

T_{lys} , a Newly Identified *Sulfolobus* Spindle-Shaped Virus 1 Transcript Expressed in the Lysogenic State, Encodes a DNA-Binding Protein Interacting at the Promoters of the Early Genes

Salvatore Fusco,^a Qunxin She,^b Simonetta Bartolucci,^a Patrizia Contursi^a

Dipartimento di Biologia, Università degli Studi di Napoli Federico II, Complesso Universitario Monte S. Angelo, Naples, Italy^a; Danish Archaea Centre, Department of Biology, University of Copenhagen, Copenhagen N, Denmark^b

While studying the gene expression of the *Sulfolobus* spindle-shaped virus 1 (SSV1) in *Sulfolobus solfataricus* lysogenic cells, a novel viral transcript (T_{lys}) was identified. Transcriptional analysis revealed that T_{lys} is expressed only in the absence of UV irradiation and is downregulated during the growth of the lysogenic host. The corresponding gene *f55* lies between two transcriptional units (T_6 and T_{ind}) that are upregulated upon UV irradiation. The open reading frame *f55* encodes a 6.3-kDa protein which shows sequence identity with negative regulators that fold into the ribbon-helix-helix DNA-binding motif. DNA-binding assays demonstrated that the recombinant F55, purified from *Escherichia coli*, is indeed a putative transcription factor able to recognize site specifically target sequences in the promoters of the early induced T_5 , T_6 , and T_{ind} transcripts, as well as of its own promoter. Binding sites of F55 are included within a tandem-repeated sequence overlapping the transcription start sites and/or the B recognition element of the pertinent genes. The strongest binding was observed with the promoters of T_5 and T_6 , and an apparent cooperativity in binding was observed with the T_{ind} promoter. Taking together the transcriptional analysis data and the biochemical evidences, we surmise that the protein F55 is involved in the regulation of the lysogenic state of SSV1.

Sulfolobus shibatae B12 isolated from a hot spring in Beppu, Japan (1), is the natural host of the *Sulfolobus* spindle-shaped virus 1 (SSV1). This archaeal virus represents a prototype of nine such viruses belonging to the *Fuselloviridae* family (2, 3). Most of them exhibit a spindle-shaped capsid of ca. 60 by 100 nm with a short tail at one end, but extended morphotypes have also been observed for those with bigger genomes (3). For the past several decades, SSV1 has served as a model for studying archaeal viruses as well as for developing genetic tools (4–7). The envelope of SSV1 virions is composed of three coat proteins (VP1, VP2, and VP3), among which VP2 is directly associated with viral DNA, whereas VP1 and VP3 form the surface structure of the viral particles (8).

Several putative transcription factors encoded by the SSV1 genome, include (i) E51 and C80, which are ribbon-helix-helix proteins (RHH) belonging to the CopG family, and (ii) A45, A79, and B129 containing C_2H_2 zinc finger-like motifs (9). Four putative DNA-binding proteins were structurally and/or functionally characterized, i.e., (i) the fold of D63 resembles that of the “repressor of primer” (ROP) (10), an adaptor protein that regulates colE1 plasmid copy number in *Escherichia coli* (11), (ii) F63 (12) and F112 (13) harbor a winged helix fold, which is a typical feature of proteins that belong to the SlyA (14) and MarR (15) subfamilies of winged-helix DNA-binding proteins, and (iii) E73 which displays the “RH3” domain, a variant of the RHH motif (16). Possibly, these putative transcription factors could play a function in the cascade regulation of SSV1 gene expression induced by UV irradiation. However, it is very unlikely that any of them could have a role in controlling the switch between the viral lysogenic versus UV-inducible life cycle, since they are not expressed during the lysogenic growth (unpublished data).

SSV1 has a narrow host range, i.e., among the three *Sulfolobus* model organisms, including *S. solfataricus*, *S. acidocaldarius*, and *S. islandicus* (17, 18), this virus only infects *S. solfataricus* strains

isolated from the solfataric field of Pisciarelli near Naples (Italy) (19).

Moreover, SSV1 is the only known crenarchaeal virus that exhibits a genomic region similar to that of the bacteriophage lambda and shows an UV-inducible life cycle (20). Instead, the closely related SSV2 shows a growth-phase-dependent induction of its replication (21–23).

Upon infection, the double-stranded SSV1 genome of 15,465 bp integrates into the host chromosome within an arginyl-tRNA gene via the integrase D355 encoded by the viral genome (17). Since attachment sites are located within the *d355* coding sequence, the integration results in partitioning of this gene that inactivates the enzyme, leading to gene capture events in the host chromosome (24, 25).

A previous analysis of the SSV1 transcriptome revealed a well-coordinated temporal expression of nine transcripts in the native host after UV irradiation (26). Subsequently, a more detailed characterization of the UV induction process was conducted using microarray analysis, which largely confirmed the previous results. Upon UV irradiation, the expression of SSV1 genome starts from an UV-inducible immediate-early transcript (T_{ind} , which is located between the two “back-to-back” oriented early transcripts T_5 and T_6). This is then followed by the synthesis of the early (T_5 , T_6 , and T_9), the late ($T_{1/2}$, T_3 , T_x , and $T_{4/7}$) transcripts and finally of the late-extended polycistronic messenger ($T_{4/7/18}$). These transcripts are organized in the viral genome according to the chrono-

Received 15 February 2013 Accepted 6 March 2013

Published ahead of print 20 March 2013

Address correspondence to Patrizia Contursi, contursi@unina.it.

Copyright © 2013, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JVI.00458-13

logical order of their expression; this fashion of regulation is reminiscent of that used by many bacteriophages and eukaryotic viruses (27). The first insight into archaeal transcription was gained by mapping minimal promoter elements and transcription terminators of SSV1 (28, 29).

Despite the extensive characterization of the viral expression upon UV exposure, the molecular components and the mechanisms underpinning the maintenance of the SSV1 lysogenic state are still poorly understood.

We recently investigated the regulation of gene expression in the lysogenic state of this virus by microarrays using as a host SSV1-InF1, an infected *S. solfataricus* P2-derived strain (30). It has been revealed that a region of the SSV1 genome located nearby the T_{ind} transcript, previously considered not transcribed (26), was indeed actively expressed. This led to the identification of a novel uncharacterized transcript, named T_{lys} , which encodes for a 6.3-kDa protein, termed F55. In the present study, the analysis of T_{lys} transcription, as well as the functional characterization of the protein F55, is reported. The results of these analyses provide further insights into the SSV1 life cycle and suggest that F55 might be the regulator of the lysogenic state of this fusellovirus.

MATERIALS AND METHODS

Strains, media, and growth conditions. *S. solfataricus* InF1, a uracil auxotrophic mutant isolated previously (30), was the host for generating an SSV1 lysogenic strain. In brief, 1 to 2 μ l of the *S. shibatae* B12 supernatant containing SSV1 virions was spotted onto the soft layer of a Gelrite plate seeded with uninfected cells of InF1. After 2 to 3 days of incubation at 75°C, turbid halos (plaques) appeared on the plate surface as result of an infection-dependent inhibition of host growth. SSV1-infected cells were extracted from plaques and revitalized in liquid medium. Thereafter, a single colony, herein named SSV1-InF1, was isolated by Gelrite plating and purified by restreaking onto plate for three times. Liquid cultures of *S. solfataricus* were grown aerobically in TYSU medium, a glycine-buffered Brock's basal salt solution, supplemented with 0.1% tryptone, 0.05% yeast extract, 0.2% sucrose (wt/vol), and 0.02 mg of uracil/ml; the pH was finally adjusted to 3.2 with concentrated H_2SO_4 . InF1 and SSV1-InF1 samples from frozen cultures were inoculated into 50 ml of TYSU medium, and culture incubation was conducted in 100- or 250-ml Erlenmeyer flasks with a long neck using an Innova 3100 water bath shaker (New Brunswick Scientific Corp.). The incubation temperature was 75°C with a shaking rate of 150 rpm, and cell growth was monitored spectrophotometrically at 600 nm. When cultures reached the logarithmic phase of growth, they were diluted to an optical density at 600 nm (OD_{600}) of 0.05 in 50 ml of fresh medium and let to grow up to an OD_{600} value of 0.3 to 0.5. Subsequently, cultures were diluted again down to 0.05 OD_{600} and split in three parallel growing cultures that were harvested at three different OD_{600} values: 0.4 (exponential phase), 0.8 (late exponential phase), and 1.2 (early stationary phase). Samples were centrifuged in 50-ml Falcon tubes at $3,000 \times g$ for 10 min using the Centrifuge 5810R (Eppendorf), and pellets were treated for total DNA (using DNeasy tissue kit; Qiagen) and RNA (using TRIzol; Sigma reagent) preparations.

