Exceptionally diverse morphotypes and genomes of crenarchaeal hyperthermophilic viruses

Prangishvili, D; Garrett, R A

Published in:
Biochemical Society Transactions

Publication date:
2004

Document version
Publisher's PDF, also known as Version of record

Exceptionally diverse morphotypes and genomes of crenarchaeal hyperthermophilic viruses

D. Prangishvili*[1] and R.A. Garrett†

*Department of Microbiology, Archaea Centre, University of Regensburg, Universitätsstrasse 31, D-93053 Regensburg, Germany, and

†Danish Archaea Centre, Institute of Molecular Biology, Copenhagen University, Sølvgade 83H, DK-1307 Copenhagen K, Denmark

Abstract

The remarkable diversity of the morphologies of viruses found in terrestrial hydrothermal environments with temperatures >80°C is unprecedented for aquatic ecosystems. The best-studied viruses from these habitats have been assigned to novel viral families: *Fuselloviridae*, *Lipothrixviridae* and *Rudiviridae.* They all have double-stranded DNA genomes and infect hyperthermophilic crenarchaea of the orders Sulfolobales and Thermoproteales. Representatives of the different viral families share a few homologous ORFs (open reading frames). However, about 90% of all ORFs in the seven sequenced genomes show no significant matches to sequences in public databases. This suggests that these hyperthermophilic viruses have exceptional biochemical solutions for biological functions. Specific features of genome organization, as well as strategies for DNA replication, suggest that phylogenetic relationships exist between crenarchaeal rudiviruses and the large eukaryal DNA viruses: poxviruses, the African swine fever virus and *Chlorella* viruses. Sequence patterns at the ends of the linear genome of the lipothrixvirus AFV1 are reminiscent of the telomeric ends of linear eukaryal chromosomes and suggest that a primitive telomeric mechanism operates in this virus.

Viral diversity in hot terrestrial aquatic ecosystems

Microbial viruses are extremely abundant on our planet and they have important and diverse roles in ecosystems and bio-geochemical processes [1]. Nevertheless, our knowledge of their degree of diversity is very limited and mainly restricted to morphological data. Recently, these morphological data were summarized for known viruses from mesophilic and moderately thermophilic bacteria and archaea, either cultivated or observed in aqueous samples [2]. The data reveal that 97% of these viruses are typical head-and-tail phages, i.e. particles with icosahedral heads and helical tails, belonging to the families *Myoviridae*, *Siphoviridae* and *Podoviridae*, while only about 3% are tail-less icosahedra, filaments or pleomorphic particles.

Recently, geothermally heated environments including hot springs, mud holes and deep-sea hydrothermal vents, all above 80°C, have been systematically screened for viruses. The results reveal a morphological diversity of virus-like particles greatly exceeding that observed in aquatic systems at lower temperatures [3–6]. Moreover, investigations of virus–host systems from such environments led to the isolation of a wide variety of viruses which infect hyperthermophilic archaea of the orders Sulfolobales and Thermoproteales and most of this pioneering work was done in the laboratory of Wolfram Zillig. Exceptional morphological and genomic features of these viruses necessitated their classification into new viral families which include: filamentous *Lipothrixvi-*

ridae, comprising viruses TTV1, TTV2, TTV3, TTV4 [7], SIFV [8] (Figure 1A) and AFV1 [9] (Figure 1B), rod-shaped *Rudiviridae*, including SIRV1 and SIRV2 [10] (Figure 1C), and spindle-shaped *Fuselloviridae*, comprising SSV1 [11,12] and SSV2 [13,14] (Figure 1D). One more family, *Guttaviridae*, has been proposed, but not yet recognized, for the droplet-shaped virus SNDV [15] (Figure 1E). All viruses exhibit double-stranded DNA genomes, some of which are modified. The genomes are linear in lipothrixviruses and rudiviruses, and circular in the other two families.

Electron microscopy and biochemical studies have revealed major differences in structures of viruses. With the exception of the rudiviruses they all have envelopes containing virus-encoded proteins [9,16,17]. In contrast, the rudiviruses lack an envelope and the virions consist of a superhelical DNA complexed with multimers of a single protein [10].

