JOURNAL OF BACTERIOLOGY, Nov. 2004, p. 7745–7753 Vol. 186, No. 22
0021-9193/04/$08.00+0 DOI: 10.1128/JB.186.22.7745–7753.2004

Transcription of the Rod-Shaped Viruses SIRV1 and SIRV2 of the Hyperthermophilic Archaeon *Sulfolobus*†

Alexandra Kessler,[1] Arie B. Brinkman,[2] John van der Oost,[2] and David Prangishvili[1,3]*

Department of Microbiology, University of Regensburg, Regensburg, Germany[1]; Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands[2]; and Institut Pasteur, Paris, France[3]

Received 28 May 2004/Accepted 16 August 2004

The double-stranded DNA genomes of the crenarchaeal rudiviruses SIRV1 (32 kb) and SIRV2 (35 kb) were previously sequenced. Here we present results of the analysis of gene expression of these viruses at different time points after infection of the host cell, *Sulfolobus islandicus*, and of the mapping of transcriptional start sites. Transcription of both genomes starts simultaneously at multiple sites spread over the total length of the genome and from both strands. The earliest time point when viral transcripts could be detected in cells was 30 min after infection. At this time point all the viral genes, except one, were transcribed. Many genes were clustered and appeared to be transcribed as polycistronic messengers. Although the coat protein-encoding gene was initially also transcribed as a polycistronic messenger, an abundant monocistronic transcript of this gene was detected 2 to 3 h after infection, just before assembly of viral particles. The expression of a single gene, adjacent to the coat protein gene, was upregulated at the late phase of infection, suggesting that it might be involved in specific processing and activation of the coat protein messenger. Start sites of 13 transcripts from the SIRV1 genome have been mapped by primer extension, and promoter sequences have been identified. Similar to host promoters, these viral promoters all contain potential binding sites for the archaeal transcription factors TATA binding protein and transcription factor B. In addition, most of them contain a virus-specific consensus element, suggesting the involvement of alternative transcription factors.

The diversity of morphotypes of double-stranded (ds) DNA viruses which infect hyperthermophilic crenarchaeota is exceptional and unprecedented for prokaryotic virus-host systems. About two dozen of these viruses have been isolated and propagated in members of the genera *Sulfolobus*, *Acidianus*, *Thermoproteus*, and *Pyrobaculum*. Due to their unique morphotypes, the viruses have been assigned to six novel families: spindle-shaped *Fuselloviridae*, flexible filamentous *Lipothrixviridae*, stiff rod-shaped *Rudiviridae*, droplet-shaped Guttaviridae (reviewed in reference 15), spherical Globuloviridae (6), and two-tailed Bicaudaviridae (M. Häring, G. Vestegaard, R. Rachel, L. Chen, R. A. Garrett, and D. Prangishvili, unpublished data). Surprisingly, members of none of these six families are found among known viruses of bacteria and euryarchaeota and, vice versa, no members of viruses of bacteria and euryarchaeota (*Myoviridae*, *Siphoviridae*, *Podoviridae*, *Tectiviridae*, *Corticoviridae*, and *Plasmaviridae* [reviewed in reference 24]) have been found which infect crenarchaeota.

Apart from the distinct morphology, the genome organization of viruses of hyperthermophilic crenarchaeaota also appears to be unique. More than 90% of all open reading frames (ORFs) of the sequenced genomes still have not been assigned functions and share no homology to sequences in public databases (reviewed in reference 13).

Viruses of *Crenarchaeaota* are also exceptional with respect to virus-host interactions. The vast majority of dsDNA viruses of bacteria and euryarchaea eventually kill the host cell during release of progeny virions; in contrast, crenarchaeal viruses establish a productive infection without killing or lysing the cell, a so-called carrier state. In the case of the fusellovirus SSV1 of *Sulfolobus*, induction of virus production in the lysogenic strain has been observed (8, 20); however, again, this does not result in cell lysis. In all cases, the infected cells continue to replicate themselves as well as the virus, although cellular growth is slowed down, suggesting modification of some specific cell functions. Modification of cellular transcription as a result of infection is well studied in bacterial virus-host systems (9, 10).

Little is known about mechanisms and controls of viral gene expression in hyperthermophilic archaea. In several cases controls of gene expression have been observed but not elucidated (reviewed in reference 15). The only systematic studies on transcription were conducted on the virus SSV1 of *Sulfolobus*; however, these studies followed UV induction of virus production in the SSV1 lysogene rather than the infection cycle. By Northern analysis eight constitutive transcripts and one UV-inducible transcript were mapped and in nuclease S1 mapping were identified promoter regions (17, 18) and termination sites (19). The results enabled identification of archaeal promoter sequences. Their similarity to TATA box-containing promoters of the eukaryal RNA polymerase II confirmed original observations of the resemblance of archaeal transcription machinery with the eukaryal RNA polymerase II apparatus (16, 27). Subsequent biochemical studies provided strong arguments for this notion (reviewed in references 1, 22, and 26).

Here we present results of the first systematic study of transcription of viruses of crenarchaeota over the replication cycle. As a model were two closely related rudiviruses of *Sulfolobus*, SIRV1 (variant VIII) and SIRV2 (12). Linear dsDNA ge-

* Corresponding author. Mailing address: Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris, Cedex 15, France. Phone: 33(0)144 38 9119. Fax: 33(0)145 68 8834. E-mail: prangish@pasteur.fr.

† Supplemental material for this article may be found at http://jb.asm.org/.

nomes of the two viruses, 32,312 and 35,502 bp long, have been sequenced and their nucleotide sequences are available in the EMBL and GenBank data libraries under accession numbers AJ344259 and AJ414696 (5, 11). The genomes are highly similar in their organization. They consist of blocks of well-conserved sequences, 74 to 94% identical, separated by sequences with low or no similarity. The genomes have covalently closed ends and carry inverted terminal repeats (ITRs) of 2,029 and 1,628 bp, respectively, which contain multiple direct repeats. G+C content of both viral genomes is extremely low, 25%, compared with 37% for the genome of *Sulfolobus solfataricus* (21).

MATERIALS AND METHODS

Propagation and purification of virus particles. The viruses SIRV1 and SIRV2 were propagated in *Sulfolobus islandicus* strains REN2H1 and LAL14/1, respectively, and purified as described by Zillig et al. (25). Virus titer was determined by plaque test on Gelrite plates as described earlier (12).

DNA and oligonucleotides. Viral DNA was isolated from the particles by treating them for 30 min with 1.0% *n*-laurylsarcosinate at room temperature, followed by phenol chloroform extraction. Virus-specific dsDNA probes used in Northern hybridizations were generated by PCR. Sequences of the primer pairs used as well as of the oligonucleotides used as single-stranded probes are provided in the supplemental material (see Tables S1 and S2).