qPCR for determination of SSV1 genome copy number. Two primer couples were designed using Primer3 software, available at the website (<http://frodo.wi.mit.edu/>), in order to amplify: (i) a 155-bp fragment of the SSV1 single-copy gene *vp2* (*vp2*-fw, 5'-TATAAATTGTTATAGACATAGAACGCTGTA-3'; *vp2*-rv, 5'-TTAAATACTTCTTGTGCCGATAGTCC-3') and (ii) a 108-bp region of the host single-copy gene *orc1* (*orc1*-fw, 5'-GGAGGGTACATCGCTACCTTATGA-3'; *orc1*-rv, 5'-CAGTAGGGCTGACAGTAAACTACG-3').

Real-time qPCR amplifications were carried out by means of an iQ5 multicolor real-time PCR detection system (Bio-Rad). The QuantiFast SYBR Green PCR kit (Qiagen) was used for the preparation of the real-

time PCR mixtures. In brief, 12.5 μ l of $2 \times$ QuantiFast SYBR green PCR buffer and primers at 1 μ M concentration were premixed and dispensed into a 96-well plate (Thermowell Gold PCR Plates). Subsequently, appropriated dilutions of standards and samples were added in RNase-free H_2O to reach a final volume of 25 μ l. The thermal cycling protocol was as follows: an initial denaturation step of 5 min at 95°C, followed by 35 cycles of 40 s at 95°C, 40 s at 62°C, and 40 s at 72°C. The fluorescence signal was measured at the end of each extension step. A final step at 72°C has been carried out for 10 min at the end of the 35th cycle. iQ5 optical system software uses the threshold cycle (C_T) value of each amplified sample to calculate the initial template copy number by means of standard curves. Total DNA samples extracted from InF1 cultures were used as negative control for *vp2* amplification.

The setting up of suitable experimental conditions for the absolute quantification was achieved by (i) testing the specificity of the amplified products by melting curves and gel electrophoresis analyses and (ii) constructing standard curves for *orc1* and *vp2* amplicons (data not shown) (31). For this purpose, 10-fold dilutions ranging from 10^8 to 10^2 molecules per μ l were used. The C_T value, which is defined as the cycle number at which the fluorescence generated within a reaction crosses the fluorescence threshold, was measured in duplicate for each dilution, and standard curves were constructed by plotting C_T values against the initial copy number of molecules.

The absolute copy number determination of *orc1* and *vp2* was obtained from two independent experiments, where each experimental point (OD_{600} values of 0.4, 0.8, and 1.2) was analyzed at least in triplicate. The SSV1 genome copy number per host cell was calculated dividing the total copy number of *vp2* by the total copy number of *orc1*. A biological confirmation of the data was achieved by testing the SSV1 genome copy number in total DNA samples derived from a parallel culture of SSV1-InF1.

Northern blot analysis. Total RNA samples (20 μ g) from three different growth phases (see above) were run on a denaturing, formaldehyde-containing 2.0% (wt/vol) agarose gel and then transferred onto a nylon membrane (Hybond-XL; Amersham-Pharmacia). T4 polynucleotide kinase (Fermentas Life Sciences) was used to label 5' ends of single-stranded oligonucleotides $T_{lys,rv}$ (5'-AAGTTCTTCAATGCGTCTTCTGATT-3') and $T_{ind,fw}$ (5'-TCTGAGCTACTAATACTGCTTGAAT-3') with radioactive [γ - 32]ATP, according to the manufacturer's instructions. The primers *Sso2359fw* (5'-AGATGAATGGGTTAATGTT-3') and *Sso2359rv* (5'-CACTAAAACATAAATATCCC-3') were used in PCR amplifications of the *Sso2359*-specific probe from total DNA of InF1. Therefore, the amplicon was 5'-end radiolabeled, purified, and used for hybridization in normalization experiments. The purification of radiolabeled oligonucleotides was achieved by gel filtration chromatography using Illustra Nick columns (Amersham Biosciences). Hybridization with single- and double-stranded DNA probes was carried out as described elsewhere (22). The probes were eventually removed by 10 min of boiling in 0.1% sodium dodecyl sulfate (SDS) to reuse the membrane for subsequent hybridizations. The relative abundance of T_{lys} and of the housekeeping *Sso2359* transcripts was evaluated by quantifying the radioactive signals using a Molecular Dynamics Bio-Rad PhosphorImager (Quantity One software). The size of the T_{lys} mRNA was determined using the RNA molecular weight markers (Roche) as standards.

Primer extension analysis. Total RNA (20 μ g) and 1 pmol (10^5 cpm) of 5'-labeled primer $T_{lys,rv}$ (5'-AAGTTCTTCAATGCGTCTTCTGATTG-3') were coprecipitated by using 2.5 volumes of 96% ethanol for 30 min at $-80^\circ C$. After centrifugation, the pellets were resuspended in the reverse transcription buffer purchased from Ambion and denatured for 3 min at 65°C, frozen in dry ice, and thawed on ice. Reactions were incubated for 30 min at 37°C for the annealing, and then deoxynucleoside triphosphates (final concentration, 2 mM each) and 20 U of RNase inhibitor (Promega) was added. The extension of the primer was performed using 8 U of Moloney murine leukemia virus reverse transcriptase (M-MLV RV; Ambion) for 1 h at 48°C. Sequencing reactions of the corresponding DNA

TABLE 1 Sequences of oligonucleotides used in band-shift assays

Name	Promoter region	Oligonucleotide		
		Strand ^a	Sequence (5'–3')	Length (nt)
T ₆ -D-TR	T ₆	PS	GATATATAGATAGAGTATAGATAGAGTAAAG	31
T ₆ -D-TR	T ₆	MS	CTTACTCTATCTATACTCTATCTATATATC	31
T _{lys} -G-TR	T _{lys}	PS	GTCATATTAATATAGTATAGATAGCGTATGC	31
T _{lys} -G-TR	T _{lys}	MS	GCATACGCTATCTATACTATATTAATATGAC	31
T _{ind} -E-TR	T _{ind}	MS	GTTGTATAAGATACATAAGATACACAG	27
T _{ind} -E-TR	T _{ind}	MS	CTGTGTATCTTATGTATCTTATACAAAC	27
T ₅ -A-SR	T ₅	MS	GATTTATAGATAGAGTGGGA	20
T ₅ -A-SR	T ₅	MS	TCCCACTCTATCTATAAATC	20

^a PS and MS indicate plus and minus strands, respectively.

region, to be used as reference ladders, were carried out with the same primer, T_{lys}rv, and the fmol DNA cycle sequencing system kit (Promega) according to the manufacturer's instructions. The resulting products were separated on denaturing 6% polyacrylamide gels, along with the DNA sequencing reaction products.

Bioinformatic analysis. An open reading frame (ORF) encoding for a 55-amino-acid protein, named F55, was identified on the T_{lys} transcript with the NCBI ORF finder software and the ExPASy translate tool. Similarity comparison of F55 was performed using BLAST (Basic Local Alignment Search Tool). The workbench Jalview 2 (32) was used for multiple sequence alignment comparison and secondary structure prediction. Regions upstream of the transcription start site (TSS) of T_{lys} and downstream of the stop codon were checked in order to identify canonical elements of archaeal promoters and terminators, respectively.