Virus–host interactions

So far, the only hyperthermophilic virus shown to exhibit lytic properties is TTV1. All of the others are present in their hosts in a stable carrier state. This preference for a stable relationship with the host cell may reflect the necessity to minimize, or avoid, direct exposure to the extreme and unstable environmental conditions. For the rudiviruses, the existence of a carrier state correlates with their relatively simple pattern of transcription of viral genes in contrast to the complex patterns observed in many bacterial phages (A. Kessler and D. Prangishvili, unpublished work).

The fusellovirus genomes exist intracellularly in a plasmid form and they can integrate specifically, and reversibly, into

Figure 1 | **Electron micrographs of representatives of different families of viruses of hyperthermophilic Archaea, negatively stained with uranyl acetate**

(**A**) Lipothrixvirus SIFV of *Sulfolobus* (reprinted from Virology, vol. 267, H.P. Arnold, W. Zillig, V. Ziese et al., "A novel lipothrixvirus, SIFV, of the extremely thermophilic crenarchaeon *Sulfolobus*", pp. 252–266, © 2000, with permission from Elsevier), (**B**) Lipothrixvirus AFV1 of *Acidianus* (reprinted from Virology, vol. 315, M. Bettstetter, X. Peng, R.A. Garrett and D. Prangishvili, "AFV1, a novel virus infecting hyperthermophilic archaea of the genus *Acidianus*", pp. 68–79, © 2003, with permission from Elsevier), (**C**) Rudivirus SIRV1 of *Sulfolobus*, (**D**) Fusellovirus SSV1 of *Sulfolobus* and (**E**) Guttavirus SNDV of *Sulfolobus* (Reprinted from Trends in Microbiology, vol. 9, D. Prangishvili, K. Stedman and W. Zillig, "Viruses of the extremely thermophilic archaeon *Sulfolobus*", pp. 39–43, © 2001, with permission from Elsevier). Scale bars, 200 nm.
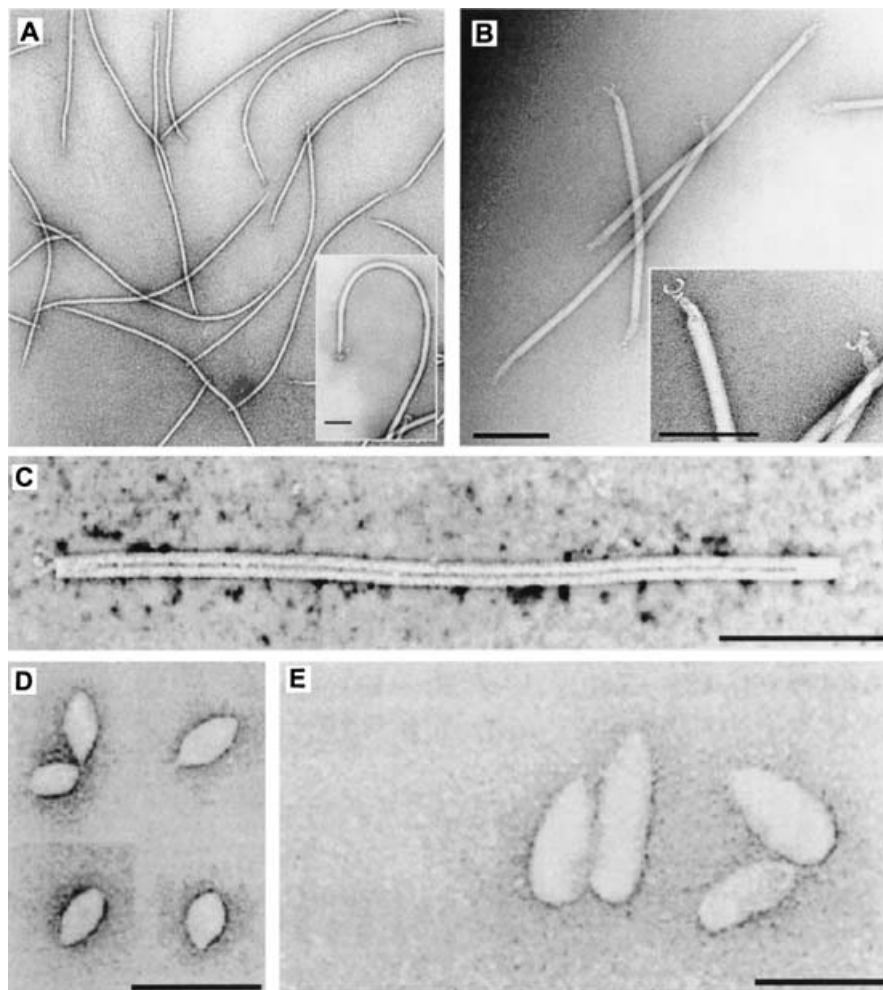
host chromosomes by means of a virus-encoded integrase. Integration occurs within a specific tRNA gene as for some temperate bacteriophages. However, in contrast to the bacterial integration mechanism which produces an intact integrase gene, integration of the archaeal viruses produces a partitioned integrase gene [18,19]. This archaeal-specific integration mechanism can lead to chromosome capture of the viral genome if the host cell is cured of the free virus, and thereby loses the capacity to express a functional integrase. Examples of this gene-capture phenomenon have been described for the genome of *Sulfolobus solfataricus* P2 [19,20]. Production of one fusellovirus, SSV1, but not SSV2, can be strongly induced by UV irradiation or mytomycin treatment, presumably as a result of a host cell SOS response,

