranded RNA in rice: a novel RNA replicon in plants. *Mol. Gen. Genet.*, **248**, 364–369.
111. Gibbs,M.J., Koga,R., Moriyama,H., Pfeiffer,P. and Fukuhara,T. (2000) Phylogenetic analysis of some large double-stranded RNA replicons from plants suggests they evolved from a defective single-stranded RNA virus. *J. Gen. Virol.*, **81**, 227–233.
112. Nieves Rodríguez-Cousiño,N., Solórzano,A., Fujimura,T. and Esteban,R. (1998) Yeast positive-stranded virus-like RNA replicons 20 S AND 23 S RNA terminal nucleotide sequences and 3′ end secondary structures resemble those of RNA coliphages. *J. Biol. Chem.*, **273**, 20363–20371.