Cloning of the ORF *f55* and overexpression and purification of the recombinant protein. The coding region *f55* was amplified from the SSV1 genome using the primers *f55*-fw (5'-TATAGTATAGATAGCATATGCC GAGGAA-3') and *f55*-rv (5'-GAGAGATAGAG TATACTCGAGCATT AAG-3'), in which restriction sites are underlined. The MinElute PCR purification kit (Qiagen) was used to purify the PCR product according to the manufacturer's instructions. After digestion with NdeI and XhoI (Roche), the amplicon was ligated to the NdeI/XhoI-digested pET30(b), yielding pET-*f55*. The cloned fragment was sequenced to verify its identity. Overexpression of F55 was carried out at 0.6 OD₆₀₀ in *E. coli* BL21-CodonPlus (DE3)-RIL cells upon addition of 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) for 3 h. Cells from 1 liter of culture were harvested by centrifugation, and pellets were resuspended in 20 ml of lysis buffer, 10 mM Tris-HCl (pH 8.0), containing Complete 600 protease inhibitor cocktail tablets (Roche). The cells were lysed by sonication for 5 min, alternating 20 s of pulse-on and 20 s of pulse-off, by means of an ultrasonic liquid processor (Heat System Ultrasonic, Inc.). Lysates were centrifuged at 30,000 × *g* (SW41 rotor; Beckman) for 30 min to clarify the crude extracts. In order to purify the recombinant F55, the lysates were dialyzed overnight against buffer A (50 mM Tris-HCl, 200 mM NaCl [pH 7.0]) and then loaded onto a 1-ml cation-exchange Resource S column (GE Healthcare) connected to a fast-performance liquid chromatography system (AKTA). After a washing step with the buffer A, the elution was performed with a linear salt gradient from 200 to 800 mM NaCl. Protein-containing fractions were collected and analyzed by SDS-PAGE to detect the protein F55 (~6.3 kDa). The F55-containing fractions were pooled, dialyzed against buffer A, and loaded onto a Superdex 75 16/60 column (GE Healthcare) for gel filtration chromatography. To determine the native molecular mass of the protein, the purified F55 at different concentrations (0.5 and 2.0 mg/ml) was applied in a volume of 200 μl to an analytical Superdex PC75 column (3.2 by 30 cm) connected to an AKTA Explorer system (GE Healthcare) equilibrated with buffer A at a flow rate of 0.04 ml/min. The column was calibrated using a set of gel filtration markers (low range; GE Healthcare), including ovalbumin (43.0 kDa), chymotrypsinogen A (25.0 kDa), RNase A (13.7 kDa), and aprotinin (6.5 kDa).

Electrophoresis mobility shift assay (EMSA). Oligonucleotides used in band-shift assays (Table 1) were annealed as follows. Equimolar amounts of plus (PS) and minus (MS) strands were mixed in annealing buffer (10 mM Tris-HCl, 50 mM NaCl, 1 mM EDTA [pH 8.0]), denatured at 95°C for 5 min, and slowly allowed to cool down at room temperature. Double-stranded DNA probes were 5'-end radiolabeled by means of a T4 polynucleotide kinase (Fermentas) in the presence of [γ -³²P]ATP. Purification of double-stranded probes was performed as described above. Thermal preincubation of the purified protein F55 was conducted for 15 min at 50°C in assay buffer (25 mM Tris-HCl, 50 mM KCl, 10 mM MgCl₂, 1 mM dithiothreitol [pH 7.0], 5% glycerol [vol/vol]) supplemented with 500 ng of salmon testes DNA (Sigma) as nonspecific competitor. The labeled probes were added to the binding reactions at a concentration of ~0.03 μM (~20,000 cpm). Binding assays were performed with increasing amounts of F55 (from 1 to 24 μM) for 30 min at 50°C. In the displacement experiments, the binding reactions were performed with 8, 10, or 12 μM F55 with the concurrent addition to the EMSA mixtures of increasing amounts of specific, unlabeled probe (i.e., 1:20, 1:50, 1:100, 1:200, 1:300, 1:400, and 1:500 ratios of labeled/unlabeled specific DNA) or salmon testes DNA as nonspecific competitor (1:50, 1:250, 1:500, 1:750, 1:1,000, 1:1,500, 1:2,000, 1:2,500, and 1:3,000 ratios of labeled specific/nonspecific DNA). Finally, samples were loaded onto a 10% polyacrylamide gel prepared in 0.5× Tris-borate-EDTA and run at 10 mA for ~1 h. Gels were transferred onto filter paper, dried, and revealed by using a Molecular Dynamics Bio-Rad PhosphorImager and/or autoradiography. Shifted signals were quantified by means of the QuantityOne software, and dissociation constants (K_d) were calculated using Prism6 (GraphPad Software) by plotting protein amounts against the percentages of shifted signals (Hill Plot). The K_d is defined as the protein concentration at which 50% of the target sequence is bound by the protein.

Nucleotide and amino acid sequences. Nucleotide and protein sequence data reported are available in the Third Party Annotation Section of the DDBJ/EMBL/GenBank databases under the accession number TPA BK008732.

RESULTS

Physiological characterization of a *S. solfataricus* SSV1 lysogen.

Upon UV irradiation of an SSV1 lysogen, virus DNA replication and virion production are strongly enhanced without apparent lysis of host cells as for most other archaeal viruses (32). The virus is able to rapidly regain lysogenic status in the absence of the UV stimulus (17), indicating that SSV1 has developed a stringent control over the lysogenic growth. To obtain an insight into the virus life cycle, *S. solfataricus* InF1 was infected with SSV1 to yield SSV1-harboring strains. One such strain (SSV1-InF1) was used for further investigations, along with a virus-free culture (InF1). Both cultures grew in a similar fashion (Fig. 1), indicating the establish-

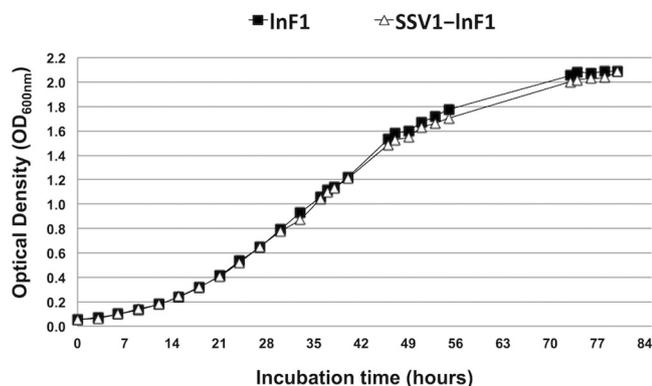


FIG 1 Growth curves of InF1 and SSV1-InF1 strains. The OD₆₀₀ values were measured and plotted versus the incubation time. Both strains grow exponentially until an OD₆₀₀ value of 1.0 and enter into the stationary phase around the 50th hour of incubation. Total RNAs used in the transcriptional analysis (see the Northern blot and primer extension experiments described in Fig. 3) were isolated from cultures harvested at OD₆₀₀ values of 0.4, 0.8, and 1.2.

ment of a harmonious coexistence between host and virus in the lysogenic state.

The copy number of SSV1 genome in the SSV1-InF1 lysogen was estimated by quantitative PCR (qPCR) at different growth phases (OD₆₀₀ values of 0.4, 0.8, and 1.2), through quantification of the amounts of *vp2* and *orc1* genes in the total DNA samples. Since these two genes are present as single copies on viral and host genomes, respectively (24, 33), the ratio between the amount of viral and host genes represents the copy number of SSV1. We found about five copies of SSV1 per chromosome, and this copy number was kept constant during the entire growth (Table 2). Therefore, the variation in transcripts level described below is independent from gene dosage (copy number).

Identification of a novel transcript highly expressed only in the SSV1 lysogenic cells. To gain an insight into the maintenance of the SSV1 lysogenic state, total RNAs were prepared from SSV1-InF1 and InF1 cultures and used for whole-genome microarray analyses. A few viral genes were found to be expressed in the SSV1-harboring cells. These genes could be divided into two categories: (i) those expressed during the entire lysogenic growth, including the ones encoding the structural proteins VP1 and VP3, the integrase D335, A291, and C124, and (ii) those expressed only in the exponential growth phase, i.e., E96, A82, C84/A92, and the capsid-associated VP2 (data not shown).