and this does not result in cell lysis [14,16]. In contrast, the linear viral genomes of hyperthermophilic viruses do not encode integrases [8,9,21] and there is no evidence from whole-genome analyses for their integrating into *Sulfolobus* chromosomes [22]. This correlates with the observation that these viruses with linear genomes are not induced by environmental factors.

The genetic factors determining virus–host compatibility are not understood. However, some insight was gained from recent studies of the two rudiviruses SIRV1 and SIRV2. Both are stable in their natural host, but when transformed into foreign *Sulfolobus* hosts, the genome of SIRV2 remains stable whereas that of SIRV1 undergoes very rapid mutation estimated at about 10^{-3} substitutions/nucleotide

per replication cycle [10]. This corresponds to the mutational rate of the most rapidly evolving RNA viruses [23] but is unprecedented for DNA viruses. In order to investigate this phenomenon, and to obtain a comprehensive picture of the genome changes, about a dozen of the SIRV1 variant genomes were examined in some detail. The results demonstrate the presence of mutational hotspots in the genome concentrated within three or four short regions whereas the remainder of the genome is either stable or undergoes mutation at a low rate (X. Peng, H. Phan, A. Kessler, M. Häring, R.A. Garrett and D. Prangishvili, unpublished work). This study also provided strong evidence for the existence of quasi-species of the SIRV1 virus. Our working hypothesis is that the genes in these few hypervariable regions are crucial for the virus–host relationships.

Organization of viral genomes

The genomes of several viruses of Sulfolobales, including the rudiviruses SIRV1 and SIRV2 [21], the lipothrixviruses SIFV [8] and AFV1 [9], and the fuselloviruses SSV1 [24] and SSV2 [14], have been sequenced. A partial nucleotide sequence is also available for the lipothrixvirus TTV1 of *Thermoproteus* [25].

The genomes of the two fuselloviruses, one of which, SSV1, was isolated from a hot spring in Japan, and the other, SSV2, from a hot spring in Iceland, are closely related showing 55% nucleotide sequence identity. Large sections of the two genomes are clearly homologous although the level of sequence identity of the homologous ORFs (open reading frames) varies over the range 16–76% [14]. Three ORFs, VP1, VP2 and VP3, correspond to coat proteins in SSV1 but only VP1 and VP3 are present in SSV2. Only two ORFs show matches with other archaeal viral genomes. ORF88a, which is exclusive to SSV2, is homologous to ORF103b that is present in SIRV2 but not SIRV1. In addition, ORFa45 of SSV1 is homologous to ORF59a of the lipothrixvirus AFV1 that has been annotated as a member of the CopG protein family. The integrase encoded in SSV1 has been expressed heterologously [18].