RNA isolation. Cells of *S. islandicus* LAL14/1 or REN2H1 were grown to an optical density at 600 nm of 0.2 (10^8 cells/ml) and infected with SIRV1 or SIRV2 at a multiplicity of infection (MOI) of 5. Following the time course of infection, 20-ml aliquots were taken at 0, 30, 60, 120, 180, and 240 min postinfection (p.i.). Cells were pelleted, and the RNA was extracted using the RNeasy system from QIAGEN. Thereby required complete homogenization of the samples was reached by passing them through a Qiashredder column (QIAGEN). To remove DNA contaminations, the RNA was treated with DNase I (QIAGEN) during the purification procedure directly on the column. Concentration and purity of the RNA were determined by the absorbances at 260 and 280 nm.

Northern blotting and hybridizations. RNA samples were analyzed by denaturing electrophoresis through 1.2% agarose gels containing 0.66% formaldehyde. The running buffer contained 20 mM morpholinepropanesulfonic acid, 5 mM sodium acetate, and 1 mM EDTA. After electrophoresis for 2.5 h at 120 V, gels were soaked in 0.05 N sodium hydroxide for 20 min, rinsed in RNase-free water, soaked in 20× SSC (0.3 M Na_3-citrate, 3 M NaCl; pH 7.0), and then transferred to Biodyne B membranes (Pall) by capillary transfer with 20× SSC. RNA was fixed to the membrane by baking for 30 min at 80°C. Hybridizations with dsDNA probes were done in hybridization buffer containing 50% formamide, 5× SSC, 5× Denhardt's reagent (50× Denhardt's reagent is 5 g of Ficoll type 400, 5 g of polyvinylpyrrolidone, 5 g of bovine serum albumin, and H_2O to 500 ml), and denatured, fragmented salmon sperm DNA (100 μg/ml) at 42°C overnight. Hybridizations with oligonucleotides were done in hybridization buffer containing 5× SSC, 5× Denhardt's reagent, 0.05 M sodium phosphate (pH 6.5), 0.1% sodium dodecyl sulfate (SDS), and denatured, fragmented salmon sperm DNA (100 μg/ml) at 42°C overnight. Membranes were washed two times for 5 min in washing buffer (0.2× SSC, 0.1% SDS) at room temperature, two times for 15 min in washing buffer at 60°C, and finally rinsed in 2× SSC. For the oligonucleotides, washing of the membranes was done for 15 min in washing buffer (1× SSC, 0.1% SDS) at room temperature, two times for 20 min in washing buffer at 55°C, and finally rinsed in 2× SSC. Probes were ^{32}P labeled with the Megaprime DNA labeling system (Amersham) (dsDNA probes) or with T4 polynucleotide kinase (oligonucleotides).

Primer extension analysis. Transcription start sites were determined by using the reverse transcription system of Promega according to a modified protocol. Primer annealing was done in a separate reaction for 10 min at 70°C and then slowly cooled down to room temperature. The components for the reverse transcription reaction were added, and the primer extension was exceeded for 30 min at 42°C. Reactions were terminated by adding 250 μl of 10 mM Tris, pH 8.0, 10 mM EDTA, 750 mM NaCl, and 1% SDS. Afterwards, 10 μg of glycogen was added and DNA was extracted with phenol-chloroform. The used oligonucleotides are listed in the supplemental material (see Table S3).

RESULTS

The genomes of viruses SIRV1 and SIRV2 contain 45 and 54 ORFs longer than 150 nucleotides, respectively, of which 44 are homologous to each other. ORF maps of both genomes (Fig. 1) were modified from those reported by Peng et al. (11). On the basis of homology to well-characterized counterparts, functions were assigned and later confirmed for ORF158a (dUTPase) (14) and ORF121 (Holliday junction resolvases) (4). On the basis of N-terminal sequencing, the gene encoding the coat protein, ORF 134, has been identified (11). Putative encoded functions include two glycosyl transferases of group 1 from each virus (SIRV1 ORFs 356 and 335 and SIRV2 ORFs 356 and 335).

Transcription maps. Transcription of SIRV1 variant VIII (SIRV1/VIII) was studied in *S. islandicus* REN2H1, the host in which this otherwise highly mutable virus remains conditionally stable (12). SIRV2 was replicated in *S. islandicus* LAL14/1. For synchronization of host cell populations, a simple method was used, based on dilution of stationary-phase cultures into fresh medium (7). Exponentially growing cells were infected by viruses at an MOI of 5. Previously, characteristics of interactions of the viruses with these hosts have been determined (12). The time interval required for infection of 50% of host cells by both viruses, at an MOI of 5, was about 14 min for both viruses. The eclipse period, the time interval between infection and the appearance of intracellular virus particles, was 4.0 h for SIRV1 and 2.5 h for SIRV2. The latent period, the time interval between infection and release of the first virus particles, was 8 h for SIRV1 and 6 h for SIRV2. Based on these results, we have chosen time intervals for isolation of RNA from infected host cells. For this, we took aliquots of growing cultures of infected cells 30 min p.i., 1 h p.i., and later at hourly intervals up to 4 h p.i. As a control, RNA was prepared from noninfected cells.

The same amount of RNA (5 μg) prepared from virus-infected cells at different time points p.i. was separated by agarose gel electrophoresis and blotted onto nylon membranes for Northern hybridization. In hybridization experiments with whole-genome probes, numerous virus-specific transcripts were detected which were difficult to interpret and assign to specific ORFs (data not shown). For more precise identification of transcripts in hybridization experiments we used fragments of the viral genomes, produced by PCR amplification. In total, 11 fragments of the SIRV1 genome and 12 fragments of the SIRV2 genome were produced. Coordinates of terminal nucleotides of these fragments in the complete nucleotide sequences of the genomes are presented in Table 1. In the same table, the results of 23 Northern hybridization experiments are summarized, indicating sizes of RNA detected with different DNA probes at different time intervals p.i. All results were evaluated, and putative ORFs, transcription of which should have caused detected signals, are also indicated in Table 1. In several cases we observed very weak signals from RNAs longer than 3 kilonucleotides (knt). Because in these cases no corresponding operons could be identified, we did not consider them in evaluation of the data. In some cases, there were ambiguities in interpretation of data. These were resolved by designing single-stranded ORF-specific oligonucleotide probes and using them in Northern hybridization experiments. In to-