Whereas the UV-inducible T_{ind} transcript was never detected by Northern blotting and microarray analyses in the lysogenic state (data not shown), a probe derived from the intergenic region of SSV1 located between the promoters of T₆ and T_{ind} (Fig. 2A) showed the presence of a highly expressed transcript, named T_{lys} (“lys” stands for lysogeny), which has never been identified in previous analyses (26, 27). To determine its size, total RNAs were analyzed by Northern blotting with a 5'-radiolabeled single-stranded DNA probe, matching against the T_{lys} transcript. The probe detected a transcript of about 0.25 to 0.30 kb which was constantly expressed during the lysogenic state, albeit its abundance decreased to 20% of the initial amount at a late growth phase (Fig. 3A).

Analyzing the sequence of the intergenic region led to the identification of an ORF (*f55*), which is oriented in the opposite direc-

tion with respect to the ORF *b49* lying on the UV-inducible transcript T_{ind} (27) (Fig. 2B). The name *f55* was chosen according to the annotation method by Palm et al. (33). Putative canonical promoter elements, i.e., TATA box and B recognition element (BRE) (28), as well as transcription terminators (23, 29), were identified upstream and downstream of *f55*, respectively (Fig. 2B).

Since the detected mRNA was ~100 nucleotides (nt) longer than the ORF *f55* (168 nt), we determined by primer extension its transcription start site (TSS), which mapped to a T residue located 10 bp upstream of the putative start codon (Fig. 2B). Transcription termination signals are located between 80 and 120 bp downstream of the stop codon. Therefore, the deduced size of T_{lys} was ~300 nt, which is in good agreement with the results of Northern analysis. It must be pointed out the T_{lys} was undetectable upon UV irradiation (26), indicating that its role is specifically related to the lysogenic state of SSV1.

T_{lys} encodes a putative RHH transcription factor that binds to its own promoter. The F55 protein exhibits 40 to 50% sequence identity to several transcription regulators containing the RHH motif and belonging either to the NikR (nickel responsive) (34) or CopG (35) families (Fig. 4). This suggested that F55 could function as a transcription factor as well.

To study this, the gene was cloned into pET30(b) giving pET-*f55* with which the recombinant F55 was expressed in *E. coli* BL21-CodonPlus (DE3)-RIL cells. After purification to homogeneity via two-step procedure, i.e., cation-exchange and gel filtration chromatographies, the purified protein was analyzed by SDS-PAGE. A single band with an apparent molecular mass of ~6.3 kDa appeared (Fig. 5A); this is in good agreement with its predicted molecular mass (6,261 Da). Analytic size-exclusion chromatography of the recombinant protein revealed a size of ~12.4 kDa, indicating that F55 forms a dimer in solution (Fig. 5B). The protein was also found to form dimers in a multi-angle light scattering analysis (data not shown) using a MiniDAWN Treos light-scattering system (Wyatt Technology) as previously described (36).

In the first attempt to study the propensity of F55 to specifically bind to DNA, EMSAs were performed with a 93-bp double-stranded DNA fragment of the predicted promoter region of T_{lys}. The purified protein retarded the migration of the labeled DNA substrate (data not shown), suggesting that the promoter of T_{lys} contains binding motifs recognized by F55. An imperfect tandem-repeated sequence of 22 nt, herein named T_{lys}-G-TR (TR, i.e., tandem repeat), was found overlapping the TSS of T_{lys} (Fig. 6A). The labeled DNA fragment was shifted up when incubated with

TABLE 2 SSV1 copy number estimation by absolute quantification method

Culture ^a	SSV1 copy no. (copies/cell)		
	OD ₆₀₀ 0.4	OD ₆₀₀ 0.8	OD ₆₀₀ 1.2
1	6.1	6.7	5.4
2	5.4	4.8	5.3
3	4.6	4.9	6.5
Avg ± SD (n = 3)	5.4 ± 0.61	5.5 ± 0.87	5.8 ± 0.54

^a In rows 2 and 3 are listed the results obtained from two independent qPCR experiments on the same total DNA samples, whereas row 3 corresponds to results from another culture of the SSV1-InF1 strain (biological confirmation of the data). The average along with the corresponding standard deviation (SD) is reported in the last row.

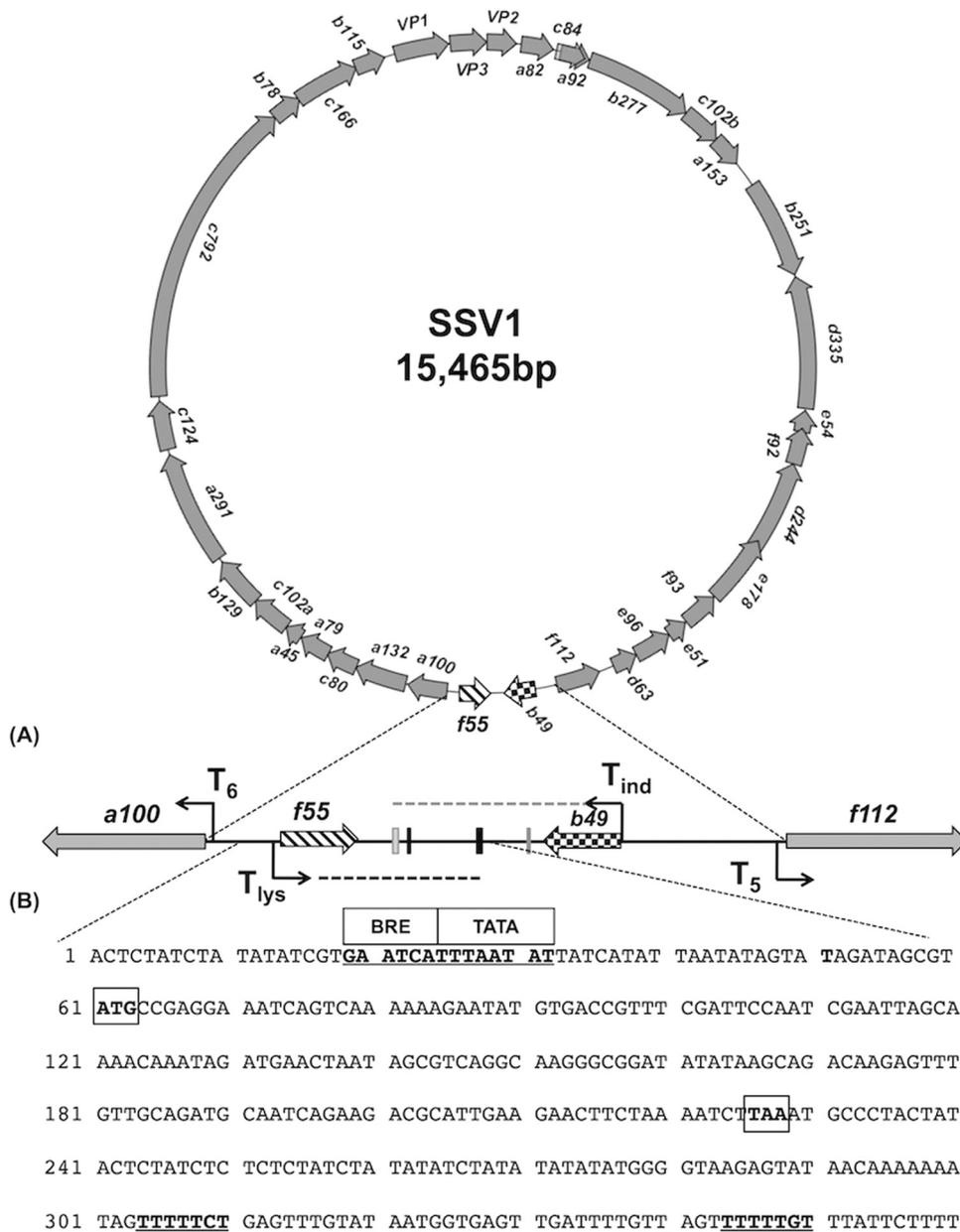


FIG 2 Scheme of the genomic region critical for the UV induction of SSV1. (A) The SSV1 genome map is schematized. ORFs lying on the plus and minus strands are clockwise and anticlockwise oriented. The head-to-head oriented ORFs *b49* (T_{ind}) and *f55* (T_{lys}) are checked and striped, respectively. The 840-bp genomic region, which is included between the transcription start sites of the early transcripts T₆ and T₅, is shown. The TSSs are indicated as thin black bent arrows. Transcripts T_{ind} and T_{lys} are represented as gray and black dashed lines, respectively. The previously mapped termination signals of T_{ind} and the putative ones of T_{lys} are indicated by gray and black bars. (B) The 5'-3' nucleotide sequence of *f55*, as well as the immediately upstream and downstream regions, are shown. The TATA box and the BRE element are centered at -23 nt and -28 nt, respectively. Start and stop codons are in boldface and boxed, while predicted transcription termination signals are in boldface and underlined.