The genomes of rudiviruses SIRV1 and SIRV2, both isolated from hot springs in Iceland, are even more closely related. The genomes consist of blocks with well-conserved sequences separated by non-conserved or less-conserved sequences. Sequence comparisons revealed that recombination, gene duplication, horizontal gene transfer and substitution of viral genes by homologous host genes have contributed to their genome evolution.

A genome map of the lipothrixvirus AFV1, isolated from a North American hot spring, is depicted in Figure 2(A), in which ORFs that are homologous to other hyper-thermophilic archaeal viruses are colour coded. At least seven homologues are shared with the rudiviruses. Two of these are putative glycosyl transferases which can transfer nucleotide-linked sugars to substrates such as glycogen and lipopolysaccharides and another is a putative DNA helicase. Functions have been assigned unequivocally to the genes

encoding the structural proteins of AFV1 and the rudiviruses. Moreover, genes encoding a Holliday junction resolvase and a dUTPase are present in each rudivirus, and they have been expressed heterologously [26,27]. However, most of the viral ORFs, from all three families, have still not been assigned functions and we still have little insight into protein components that contribute to viral–host interactions and virus replication.

Relationships with other viruses

The very low level of significant sequence similarity between viruses of hyperthermophilic Archaea and other known viruses correlates with the hypothesis that the rapid evolution of viral genes precludes the detection of relationships over large evolutionary distances [28]. However, for viruses with linear genomes a comparison of the DNA ends and the replication mechanism could provide a basis for following phylogenetic lineages. Initiating DNA synthesis from the termini of linear DNA strands is necessarily complex and different mechanisms have evolved to deal with this, including covalently closed ends, cohesive ends, telomeric ends or ends with covalently attached proteins [29].

The termini of linear viral genomes are difficult to sequence. Terminal fragments are not ligated into vectors if they are modified and, therefore, they are not represented in genomic libraries. Moreover, given the low production of the hyperthermophilic viruses, it is difficult to obtain sufficient amounts of virus to perform primer-induced sequencing on the ends of the viral DNA. Nevertheless, the termini of two viral genomes have been sequenced and analysed, one from the lipothrixvirus AFV1 [9] and another from the rudivirus SIRV1 [30].

The AFV1 termini of the A + T-rich genome exhibit an unusual 11 nucleotide G-C inverted repeat. This possibly constitutes a kind of clamp that ensures the stability of the DNA termini at very high temperatures. It could also represent a transposase recognition site. In addition, the 300 bp at each end of the genome contains clusters of multiple short direct repeats of the pentanucleotide TTGTT (and its complementary sequence), and close variants thereof (Figure 2B). Such an organization is reminiscent of the telomeric ends of linear eukaryal chromosomes where the telomerase produces multiple short and imperfect repeat sequences, when generating a 3′-overhang on the lagging strand, in order to prevent shortening of the linear DNA during each round of replication. Possibly, a less complex telomeric mechanism operates on the linear genomes of some archaeal viruses.

The terminal structures of the rudiviral genomes are different. Both SIRV1 and SIRV2 exhibit inverted terminal repeats of about 2 kb, albeit with a large insert in one repeat sequence in SIRV2 [21,30]. Moreover, these large repeat structures each exhibit regularly spaced perfect, and imperfect, direct repeat structures. Chemical sequencing of the termini of the SIRV1 genome demonstrated that the two DNA strands are complementary, and covalently linked,