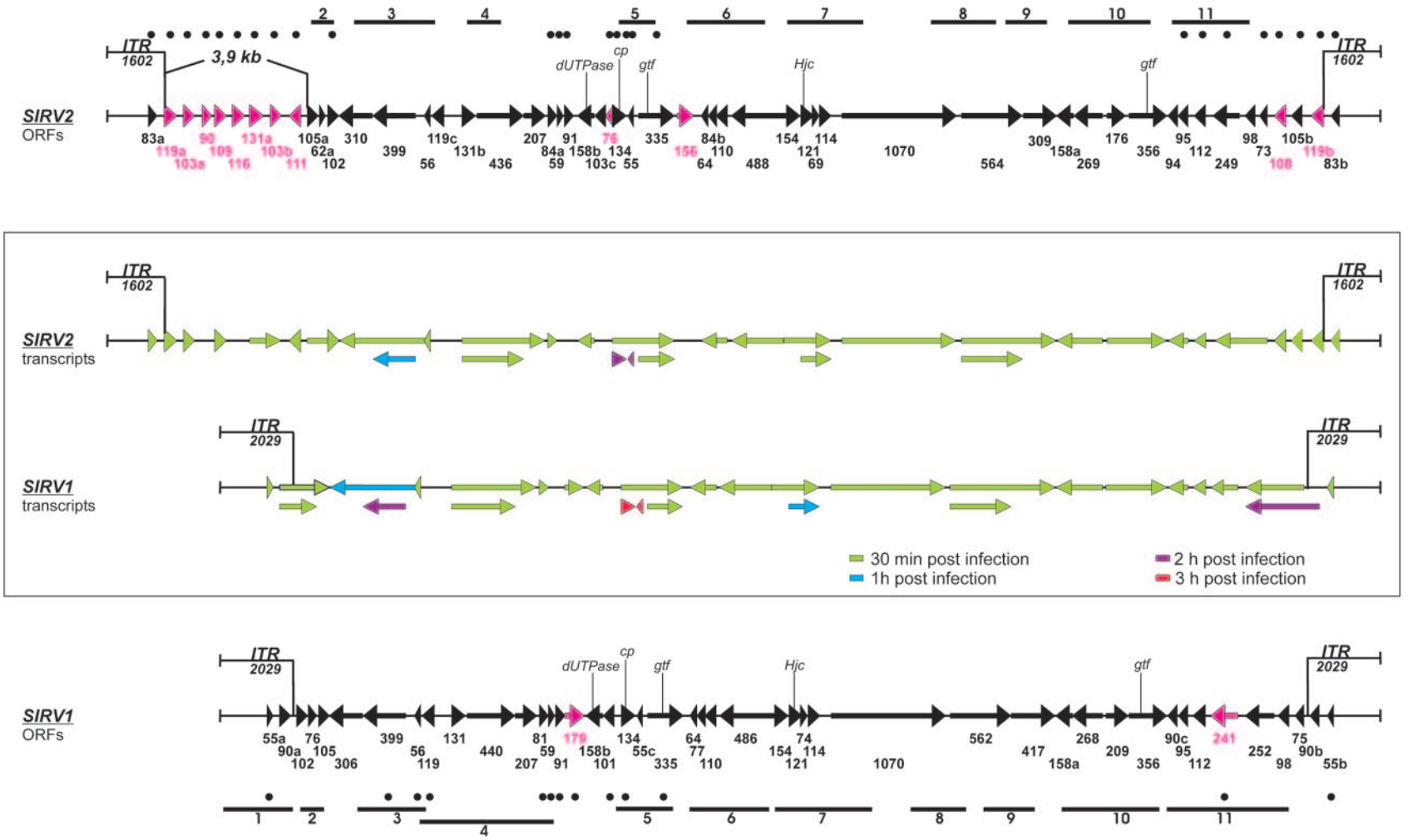

FIG. 1. ORF and transcription maps of SIRV1 (variant VIII) and SIRV2. ORF maps were modified with permission from those of Peng et al. (11). The numbers used to identify ORFs refer to the number of codons. Specific ORFs for either SIRV1 (variant VIII) or SIRV2 are highlighted in pink. Genes with identified or putative functions are indicated as follows: cp, coat protein; Hjc, Holliday junction resolvase; dUTPase, deoxyuridine triphospatase; gtf, glycosyl transferase. dsDNA probes used in the Northern hybridization are represented by bars and are numbered according to the information in Table 1. Single-stranded DNA probes are represented by dots. The transcription maps are surrounded by a frame.

TABLE 1. Transcripts from genomes of the viruses SIRV1 and SIRV2 identified by Northern hybridization with double-stranded and single-stranded DNA probes[a]