F55, producing two distinct DNA-protein complexes (A1 and A2). The band intensity was found to be proportional to the amounts of protein used in the reactions (from 3 to 16 μ M), suggesting that there were two F55 binding sites in the probe (Fig. 6B).

In order to determine the F55 dissociation constant (K_d) of the DNA-protein complexes, the relative radioactive signals were quantified, and data were used to generate binding curves (Hill plots). We found that the K_d was 5.2 μ M for the faster-migrating

complex A1 (K_{dA1}) and 9.3 μ M for the slower-migrating complex A2 (K_{dA2}) (Fig. 6B). A third signal A3 was very weak in intensity and only detectable at a high concentration of F55 (8 to 16 μ M), which more likely represented unspecific DNA-protein aggregates. Consequently, we only studied the main A1 and A2 complexes and their equivalents further.

The A1 and A2 complexes exhibited different stability as confirmed by specific competition assays that abolished first the signal of A2 and then that of A1 (Fig. 6B). The shifts almost completely

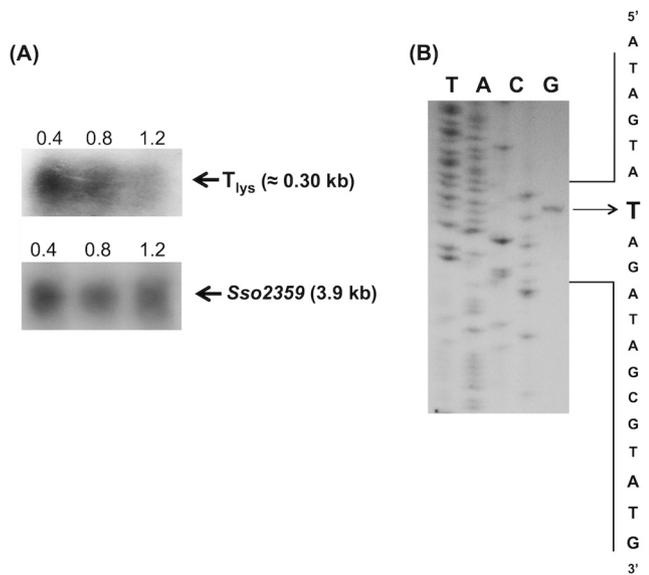


FIG 3 Transcription analysis of T_{lys} . (A) SSV1-InF1 total RNAs isolated at three different growth stages ($OD_{600} = 0.4, 0.8,$ and 1.2) that highlight the variation of gene expression at the early exponential, late exponential, and stationary phases of growth, respectively, were analyzed by Northern blotting and primer extension experiments. The oligonucleotides used for the analysis of T_{lys} transcript is $T_{lys}Fv$ (see Material and Methods). The cDNA products were electrophoresed with the sequence ladder generated by the same primer on the noncoding strand of the $f55$ gene. The mapped TSS is indicated by the arrow. Detection of the housekeeping gene $Sso2359$ was performed as described above.

disappeared at 500 molar excess of the cold DNA, as judged by quantification of the free probe signal (>90% of the initial amount).

The protein F55 binds to several operator sequences located in the promoters of the early transcripts. The high expression level of T_{lys} in the SSV1-InF1 lysogenic strain suggested that it could play a key role in controlling SSV1 lysogeny. To yield an insight into this hypothesis, the whole SSV1 genome was scanned to identify tandem-repeated sequences similar to those present in the T_{lys} promoter. This element was identified in three other promoters, i.e., of the early transcripts T_5 and T_6 and of T_{ind} , a transcript that is highly induced upon UV irradiation. Interestingly, all of these repeated sequences are found within a small region (~840

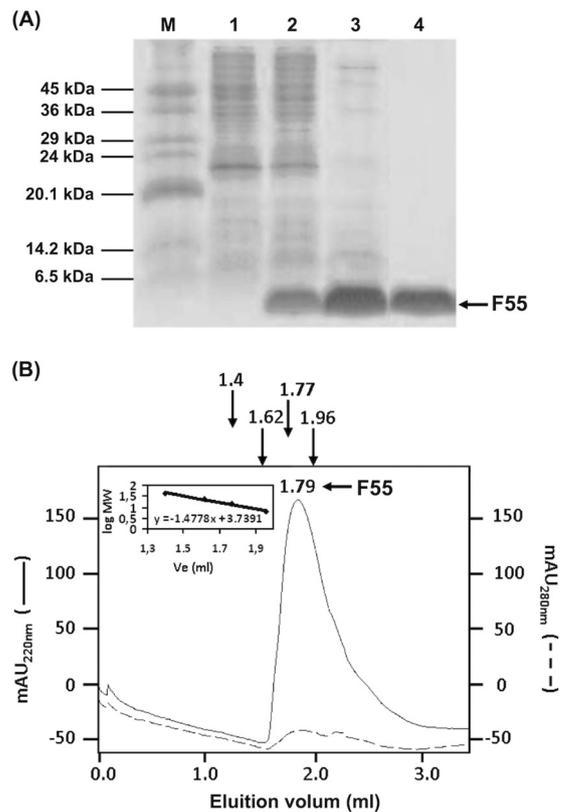


FIG 5 SDS-PAGE and quaternary structure analyses of the recombinant F55. (A) SDS-PAGE of protein extracts at each step of the protein purification. Lane M, molecular-mass markers; lane 1, crude extract from not induced cells; lane 2, crude extract from cells induced with 1 mM IPTG; lanes 3 and 4, samples after cationic exchange and gel filtration chromatography. (B) Elution profile from gel filtration chromatography on a Sephadex PC75 16/60 column. The elution volume of F55 (1.79 ml) corresponds to a molecular mass of 12.4 kDa. Arrows indicate the elution volumes of the protein standards in the relative calibration of the column.

bp) of the SSV1 genome that is included between the promoter sequences of transcripts T_5 and T_6 (Fig. 6A).

Two copies of the tandem-repeated element were found in all of these promoters. T_5 -A-TR and T_6 -C-TR encompass the TSS of T_5 and T_6 , respectively. Instead, T_5 -B-TR and T_6 -D-TR lie immediately upstream of the core promoters of these transcripts and

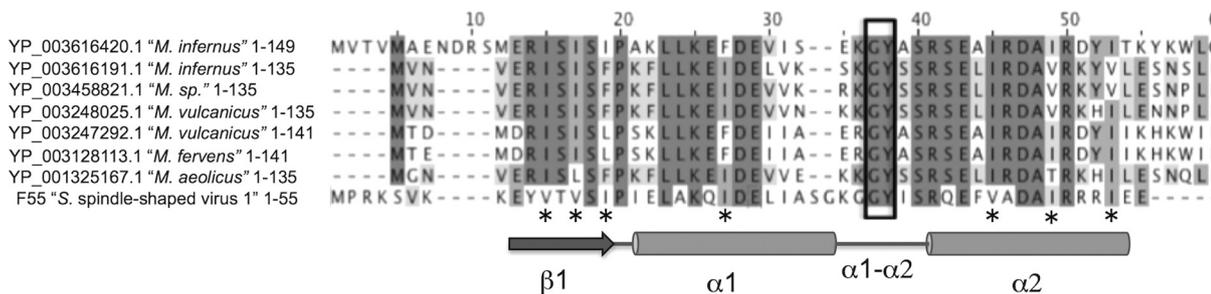


FIG 4 Multiple sequence alignment and secondary structure prediction of F55. The sequence of F55 aligns with the N termini of NikR proteins encoded by the genome of *Methanocaldococcus* spp. and *Methanococcus* spp. Residues conserved in all of the sequences analyzed are shaded in dark gray. The dipeptide Gly-Tyr (GY), typically located between helices $\alpha 1$ and $\alpha 2$ in the RHH proteins, is boxed. The intensity of the gray color decreases proportional to the percentage of conservation. Asterisks indicate conserved hydrophobic, primarily branched-chain amino acids. The secondary structure prediction based on Jpred algorithm, is also schematized.