Figure 2 | Genome organization of the lipothrixvirus AFV1

(**A**) Genome map of lipothrixvirus AFV1 showing the size and direction of transcripts of the putative genes present on the two DNA strands. Homologues that are shared with other *Sulfolobus* viruses, SIFV, SIRV1, SIRV2 and SSV viruses, are colour coded. (**B**) Terminal sequences of one strand of AFV1. The terminal 350 nucleotides contain numerous short direct repeats, commonly TTGTT and close variants thereof at one end and the complementary sequence AACAA at the other. The variants of the pentanucleotide repeats are colour coded. An 11 bp G-C inverted repeat exists at the extremities. The terminal regions exhibit larger imperfect inverted repeats marked by arrows. Direct 37 bp repeats are underlined without arrows at the left-hand end. Reprinted from Trends in Microbiology, vol. 9, D. Prangishvili, K. Stedman and W. Zillig, "Viruses of the extremely thermophilic archaeon *Sulfolobus*", pp. 39–43, © 2001, with permission from Elsevier.
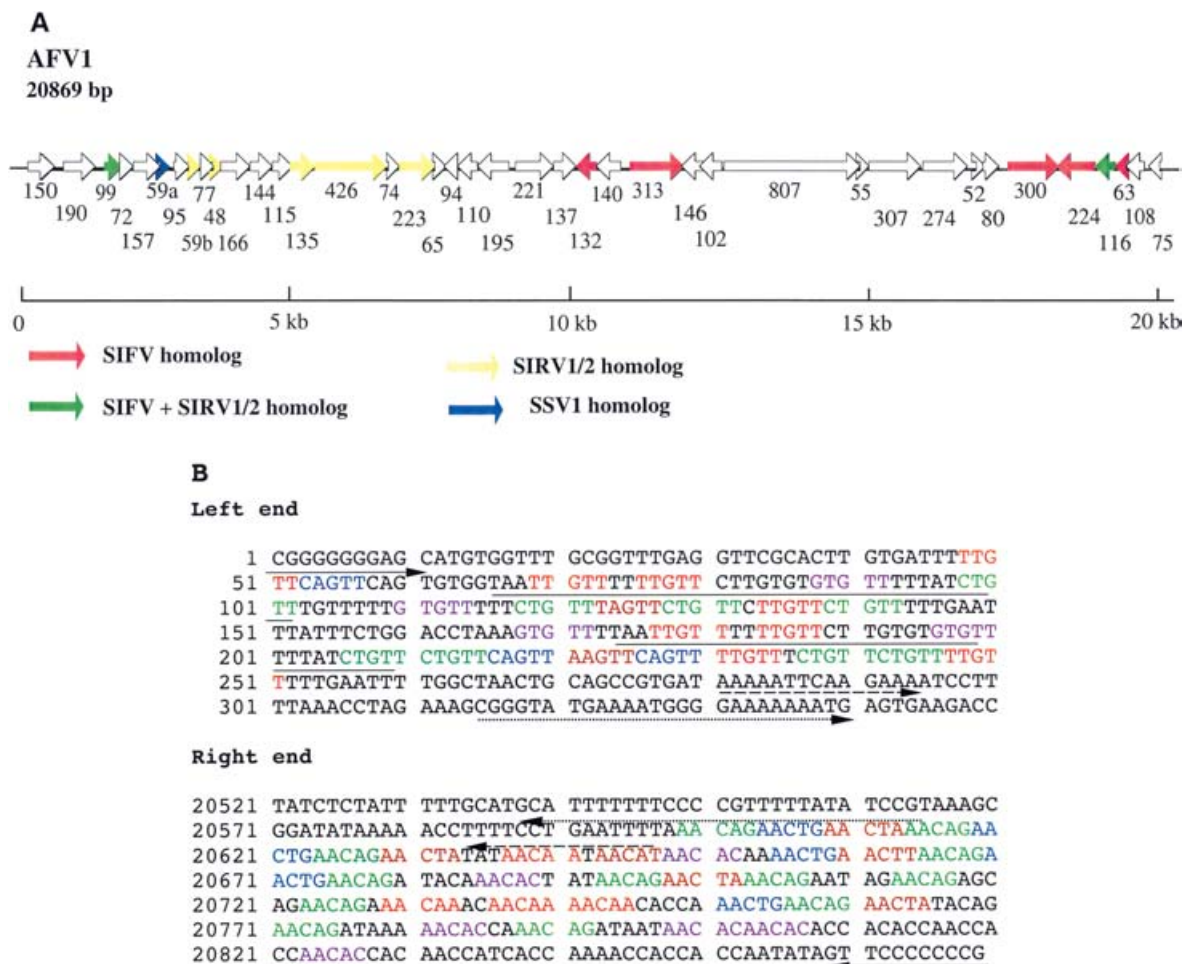
producing a continuous polynucleotide chain [30]. Although the nature of the chemical linkage is unknown, it was possible to sequence through one end of SIRV2, using the *Taq* DNA polymerase, which suggests that it is a 3′–5′ linkage (X. Peng, unpublished work). If so, then stereochemical constraints require that despite the complementarity of the bases at the termini, a hairpin structure with a loop region of at least four nucleotides is present.