Virus and dsDNA probe (coordinates)	Size of transcript(s) in knt (included ORFs) p.i. at:				
	30 min	1 h	2 h	3 h	4 h
SIRV1					
1 (108–2024)	1.9 (90a, 102, 76)	1.9 (90a, 102, 76)	1.9 (90a, 102, 76)	1.9 (90a, 102, 76)	1.9 (90a, 102, 76)
	1.2 (90a, 102, 76, 105)	1.2 (90a, 102, 76, 105)	2.2 (90b, 75, 98, 252)	2.2 (90b, 75, 98, 252)	2.2 (90b, 75, 98, 252)
2 (2211–2890)	1.0 (90a, 102, 76)	1.0 (90a, 102, 76)	1.0 (90a, 102, 76)	1.0 (90a, 102, 76)	1.0 (90a, 102, 76)
	1.2 (90a, 102, 76, 105)	1.2 (90a, 102, 76, 105)			
3 (3821–5720)	<0.24 (*56*)	<0.24 (*56*)	<0.24 (*56*)	<0.24 (*56*)	<0.24 (*56*)
		2.3 (56, 399, 306)	2.3 (56, 399, 306)	2.3 (399, 306)	2.3 (399, 306)
			1.1 (*399*)	1.1 (*399*)	1.1 (*399*)
4 (5547–9279)					
	1.6 (131, 440)	1.6 (131, 440)	1.6 (131, 440)	1.6(131, 440)	1.6(131, 440)
	2.4 (131, 440, 207)	2.4 (131, 440, 207)	2.4 (131, 440, 207)	2.4 (131, 440, 207)	2.4 (131, 440, 207)
5 (11173–12738)	1.2 (*335*)	1.2 (*335*)	0.6 (*134*)	0.6 (*134*)	0.6 (*134*)
	1.8 (134, 335)	1.8 (134, 335)	1.2 (*335*)	1.2 (*335*)	1.2 (*335*)
			1.8 (134, 335)	1.8 (134, 335)	1.8 (134, 335)
6 (13062–15281)	0.65 (64, 77, 110)	0.65 (64, 77, 110)	0.65 (64, 77, 110)	0.65 (64, 77, 110)	0.65 (64, 77, 110)
	1.5 (488)	1.5 (488)	1.5 (488)	1.5 (488)	1.5 (488)
7 (15461–18169)	1.5 (154, 121, 74, 114)	1.0 (121, 74, 114)	1.0 (121, 74, 114)	1.0 (121, 74, 114)	1.5 (154, 121, 74, 114)
	3.0 (1070)	1.5 (154, 121, 74, 114)	1.5 (154, 121, 74, 114)	1.5 (154, 121, 74, 114)	3.0 (1070)
		3.0 (1070)	3.0 (1070)	3.0 (1070)	
8 (19241–20762)	1.5 (562)	1.5 (562)	1.5 (562)	1.5 (562)	1.5 (562)
	3.0 (1070/562, 417)	3.0 (1070/562, 417)	3.0 (1070/562, 417)	3.0 (1070/562, 417)	3.0 (1070/562, 417)
9 (21278–22668)	1.5 (562)	1.5 (562)	1.5 (562)	1.5 (562)	1.5 (562)
	2.8 (562, 417)	2.8 (562, 417)	2.8 (562, 417)	2.8 (562, 417)	2.8 (562, 417)
10 (23429–26131)	1.3 (158a, 268)	1.3 (158a, 268)	1.3 (158a, 268)	1.3 (158a, 268)	1.3 (158a, 268)
	2.3 (209, 356)	2.3 (209, 356)	2.3 (209, 356)	2.3 (209, 356)	2.3 (209, 356)
11 (26381–29772)	0.65 (90c, 95/*241*)	0.65 (90c, 95/*241*)	0.65 (90c, 95/*241*)	0.65 (90c, 95/*241*)	0.65 (90c, 95/*241*)
	1.4 (252, 98, 75)	1.4 (252, 98, 75)	1.4 (252, 98, 75)	1.4 (252, 98, 75)	1.4 (252, 98, 75)
			2.2 (90b, 75, 98, 252)	2.2 (90b, 75, 98, 252)	2.2 (90b, 75, 98, 252)
SIRV2					
2 (5681–6307)	1.0 (105a, 62a, 102)	1.0 (105a, 62a, 102)	1.0 (105a, 62a, 102)	1.0 (105a, 62a, 102)	1.0 (105a, 62a, 102)
3 (6871–9127)	<0.24 (*56*)	<0.24 (*56*)	<0.24 (*56*)	<0.24 (*56*)	<0.24 (*56*)
	2.2 (56, 399, 310)	2.2 (56, 399, 310)	2.2 (56, 399, 310)	2.2 (56, 399, 310)	2.2 (56, 399, 310)
		1.4 (*399*)	1.4 (*399*)	1.4 (*399*)	1.4 (*399*)
4 (10016–10980)					
	1.6 (131b, 436)	1.6 (131b, 436)	1.6 (131b, 436)	1.6 (131b, 436)	1.6 (131b, 436)
	2.8 (131b, 436, 207)	2.8 (131b, 436, 207)	2.8 (131b, 436, 207)	2.8 (131b, 436, 207)	2.8 (131b, 436, 207)
5 (14231–15264)	1.2 (*335*)	1.2 (*335*)	0.6 (*134*)	0.6 (*134*)	0.6 (*134*)
	1.8 (134, 335)	1.8 (134, 335)	1.2 (*335*)	1.2 (*335*)	1.2 (*335*)
			1.8 (134, 335)	1.8 (134, 335)	1.8 (134, 335)
6 (16145–18324)	0.9 (64, 84b, 110)	0.9 (64, 84b, 110)	0.9 (64, 84b, 110)	0.9 (64, 84b, 110)	0.9 (64, 84b, 110)
	1.8 (488)	1.8 (488)	1.8 (488)	1.8 (488)	1.8 (488)
7 (18931–21070)	0.65 (121, 69, 114)	0.65 (121, 69, 114)	0.65 (121, 69, 114)	0.65 (121, 69, 114)	0.65 (121, 69, 114)
	1.5 (154, 121, 69, 114)	1.5 (154, 121, 69, 114)	1.5 (154, 121, 69, 114)	1.5 (154, 121, 69, 114)	1.5 (154, 121, 69, 114)
	3.0 (1070)	3.0 (1070)	3.0 (1070)	3.0 (1070)	3.0 (1070)
8 (22946–24770)	1.5 (564)	1.5 (564)	1.5 (564)	1.5 (564)	1.5 (564)
	3.2 (1070)	3.2 (1070)	3.2 (1070)	3.2 (1070)	3.2 (1070)
9 (25021–26180)	1.6 (564)	1.6 (564)	1.6 (564)	1.6 (564)	1.6 (564)
	3.1 (564, 309)	3.1 (564, 309)	3.1 (564, 309)	3.1 (564, 309)	3.1 (564, 309)
10 (26771–29080)	1.0 (158a, 269)	1.0 (158a, 269)	1.0 (158a, 269)	1.0 (158a, 269)	1.0 (158a, 269)
	1.5 (176, 356)	1.5 (176, 356)	1.5 (176, 356)	1.5 (176, 356)	1.5 (176, 356)
11 (29649–31841)	0.4 (*112*)	0.4 (*112*)	0.4 (*112*)	0.4 (*112*)	0.4 (*112*)
	0.8 (94, 95)	0.8 (94, 95)	0.8 (94, *95*)	0.8 (94, 95)	0.8 (94, 95)
	1.2 (*249,* 98, 73)	1.2 (*249,* 98, 73)	1.2 (*249,* 98, 73)	1.2 (*249,* 98, 73)	1.2 (*249,* 98, 73)

[a] Nucleotide positions are according to EMBL/GenBank data libraries under accession nos: AJ344259 and AJ414696. The average transcript size is given. In parentheses following transcript sizes are the indicated ORFs which most likely were encompassed by the transcript. ORFs are referred to by numbers of amino acids in the predicted proteins. In italics are shown ORFs for which transcripts were also demonstrated by hybridization with single-stranded oligonucleotide probes.

tal, 27 hybridizations were conduced with ORF-specific single-stranded probes, the results of which are also summarized in Table 1.

As an example of evaluation of the data, results of the Northern hybridization with probe 3 for SIRV1 are presented in Fig. 2A. This DNA fragment encompassed ORF56, ORF399, ORF306, and ORF119. Small abundant RNA, less than 0.24 knt in length, which hybridized with the probe should be a transcript of ORF56. This was confirmed by hybridizations with an ORF56-specific oligonucleotide. A signal in the range of bands with the length of 2.0 to 2.3 knt, appearing very early, could be produced by a long transcript encompassing ORF119,