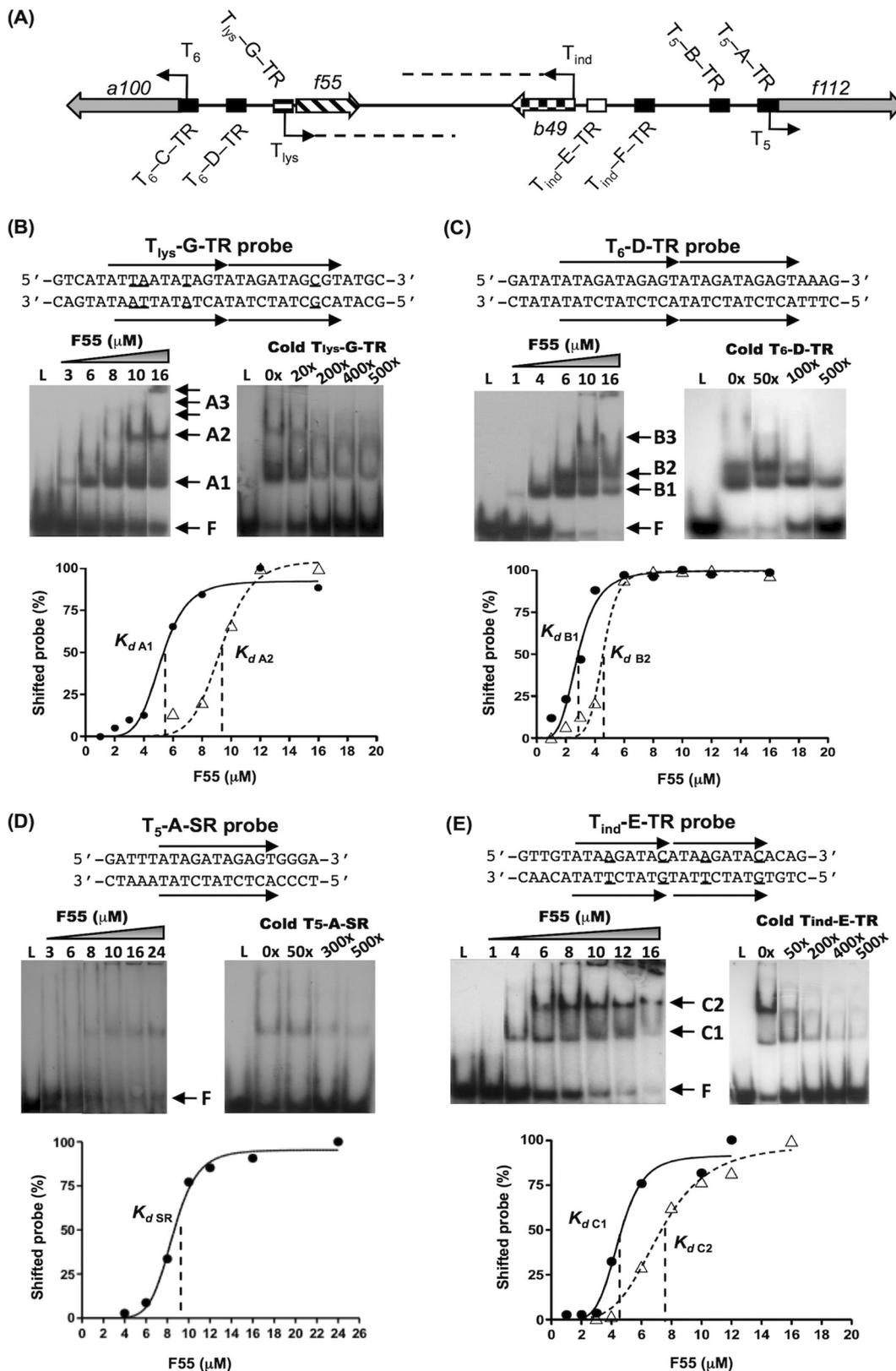


FIG 6 Analysis of the F55 interaction with operator sequences localized in the key regulative region of SSV1. (A) Graphic representation of the region included between the TSS of the early transcripts T₅ and T₆. ORFs are represented as straight arrows, and TSSs are indicated as thin black bent arrows. Tandem-repeated sequences (TR) are reported as black, white, gray, or striped boxes located upstream of the relative ORFs. (B to E) Band-shift assays results. For each experiment, the name and the sequence of the probe tested are indicated. Black arrows highlight the tandem-repeated sequences in which underlined letters indicate the mismatches compared to the consensus sequence. The two panels indicate the EMSAs performed with increasing concentrations of F55 (left) and with increasing concentrations of the specific cold probe (right), whose molar excess is reported on the top. Different DNA-protein complexes are indicated, and “F” means free probe. Binding curves relative to the formation of the DNA-protein complexes are reported. Densitometric data from EMSA obtained as described in Materials and Methods are plotted versus the concentration of F55.

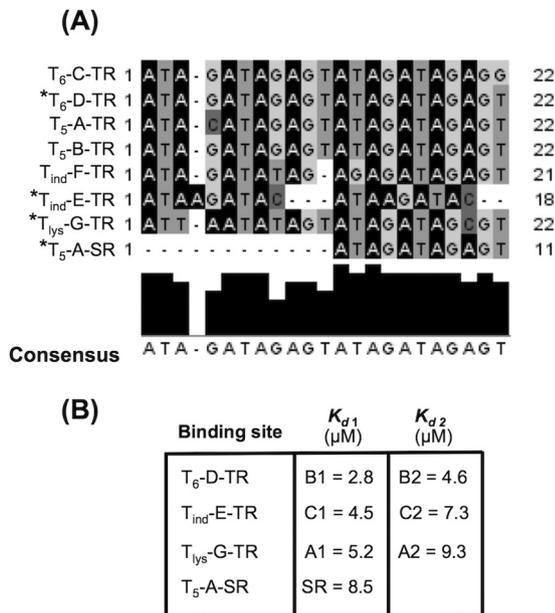


FIG 7 Sequence alignment of the F55 binding sites and comparison of their dissociation constants. (A) F55 binding sites, the relative sequences, and the lengths are shown. Asterisks indicate the sequences tested in EMSAs. Nucleotides conserved in a specific position are highlighted in the same color, while substitutions are shown in a different color. The height of the black histograms is proportional to the percentage of the conservation of each nucleotide position; the consensus sequence is shown. (B) For the tested sequences, the K_d values relative to the faster-migrating (A1, B1, and C1) and slower-migrating (A2, B2, and C2) complexes are reported.

partially overlap their BRE elements (Fig. 6A). Despite the absence of canonical basal promoter elements (TATA box and BRE), the T_{ind} promoter region contains two copies of the tandem-repeated sequence that are similarly spaced as those of T_5 and T_6 . The multiple alignment of these sequences (Fig. 7A) allowed the identification of a 22-bp consensus element (5'-ATAGATAGAGTATAGATAGAGT-3').

When T_6 -D-TR was tested, two differently migrating complexes were identified. The complex B1 appeared to precede B2, indicating that the formation of B2 is a multistep process (Fig. 6C). The K_d values of the complexes ($K_{d_{B1}} = 2.8 \mu\text{M}$ and $K_{d_{B2}} = 4.6 \mu\text{M}$), indicated a better affinity of F55 for T_6 -D-TR than for T_{lys} -G-TR (Fig. 7B). The complexes B1 and B2 displayed different stability ($B2 < B1$) as inferred by displacement experiments with specific cold DNA (Fig. 6C). Comparing the results obtained with T_{lys} -G-TR (complexes A1 and A2), the faster-migrating signal B1, although decreased upon addition of cold specific DNA, still persists at 500 molar excess, confirming that the interaction between F55 and T_6 -D-TR is stronger than that with T_{lys} -G-TR (Fig. 6B and C).