The structural features of the rudiviral genomic termini are shared by genomes of a family of large double-stranded DNA eukaryal viruses including the poxviruses, *Chlorella* viruses and the African swine fever virus. The similarities extend further to the mechanism of DNA replication. A self-priming model for replication of these eukaryal viruses postulates that

replication is initiated and primed by generating a free 3′-OH near each terminus with the subsequent formation of specific head-to-head and tail-to-tail linked replicative intermediates [31]. For SIRV1, a small proportion of DNA molecules was shown to be nicked 11 nucleotides from the termini [30]. Moreover, similar replicative intermediates, linked head-to-head and tail-to-tail, were detected in *Sulfolobus* cells infected with SIRV1 and SIRV2 [21]. This linkage between two DNA molecules with inverted terminal repeats generates a cruciform structure which, for the eukaryal viruses, is resolved by a viral-encoded Holliday junction resolvase. This enzyme is also encoded in the genomes of archaeal rudiviruses SIRV1 and SIRV2 [27]. Moreover, the probable recognition sequences for the archaeal and eukaryal resolvases are closely

similar, each exhibiting an A_7 tract, and they are located close to the ends of the genome [21].

Thus rudiviruses appear to be similar to eukaryal double-stranded DNA viruses not only in the structural properties of their genomic termini but also in their mechanisms of initiation of DNA replication and their mode of resolving replication intermediates. Since it is improbable that all these similarities resulted from convergent evolution, a common origin of the replication machineries is likely. The results also suggest that the genomes of archaeal rudiviruses and the eukaryal DNA viruses share a common lineage. The latter supposition is supported by the detection of 14 homologous genes shared by the rudiviruses and the eukaryal viruses, and by the location of the most conserved genes in the centres of their genomes [21].

The results are in line with the emerging picture of viral relationships in the three domains of life [32] although, to date, they provide the only substantial evidence for a phylogenetic relationship between archaeal and eukaryal viruses. The results also correlate with the hypothesis of viral eukaryogenesis, which proposes that the eukaryal nucleus evolved from an archaeal DNA virus and that some specific features of eukaryotic cell including linear chromosomes, telomeres, mRNA capping and α-polymerase have a viral ancestry [33,34]. Thus although the evolutionary implications of the exceptional diversity of the viruses of hyperthermophilic archaea remain obscure, we feel strongly that a better understanding of these will yield unexpected insights into an understanding of the origin and evolution of viruses and of early life in general.