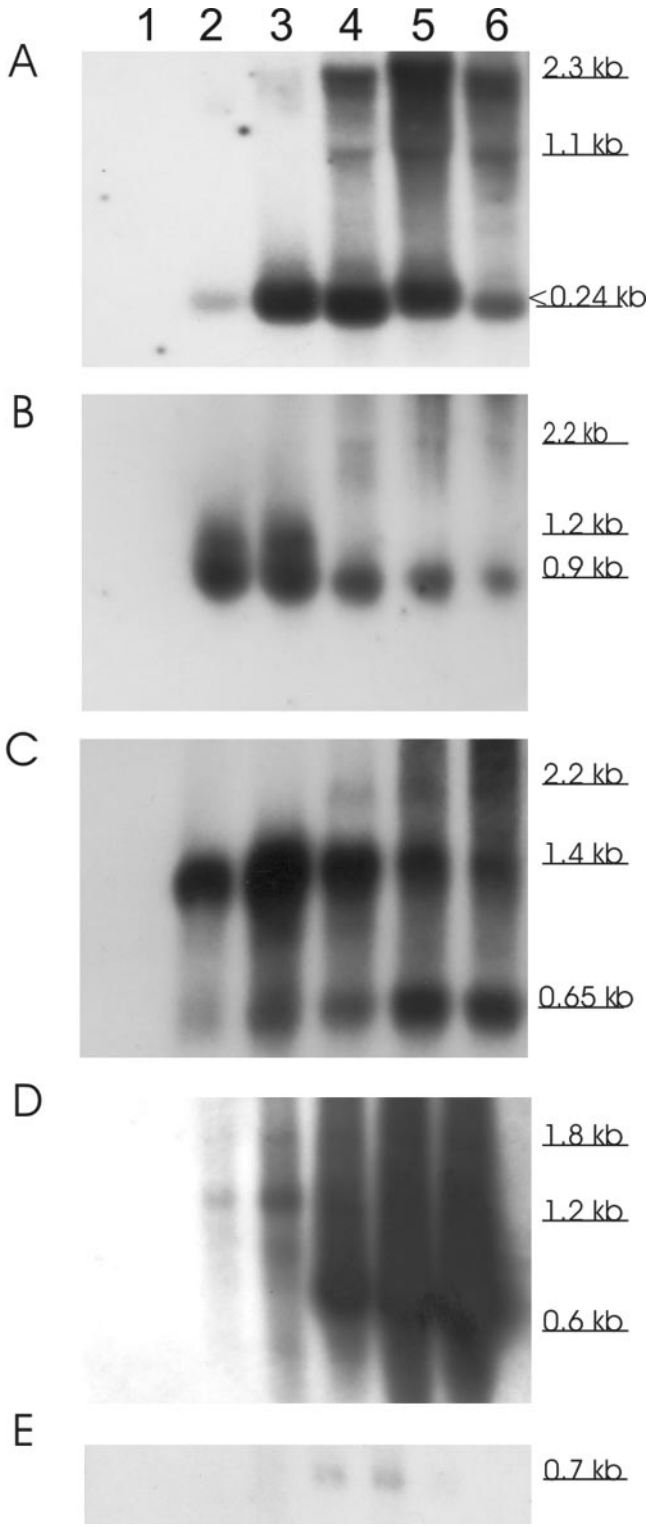

FIG. 2. Northern blot hybridization of infected cell RNA to specific SIRV1 and SIRV2 genome probes. RNA was extracted from cells at various times after infection. Lane 1, uninfected cell RNA; lanes 2 to 6, 30 min, 1 h, 2 h, 3 h, and 4 h p.i., respectively. Double-stranded probes were following: SIRV1 probe 3, covering ORFs 306 and 119 (A); SIRV1 probe 1, covering ORFs 55a and 90a (B); SIRV1 probe 11, covering ORFs 90c and 98 (C); SIRV2 probe 5, covering ORFs 134 and 355 (D); hybridization with an oligonucleotide complementary to a portion of SIRV2 ORF 55 (E).

ORF56, ORF399, and ORF306. However, ORF119 apparently was not transcribed, as suggested by the absence of any signal of hybridization with ORF119-specific oligonucleotide and, thus, it was excluded from the operon. RNA of about 1.1 knt, appearing 2 h p.i. should be a transcript either of ORF399 or of ORF306. Hybridization with 24-mer oligonucleotide complementary to a portion of the coding sequence of SIRV1 ORF399 revealed a signal in this region, suggesting that the 1.1-knt RNA represented a transcript of ORF399. Transcription of ORF399 was also confirmed by primer extension experiments (see below).

Transcription of the ITRs of SIRV1 was studied with probes 1 and 11 (Fig. 2B and C). The two bands with lengths of about 0.9 and 1.2 knt hybridizing with probe 1 were also detected by probe 2 (Table 1) and, thus, apparently represent transcripts starting from the leftward ITR and running across the ITR border. Most likely, they both start at ORF90a. A start at ORF55a was excluded by the results of a hybridization with a probe corresponding to this ORF, showing a signal with a transcript shorter than 0.5 knt. A weak, long transcript of about 2.2 knt detected with probe 1 was not visible in hybridizations with probe 2 and, thus, most likely it is the result of cross-hybridization with the partly homologous rightward ITR. This was indeed confirmed by detection of the same transcript with probe 11 (Fig. 2C).

Transcripts were detected over the whole genomes, and all identified ORFs of both viruses, except those mentioned below, were transcribed. They included also the ORFs unique to one or the other of the highly similar genomes. However, four SIRV2-specific ORFs, ORF156, ORF90, ORF116, and ORF76, were apparently not transcribed. By using either the above-mentioned probes or ORF-specific oligonucleotides, no transcription was detected for SIRV1 and SIRV2 ORF119c, ORF59, and ORF 91 and SIRV1 ORF 101/SIRV2 ORF103c.

The earliest time point when any transcript could be clearly detected was 30 min p.i. However, at this time point most of the ORFs of both genomes appeared to be transcribed. Transcription started nearly simultaneously from 21 sites on the genome of SIRV1 and from 29 sites in the genome of SIRV2. No strand specificity was observed in early transcription: both strands were transcribed with nearly equal efficiency. Transcripts were often large enough to span several ORFs, suggesting many ORFs were clustered into operons. Relative abundance of each transcript could be estimated only approximately; moreover, it was estimated only in those cases when compared signals were revealed in the same hybridization experiment. Generally, within 30 min all transcripts were relatively weak and reached higher levels 1 h p.i. Later, decrease of signals from some transcripts was observed, e.g., from SIRV1 ORF56 and the two transcripts starting from the leftward ITR of SIRV1 (Fig. 2B). The longer of these transcripts could not be detected at later stages of the replication cycle.

Very few transcripts appeared later, in addition to those observed 30 min p.i. These included monocistronic transcripts of SIRV1/SIRV2 ORF399 and those starting from SIRV1 ORF90b in the rightward ITR. In the latter case the signal was relatively weak, and we cannot completely exclude the possibility that a small amount of the transcript was also present earlier in the infection cycle. Moreover, the transcript starting from the identical site on the leftward ITR appeared already at

30 min p.i. However, the monocistronic transcript of ORF134 (coat protein gene) was clearly absent among early transcripts of both viruses (shown in Fig. 2D for SIRV2). In the host of SIRV2 it appeared 2 h p.i., and in SIRV1 host cells it appeared at 3 h p.i., in both cases close to the eclipse period of the infection cycle. It is noteworthy that results of hybridization with corresponding probes (probe 5) suggested that the coat protein gene of both viruses was transcribed in the context of the operon already at 30 min p.i. (Table 1 and Fig. 2D; data not shown for SIRV1).

The only ORF which apparently was transcribed exclusively late in the infection cycle was SIRV2 ORF55/SIRV1 ORF55c. Appearance of a weak transcript of this ORF coincided with the late appearance of the monocistronic transcript of the coat protein gene (ORF134; Fig. 2E). It is noteworthy that transcription of ORF55/ORF55c starts immediately downstream of the coat protein gene and proceeds in the other direction. Thus, the RNA would be complementary to the intergenic region of the early polycistronic transcript of ORF134-ORF335.