We also tested whether F55 could bind to a single consensus module (5'-ATAGATAGAGT-3') of the tandem repeat (Fig. 7A). With this aim, the 20-bp probe T_5 -A-SR (SR, i.e., single repeat) was labeled and used in band-shift assays as described above (Fig. 6D). Only one shifted signal was detected, indicating that one binding site of F55 is included in the single module. The construction of a binding curve revealed a higher dissociation constant ($K_{d_{SR}} = 8.5 \mu\text{M}$; Fig. 7B); therefore, the DNA-protein complex formed with a single module is less stable than the complexes formed with a dual module. Furthermore, specific displacement

experiments indicate that F55 specifically binds to the sequence 5'-ATAGATAGAGT-3' (Fig. 6D), which is present in slightly different variants in every promoter tested (Fig. 7A). According to these EMSA data, there is a correlation between the presence of mismatches in the tandem-repeated sequence (as in T_{lys} -G-TR) and the decreased affinity of F55 toward its targets.

To better define the target sequence of F55, T_{ind} -E-TR was used as a DNA substrate. A similar binding pattern with two shifted bands (C1 and C2) was observed (Fig. 6E). Further, $K_{d_{C1}}$ and $K_{d_{C2}}$ values were in the same range as those determined for T_{lys} -G-TR and T_6 -D-TR (Fig. 7B). Interestingly, the formation of the complex C2 occurred at a F55 concentration that was lower than the $K_{d_{C1}}$ (4.5 μM) (Fig. 6E, dashed curve), which contrasted to the results obtained with other probes. Indeed, for all of the other sequences tested, the appearance of the slowest-migrating complexes A2 and B2 (Fig. 6B and C, dashed curves) occurred as soon as the F55 concentration reached the K_d values of A1 and B1, respectively (Fig. 6B and C, solid curves). A possible explanation for this difference is that the two modules of the tandem-repeated sequence are closer to each other, thus influencing positively the interaction of F55 to the binding sites of T_{ind} -E-TR (Fig. 7A).

Displacement experiments performed with salmon testes DNA (data not shown) confirmed that F55 exhibited the highest binding specificity toward T_6 -D-TR, which contains the perfect tandem-repeated sequence (Fig. 7A). However, under our experimental conditions F55 showed a not-pronounced difference in the apparent affinity to specific and nonspecific DNA, in accordance to what already reported for other RHH-containing transcription regulators, such as Lrp from *S. folfataricus* (37) and SvtR from virus SIRV1 (38).

Taken together, these findings show that F55 interacts with differential affinity to all of the identified repeated sequences in the early UV-inducible region of the SSV1 genome.

DISCUSSION

In this study, we have detected a transcriptional activity within the intergenic region ranging from the T_6 to the T_{ind} promoters. This novel transcript contains an ORF, which encodes a putative transcription factor (F55). This gene is highly expressed in the SSV1-InF1 strain, only in the absence of UV stimulus (lysogenic state), suggesting a key role for F55 in the SSV1 lysogeny. Intriguingly, the protein F55 binds specifically to a direct-repeated motif that recurs in its own promoter, as well as in the promoters of the early T_5 and T_6 (26, 27, 39), and of the UV-inducible T_{ind} transcripts.

While the UV induction of SSV1 has been characterized in the native and foreign hosts at physiological and transcriptional levels (17, 18, 26–29), the lysogenic growth of the virus has never been studied extensively. Nevertheless, our experimental evidences, such as the stringent control of the SSV1 copy number during the lysogenic growth, suggested the occurrence of a tight regulation of the SSV1 replication in the absence of UV stimulus. Therefore, we resolved to unravel the molecular components underpinning the maintenance of the lysogenic state.

The *f55* gene was first identified here. There are two reasons that could account for the delayed discovery of this gene: (i) when UV-inducible expression was investigated in the natural host, the expression of T_{lys} was inhibited (26), and (ii) in the microarray analysis by Fröls et al., no probe was designed for the above-mentioned intergenic region (27).

The primary structure of F55 shares a good degree of sequence

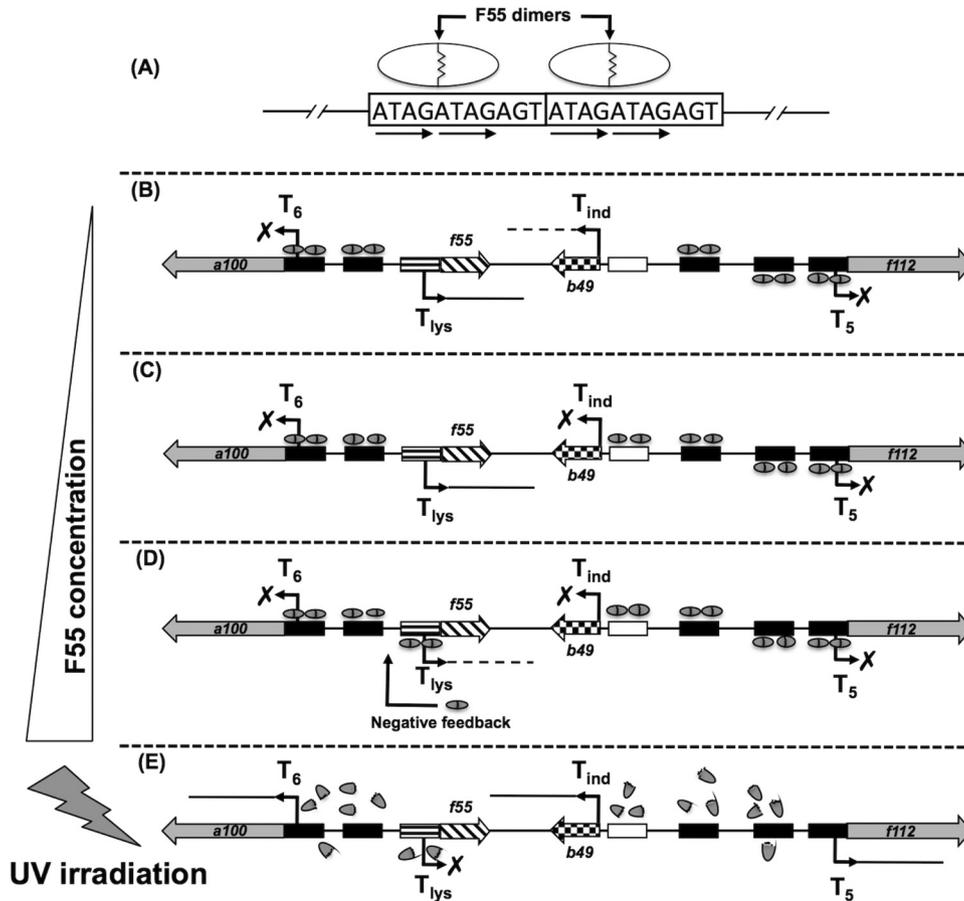


FIG 8 Model of the F55 interaction at its binding sites in the key regulative region. (A) The consensus tandem-repeated sequences is framed. Black arrows indicate the adjacent binding subsites (5'-ATAG-3'). Each F55 dimer (white ovals) binds to two adjacent subsites (5'-ATAG-3') within a single repeat. (B to D) Progressive concentration-dependent saturation of the F55 binding sites as inferred from their K_d values. (B) At low concentrations, F55 binds to two tandem repeats of T_5 and T_6 , as well as to one of the T_{ind} binding sites (black boxes). (C and D) As the F55 concentration increases, the second T_{ind} binding site (white box) and that of T_{lys} are saturated sequentially. ORFs and TSSs are indicated as large solid arrows and thin black bent arrows, respectively. The interaction of F55 at its binding sites causes complete transcriptional abrogation (black crosses) or downregulation (dashed lines starting from TSS). In this latter case (D), a negative-feedback regulation mechanism operates on the T_{lys} expression. (E) Upon UV irradiation, F55 is degraded and/or inactivated and, in turn, the transcription of the early induced T_{ind} , T_5 , and T_6 transcripts is unlocked.

identity with several transcription repressors such as those belonging to NikR or CopG families. These regulators form a dimeric DNA-binding domain constituted by a pair of antiparallel β -strands framed within an α -helical scaffold (34, 35). Accordingly, our secondary structure prediction indicated that F55 also folds into the RHH motif and size exclusion chromatography demonstrated that it exists as dimer in solution.