References

1 Wommack, K.E. and Colwell, R.R. (2000) Micr. Mol. Biol. Rev. **64**, 9–114
2 Ackermann, H.-W. (2001) Arch. Virol. **146**, 843–857
3 Zillig, W., Kletzin, A., Schleper, C., Holz, I., Janekovic, D., Hain, H., Lanzendörfer, M. and Kristjansson, J.K. (1994) System. Appl. Microbiol. **16**, 609–628
4 Rice, G., Stedman, K.M., Snyder, J., Wiedenheft, B., Brumfield, S., McDermott, T. and Young, M. (2001) Proc. Natl. Acad. Sci. U.S.A. **98**, 13341–13345
5 Rachel, R., Bettstetter, M., Hedlund, B.P., Häring, M., Kessler, A., Stetter, K.O. and Prangishvili, D. (2002) Arch. Virol. **147**, 2419–2429
6 Geslin, C., Le Romancer, M., Gaillard, M., Erauso, G. and Prieur, D. (2003) Res. Microbiol. **154**, 303–307
7 Janekovic, D., Wunderl, S., Holz, I., Zillig, W., Gierl, A. and Neumann, H. (1983) Mol. Gen. Genet. **192**, 39–45
8 Arnold, H.P., Zillig, W., Ziese, U., Holz, I., Crosby, M., Utterback, T., Weidmann, J.F., Kristjansson, J., Klenk, H.-P., Nelson, K.E. and Fraser, C.M. (2000) Virology **267**, 252–266
9 Bettstetter, M., Peng, X., Garrett, R.A. and Prangishvili, D. (2003) Virology **315**, 68–79
10 Prangishvili, D., Arnold, H.P., Götz, D., Ziese, U., Holz, I., Kristjansson, J.K. and Zillig, W. (1999) Genetics **152**, 1387–1396
11 Martin, A., Yeats, S., Janekovic, D., Reiter, W.-D., Aicher, W. and Zillig, W. (1984) EMBO J. **3**, 2165–2168
12 Schleper, C., Kubo, K. and Zillig, W. (1992) Proc. Natl. Acad. Sci. U.S.A. **89**, 7645–7649
13 Arnold, H.P., She, Q., Phan, H., Stedman, K., Prangishvili, D., Holz, I., Kristjansson, J.K., Garrett, R.A. and Zillig, W. (1999) Mol. Microbiol. **34**, 217–226
14 Stedman, K., She, Q., Phan, H., Arnold, H.P., Holz, I., Garrett, R.A. and Zillig, W. (2003) Res. Microbiol. **154**, 295–302
15 Arnold, H.P., Ziese, U. and Zillig, W. (2000) Virology **272**, 409–416
16 Zillig, W., Prangishvili, D., Schleper, C., Elferink, M., Holz, I., Albers, S., Janekovic, J. and Holz, I. (1996) FEMS Microbiol. Rev. **18**, 225–236
17 Prangishvili, D., Stedman, K.M. and Zillig, W. (2001) Trends Microbiol. **9**, 39–43
18 Muskhelishvili, G., Palm, P. and Zillig, W. (1993) Mol. Gen. Genet. **237**, 334–342
19 She, Q., Brügger, K. and Chen, L. (2002) Res. Microbiol. **153**, 325–332
20 She, Q., Peng, X., Zillig, W. and Garrett, R.A. (2001) Nature (London) **409**, 478
21 Peng, X., Blum, H., She, Q., Mallok, S., Brügger, K., Garrett, R.A., Zillig, W. and Prangishvili, D. (2001) Virology **291**, 226–234
22 She, Q., Singh, R.K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M.J., Chan-Weiher, C.C.-Y., Clausen, I.G., Curtis, B.A. et al. (2001) Proc. Natl. Acad. Sci. U.S.A. **98**, 7835–7840
23 Holland, J.J., De La Torre, J. and Steinhauer, D.A. (1992) Curr. Top. Microbiol. Immunol. **176**, 1–20
24 Palm, P., Schleper, C., Gramp, B., Yeats, S., McWilliam, P., Reiter, W.-D. and Zillig, W. (1991) Virology **185**, 242–250
25 Neumann, H. (1988) Ph.D. Dissertation, Ludwig-Maximilians-Universität München, Munich, Germany
26 Prangishvili, D., Klenk, H.-P., Jacobs, G., Schmiechen, A., Hanselman, C., Holz, I. and Zillig, W. (1998) J. Biol. Chem. **273**, 6024–6029
27 Birkenbihl, R.P., Neef, K., Prangishvili, D. and Kemper, B. (2001) J. Mol. Biol. **309**, 1067–1076
28 Pagel, W. (1999) Nature (London) **401**, 877–884
29 Kornberg, A. and Baker, T.A. (1992) DNA Replication, Freeman, New York
30 Blum, H., Zillig, W., Mallok, S., Domdey, H. and Prangishvili, D. (2001) Virology **281**, 6–9
31 Moss, B. (1996) in Fields Virology (Fields, B.N., Knipe, D.M. and Howley, P.M., eds.), pp. 2637–2671, Lippicott-Raven, Philadelphia
32 Bamford, D.H. (2003) Res. Microbiol. **154**, 231–236
33 Bell, P.J.L. (2001) J. Mol. Evol. **53**, 251–256
34 Takemura, M. (2001) J. Mol. Evol. **52**, 419–425