Results of evaluation of all hybridization experiments are schematically represented as transcription maps of the two viruses in Fig. 1. The majority of identified transcripts, shown in Fig. 1, could be terminated at T-rich sequences which were found downstream of stop codons of translation of the last (or only) ORF in an operon. However, in extremely A+T-rich genomes of SIRV1 and SIRV2 (75% A+T), it is not justified to speculate about putative functions of T-rich segments.

Primer extension and sequence of promoters. To confirm transcription start sites and identify promoters, primer extension experiments were conducted. RNA isolated from SIRV1-infected cells 3 h p.i. was used in reverse transcription experiments. At this time, transcripts from all start sites were available in host cells. The 5′ termini for 13 transcripts from the virus SIRV1 were precisely mapped by comparing the primer extension product with a sequencing ladder generated with the same primer. Some typical examples are shown in Fig. 3, and the results are summarized in Fig. 4. Transcripts were termed according to the first (or only) ORF which they encompass, and those for which 5′ termini have been mapped are in italics. Comparison of the sequences around transcriptional initiation sites revealed the presence of three conserved sequence elements, as indicated in Fig. 4: (i) an AT-rich pentanucleotide sequence centered 20 to 25 nucleotides upstream of the transcription start site, presumably corresponding to the TATA box; (ii) two purines, presumably representing the transcription factor B-responsive element (BRE), immediately upstream of the potential TATA box; and (iii) a trinucleotide GTC or, more generally, a pentanucleotide A/TGTCA/T, immediately downstream of presumable TATA boxes in 70% of cases. In four out of five putative promoters of SIRV1 which have not been mapped, these three elements could all be identified (Fig. 4). Putative promoters of SIRV1 ORFs 179, 562, and 112 could not be identified by sequence analysis; due to a very high AT content (75%) in the regions upstream of these ORFs, and also the absence of the trinucleotide GTC, it is difficult to identify putative TATA boxes.

Transcripts from well-conserved regions of the SIRV1 and SIRV2 genomes were very similar in size and ORF content (Fig. 1). It is most likely that start sites of these transcripts are also similar. At least, putative promoters of these SIRV2 transcripts contained conserved elements identical to those found in corresponding SIRV1 promoters (Fig. 4). Putative TATA box and BRE could be recognized also in promoter regions of most of SIRV2-specific ORFs, which were shown to be transcribed (Fig. 1 and 4). In 65% of these, immediately downstream of the TATA box was also present the trinucleotide GTC (Fig. 4).

DISCUSSION

In the present communication, we describe results of the first systematic study of transcription of genomes of viruses of *Crenarchaeota*, one of the two kingdoms of the domain *Archaea*, over the infection cycle. The rod-shaped viruses SIRV1 and SIRV2, transcription of which has been studied, are the only two described representatives of the family *Rudiviridae*, which infect hyperthermophilic crenarchaeaota of the genus *Sulfolobus* (12). The general picture of transcription of genomes of these viruses differs from that of dsDNA genomes of viruses of *Euryarchaeota*, *Bacteria*, or *Eukarya*, for which this has been studied in detail. Generally, in the latter cases transcription of viral genes is temporally regulated, proceeds in waves, and two main classes of genes are recognized (early and middle-late). At the earliest stage of infection with rudiviruses when any transcript could be observed in our experiments, 30 min p.i., 24 transcripts from 21 promoters of SIRV1 and 31 transcripts from 29 promoters of SIRV2 were detected which included all ORFs that are generally transcribed, except ORF55 (ORF55c). Thirty minutes p.i. is an early stage of the viral replication cycle, considering that 15 min was required for infection at an MOI of 3 of 50% of host cells and that the eclipse period was 4 h for SIRV1 and 2.5 h for SIRV2. Moreover, hosts were growing rather slowly, with a generation time of about 6 h. At later stages of the infection cycle, in addition to already existing transcripts, monocistronic transcripts of ORF399 and ORF134 appeared, which in an earlier stage were transcribed as parts of operons.

The rather uniform pattern of transcription in the course of the whole replication cycle fits the characteristics of interactions of the rudiviruses SIRV1 and SIRV2 with their hosts and their presence in host cells in a stable carrier state. This apparently does not require as much transcription control as in, for example, the case of temperate bacteriophages.

The two sites where we presume operation of transcriptional controls are rather long intergenic regions between ORF56 and ORF131(b) and between ORF158b and ORF134. From here, in addition to polycistronic RNAs, later in the infection cycle single gene transcripts are initiated. Termination of transcription downstream of the coat protein gene could be linked with transcription of a small ORF55 (ORF55c) from the counter strand (Fig. 2E), which can function as an antisense RNA. Late appearance of the transcripts of ORF55 (ORF55c) and the coat protein gene coincides with an increased transcription of the latter, suggesting additional positive transcription regulation. The suggestion is supported by the presence of multiple inverted and direct repeats in the long noncoding region upstream of ORF134. It seems reasonable to induce expression of the coat protein at the final stage of the viral replication cycle close to the time of appearance of the first

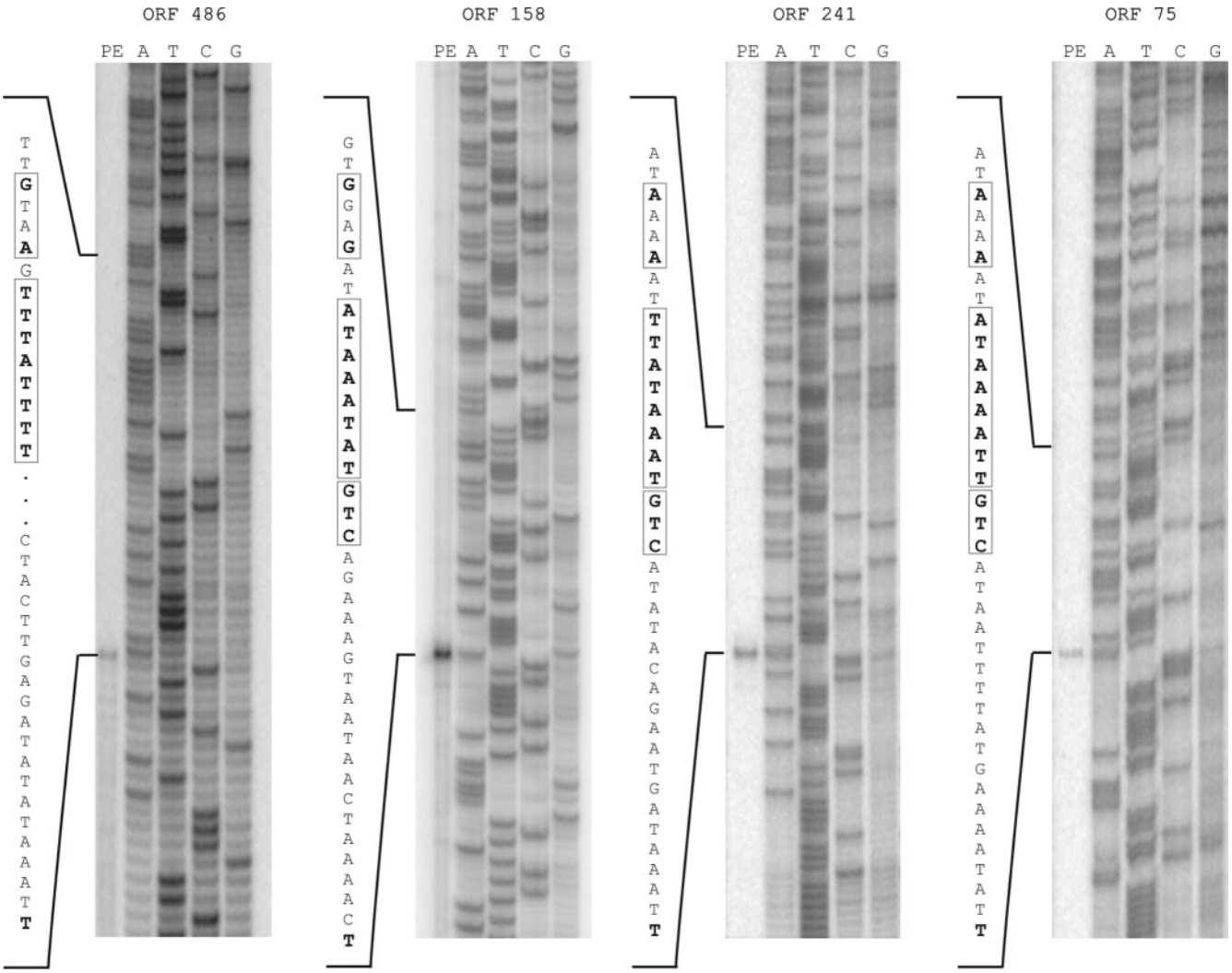

FIG. 3. Mapping of the initiation sites of four transcripts of the SIRV1 virus. The first (or only) ORF of the operon is indicated. Primer extension analysis was carried out using sequence-specific primers (see Materials and Methods). Each of the runoff products (PE) was electrophoresed alongside its DNA sequencing reactions (A, T, C, and G), also carried out with the corresponding primer. The promoter sequences are written in letters next to the autoradiographs. The TATA box, BRE, and GTC feature are highlighted in boxes.

virions, the eclipse. Accumulation in the cell of the coat protein could induce the assembly of rather simply organized nonenveloped virions of rudiviruses, the body of which is just a nucleoprotein consisting of the linear dsDNA and subunits of the coat protein (12).

An apparent candidate for the role of transcription regulator is the putative protein encoded by ORF56. The transcript from this gene is abundant throughout the complete infection cycle (Fig. 2A). The small protein reveals a helix-turn-helix motif characteristic of many DNA binding proteins.

For translation initiation, members of *Sulfolobales* are known to use two different mechanisms. On distal cistrons of polycistronic mRNAs, they mostly operate via Shine-Dalgarno-dependent initiation resembling the system prevalent in extant bacteria, whereas on monocistronic mRNAs and on opening cistrons of polycistronic mRNA they use "leaderless" initiation, reminiscent of the eukaryotic pathway (3, 23). The picture is slightly different for genomes of the rudiviruses, at least of SIRV1, for which most transcription starts have been mapped. Potential Shine-Dalgarno sequences can be identified upstream of 50% of mapped or putative transcription start sites.

Transcription starts of 13 transcripts of the genome of SIRV1 have been determined by primer extension. This comprises 52% of all identified transcripts. In all cases, 20 to 25 nucleotides upstream of the transcription start potential TATA boxes have been identified, as well as an adjacent purine-rich region, which most likely corresponds to the BRE, the transcription factor B binding site identified in all well-characterized *Sulfolobus* promoters (2). Location and consensus sequence of these promoter elements in the genomes of the two rudiviruses are very similar to their counterparts in the *Sulfolobus* hosts (2). In 9 out of 13 mapped SIRV1 promoters (about 70%) and in 4 out of 5 putative promoters (80%), a pentanucleotide A/TGTCA/T is present immediately downstream of the TATA box (Fig. 4). This pentanucleotide is

FIG. 4. DNA sequences of the promoters of SIRV1and SIRV2. Promoters mapped by primer extension are shown in italics. Other promoters were determined by similarities to them. The respective positions of the TATA-box-, BRE-, and GTC elements are shown in pink, orange, and green, respectively. In case of mapped promoters the transcription initiation site is indicated by an asterisk.

present also in most of the predicted promoters of SIRV2 (Fig. 4). Conservation of such a motif has never been reported for promoters of the *Sulfolobus* hosts. However, we have detected it in two of seven mapped promoters of the virus SSV1 of *Sulfolobus* (18, 26). We infer that the GTC motif is a rudivirus-specific *cis*-regulatory element. This suggestion is strongly supported by the presence of the same trinucleotide in most of the putative promoters of yet another rudivirus, ARV1, isolated recently from Italian hot springs (M. Häring, G. Westergaard, R. A. Garrett, and D. Prangishvili, unpublished data). The GTC-containing unusual promoters can be recognized by specific virus- or host-encoded factors. The presence of the trinu-

cleotide GTC in the promoter apparently does not regulate transcription temporally: transcription started simultaneously from GTC-containing promoters and promoters without it (Fig. 1 and 4). Also, no correlation was observed between the presence of GTC in the promoter and the abundance of transcripts.

Several strategies known from prokaryotic and eukaryal viruses could explain the early transcription of genomes of the rudiviruses SIRV1 and SIRV2 and the specific nature of most of the viral promoters: (i) the viruses inject RNA polymerase and/or a transcription regulatory protein at the time of infection; (ii) viral promoters are recognized by cellular factors and as such modulate gene expression; (iii) viral infection is a general transcription stimulus, e.g., up-regulating the activity of the cellular RNA polymerase; (iv) viral genes are transcribed, at least in part, by the basal transcription system of the host, although expression of a viral regulator stimulates preferential transcription of certain viral genes. Understanding the strategies used by the rudiviruses will contribute to our knowledge of mechanisms of transcription-level gene regulation in archaea and will shed light on the evolution and diversity of transcriptional control in general.

ACKNOWLEDGMENTS

We are grateful to Karl. O. Stetter for stimulating discussions.

The work was supported by grants from the Deutsche Forschungsgemeinschaft (PR 663/1-1 and PR 663/1-2).

REFERENCES

1. **Bell, S. D., and S. P. Jackson.** 2001. Mechanism and regulation of transcription in archaea. Curr. Opin. Microbiol. **4:**208–213.
2. **Bell, S. D., P. L. Kosa, P. B. Sigler, and S. P. Jackson.** 1999. Orientation of the transcription preinitation complex in Archaea. Proc. Natl. Acad. Aci. USA **96:**13662–13667.
3. **Benelli, D., E. Maone, and P. Londei.** 2003. Two different mechanisms for ribosome/mRNA interaction in archaeal translation initiation. Mol. Microbiol. **50:**635–643.
4. **Birkenbihl, R. P., K. Neef, D. Prangishvili, and B. Kemper.** 2001. Holliday junction resolving enzymes of archaeal viruses SIRV1 and SIRV2. J. Mol. Biol. **309:**1067–1076.
5. **Blum, H., W. Zillig, S. Mallock, H. Domdey, and D. Prangishvili.** 2001. The genome of the archaeal virus SIRV1 has features in common with genomes of eukaryal viruses. Virology **281:**6–9.
6. **Häring, M., X. Peng., K. Brügger, R. Rachel., K. O. Stetter, R. Garrett, and D. Prangishvili.** 2004. Morphology and genome organization of the virus PSV of the hyperthermophilic archea genera Pyrobaculum and Thermoproteus: a novel virus family, the Globuloviridae. Virology **323:**233–242.
7. **Hjort, K., and R. Bernander.** 1999. Changes in cell size and DNA content in *Sulfolobus* cultures during dilution and temperature shift experiments. J. Bacteriol. **181:**5669–5675.
8. **Martin, A., S. Yeats, D. Janekovic, W.-D. Reiter, W. Aicher, and W. Zillig.** 1984. SAV1, a temperate u.v.-inducible DNA virus like particle from the archaebacterium Sulfolobus acidocaldarius isolate B12. EMBO J. **3:**2165–2168.
9. **Molineux, I. J.** 1999. T7 bacteriophage, p. 1388–1396. *In* R. G. Webster and A. Granoff (ed.), Encyclopedia of virology, vol. 3. Academic Press, San Diego, Calif.
10. **Mosig, G.** 1999. T4 bacteriophage and related bacteriophages, p. 1376–1383. *In* R. G. Webster and A. Granoff (ed.), Encyclopedia of virology,vol. 3. Academic Press, San Diego, Calif.
11. **Peng, X., H. Blum, Q. She, S. Mallok, K. Brügger, R. A. Garrett, W. Zillig, and D. Prangishvili.** 2001. Sequences and replication of genomes of the archaeal rudiviruses SIRV1 and SIRV2: relationships to the archaeal lipothrixvirus SIFV and some eukaryal viruses. Virology **291:**226–234.
12. **Prangishvili, D., H. P. Arnold, U. Ziese, D. Goetz, I. Holz, and W. Zillig.** 1999. A novel virus family, the Rudiviridae: structure, virus-host interactions and genome variability of Sulfolobus viruses SIRV1 and SIRV2. Genetics **153:**1387–1396.
13. **Prangishvili, D., and R. A. Garrett.** 2004. Exceptionally diverse morphotypes and genomes of crenarchaeal hyperthermophilic viruses. Biochem. Soc. Trans. **32:**204–208.
14. **Prangishvili, D., H.-P. Klenk, G. Jakobs, A. Schmiechen, C. Hanselman, I. Holz, and W. Zillig.** 1998. Biochemical and physiological characterization of the dUTPase from the archaeal virus SIRV. J. Biol. Chem. **273:**6024–6029.
15. **Prangishvili, D., K. M. Stedman, and W. Zillig.** 2001. Viruses of the extremely thermophilic archaeon Sulfolobus. Trends Microbiol. **9:**39–42.
16. **Prangishvili, D., W. Zillig, A. Gierl, L. Biesert, and I. Holz.** 1982. DNA-dependent RNA polymerases of thermoacidophilic archaeabacteria. Eur. J. Biochem. **122:**471–477.
17. **Reiter, W.-D., P. Palm, S. Yeats, and W. Zillig.** 1987. Gene expression in archaebacteria: physical mapping of constitutive and UV-inducible transcripts from the Sulfolobus virus-like particle SSV1. Mol. Gen. Genet. **209:** 270–275.
18. **Reiter, W.-D., P. Palm, and W. Zillig.** 1988. Analysis of transcription in the archaebacterium Sulfolobus indicates that archaebacterial promoters are homologous to eukaryotic polII promoters. Nucleic Acids Res. **16:**1–19.
19. **Reiter, W.-D., P. Palm, and W. Zillig.** 1988. Transcription termination in the archaebacterium Sulfolobus: signal structures and linkage to transcription initiation. Nucleic Acids Res. **16:**2445–2459.
20. **Schleper, C., K. Kubo, and W. Zillig.** 1992. The particle SSV1 from the extremely thermophilic archaeon Sulfolobus is a virus: demonstration of infectivity and of transfection with viral DNA. Proc. Natl. Acad. Sci. USA **89:**7645–7649.
21. **She, Q., R. K. Singh, F. Confalonieri, Y. Zivanovic, G. Allard, M. J. Awayez, C. C.-Y. Chan-Weiher, I. G. Clausen, B. A. Curtis, A. De Moors, G. Erauso, C. Fletcher, P. M. K. Gordon, I. Heikamp-de Jong, A. C. Jeffries, C. J. Kozera, N. Medina, X. Peng, H. P. Thi-Ngoc, P. Redder, M. E. Schenk, C. Theriault, N. Tolstrup, R. Charlebois, W. F. Doolittle, M. Duguet, T. Gaasterland, R. A. Garrett, M. A. Ragan, C. W. Sensen, and J. Van der Oost.** 2001. The complete genome of the crenarchaeon Sulfolobus solfataricus P2. Proc. Natl. Acad. Sci. USA **98:**7835–7840.
22. **Thomm, M.** 1996. Archaeal transcription factors and their role in transcription initiation. FEMS Microbiol. Rev. **18:**159–172.
23. **Tolstrup, N., C. W. Sensen, R. A. Garrett, and I. G. Clausen.** 2000. Two different and highly organized mechanisms of translation initiation in the archaeon Sulfolobus solfataricus. Extremophiles **4:**175–179.
24. **van Regenmortel, M. H. V., C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle, and R. B. Wickner.** 2000. Virus taxonomy: the classification and nomenclature of viruses. Seventh report of the international committee on taxonomy of viruses. Academic Press, San Diego, Calif.
25. **Zillig, W., A. Kletzin, C. Schleper, I. Holz, D. Janekovic, J. Hain, M. Lanzendörfer, and J. K. Kristjansson.** 1994. Screening for Sulfolobales, their plasmids and their viruses in Icelandic solfataras. Syst. Appl. Microbiol. **16:**609–628.
26. **Zillig, W., P. Palm, H.-P. Klenk, D. Langer, U. Hüdepohl, J. Hain, M. Lanzendörfer, and I. Holz.** 1993. Transcription in archaea, p. 367–391. *In* M. Kates, D. J. Kushner, and A. T. Matheson (ed.), The biochemistry of Archaea (Archaebacteria). Elsevier Science Publishers B.V., Amsterdam, The Netherlands.
27. **Zillig, W., K. O. Stetter, and D. Janekovic.** 1979. DNA-dependent RNA polymerase from the archaebacterium Sulfolobus acidocaldarius. Eur. J. Biochem. **96:**597–604.