Some examples of archaeal RHH-containing proteins have been reported and are: (i) ORF56, encoded by the *S. islandicus* pRN1 plasmid (40), (ii) the regulator SvtR encoded by the rudi-virus SIRV1 (38), (iii) the protein E73 encoded by the virus SSV-RH (16), and (iv) the recently described AvtR encoded by the lipothrixvirus AFV6 (41). Typically, multiple RHH dimers bind to an array of regularly spaced inverted or tandem repeats in the promoters of pertinent gene(s), which include their own genes, to exert transcriptional repression (42).

Here we show that the dimeric F55 binds specifically to an 11-nt sequence (5'-ATAGATAGAGT-3'), which is the minimal binding site, leading to the formation of a stable DNA-protein complex. This sequence motif contains in turn a tandem-repeated

element which fits well with the 2-fold symmetry of the dimeric (RHH)₂ domain that requires two adjacent subsites (5'-ATAG-3' in this case) to bind the DNA (Fig. 8A).

In the hypothesis that F55 were a repressor involved in the maintenance of the lysogenic state by downregulating the transcription of the early induced transcripts (T_5 , T_6 , and T_{ind}), as well as its own expression, we tested the ability of F55 to bind to the relative promoter sequences.

While the promoters of T_5 and T_6 and of the UV-inducible T_{ind} transcripts bear four F55 binding sites arranged as two distinct 22-nt tandem-repeated sequences, the promoter of T_{lys} carries only one element overlapping its own TSS. Since the F55 binding sites present in promoters of T_5 (T_5 -A-TR and T_5 -B-TR), T_6 (T_6 -C-TR and T_6 -D-TR) and of T_{ind} (T_{ind} -F-TR) are overall identical to one another (Fig. 8B), we resolved to analyze only T_6 -D-TR. The recombinant F55 interacts more strongly to the above-mentioned binding sites if compared to that of T_{ind} (T_{ind} -E-TR) and of its own promoter (T_{lys} -G-TR). This is in agreement with the fact that the former binding sites are very similar to the consensus sequence; instead, T_{ind} -E-TR and T_{lys} -G-TR differ significantly.

The substantial divergence of T_{ind} -E-TR from the consensus provides a molecular basis for the apparent cooperativity in the binding observed specifically after quantification of the relative EMSA signals. This DNA-binding mode is in accordance with the fact that T_{ind} is the first transcript to be produced upon the onset of the UV response and therefore transcriptional regulation at the T_{ind} promoter needs to be finely tuned and quickly reverted.

As in *Bacteria*, the placement of the binding sites of transcription repressors relative to promoter elements is also the primary determinant affecting transcription initiation in *Archaea* (43). The location of the F55 binding sites overlapping the TSS and the BRE element suggests a negative regulatory role of these *cis*-acting elements recognizable by transcription repressor(s) (44). Indeed, analogous to previously characterized archaeal negative regulators such as MDR1 from *Archaeoglobus fulgidus* (45) or LrpA from *Pyrococcus furiosus* (46), F55 may impinge on the activity of the transcription machinery through binding to operator sequences that overlap the transcription start site, thus abrogating RNAP (RNA polymerase) recruitment to the promoter and preventing initial steps of RNA chain elongation. Alternatively, by interacting to operator sequences that overlap the BRE element, F55 might convey its repressive effect on transcription initiation by hindering an earlier step, i.e., the formation of the ternary TBP-TFB-DNA PIC (preinitiation complex). This kind of transcriptional repression mode has been already described for the *Sulfolobus solfataricus* Lrs-14 protein (47) and for the *Thermococcus litoralis* TrmB transcription factor (48).

In vivo analysis indicates that the ORF *f55* is strongly down-regulated during the SSV1 life cycle and reaches its lowest expression level in the late-stationary phase, thus suggesting that the protein F55 is a transcription repressor that modulates its own expression. This strategy allows the cell to avoid overproducing the repressor protein and to keep its functional concentration.

The newly identified *f55* gene is located in an SSV1 genomic region which resembles that of the early induced UV genes of bacteriophage lambda. In this case the intergenic region, located between the early genes, encodes the *cI* repressor protein. This regulator simultaneously represses transcription of the early induced genes by binding to specific operator regions, thus allowing the maintenance of the lysogenic state to be governed by *cI* alone.

The lysogenic growth of lambda requires that this repressor binds to all three operators at the left and right promoters, and its oligomerization on the target sites shut off the promoters completely (20). We hypothesize that F55 controls the SSV1 lysogeny in a similar fashion. Although our data do not indicate whether F55 exerts its physiological function through binding to one or both tandem-repeated sequences within the same promoter, oligomerization most likely starts from the most affine primary sites and subsequently extend to the entire region in a concentration-dependent manner. Therefore, according to its differential affinity toward the binding sites tested (Fig. 7B), F55 would bind first to the sequences in the promoters of T_5 , T_6 , and T_{ind} transcripts (Fig. 8B, black boxes) and then extend to the less affine sequence of T_{ind} (Fig. 8C, white box) and finally to its own promoter (Fig. 8D, striped box), possibly repressing the expression of the corresponding transcripts, sequentially. Saturation of all of the binding sites may serve to enhance transcription repression.

By analogy to the lambda phage (20), it is tempting to speculate that F55 is degraded and/or inactivated upon UV irradiation in a

fashion similar to *cI*, thus unblocking the transcriptional circuit of the early genes (Fig. 8E).

In contrast to a lambda lysogen, in which episomic DNA is completely absent, both episomic and integrated forms of SSV1 exist in SSV1 lysogenic cells (17). This indicates that SSV1 has to express a minimal set of proteins responsible for a low level of DNA replication and/or for the maintenance of the lysogenic state. The absence of F55 binding sites in the promoters of the T_x , T_3 , and T_9 transcripts, which are expressed together with T_{lys} during the lysogenic growth (data not shown), further support the hypothesis that F55 functions as a regulator only toward the early UV-inducible transcripts.

ACKNOWLEDGMENTS

We thank the *Sulfolobus* Gene Chip Consortium coordinated by John van der Oost for constructing the microarrays used in this study and Luciano Pirone for performing the light-scattering analysis of F55. We are grateful to Gabriella Fiorentino for helpful scientific discussions.

This study was supported by Danish grant 11-106683 from Danish Independent Research Council–Technology and Production Sciences and from Ministero dell’Istruzione, dell’Università e della Ricerca Scientifica (Progetti di Ricerca di Interesse Nazionale E61J10000020001).

REFERENCES

- Martin A, Yeats S, Janekovic D, Reiter WD, Aicher W, Zillig W. 1984. SAV1, a temperate UV-inducible DNA virus-like particle from the archaeobacterium *Sulfolobus acidocaldarius* isolate B12. *EMBO J.* 3:2165–2168.
- Wiedenheft B, Stedman KM, Roberto F, Willits D, Gleske AK, Zoeller L, Snyder J, Douglas T, Young M. 2004. Comparative genomic analysis of hyperthermophilic archaeal *Fuselloviridae* viruses. *J. Virol.* 78:1954–1961.
- Redder P, Peng X, Brugger K, Shah SA, Roesch F, Greve B, She Q, Schleper C, Forterre P, Garrett RA, Prangishvili D. 2009. Four newly isolated fuselloviruses from extreme geothermal environments reveal unusual morphologies and a possible interviral recombination mechanism. *Environ. Microbiol.* 11:2849–2862.
- Stedman KM, Schleper C, Rumpf E, Zillig W. 1999. Genetic requirements for the function of the archaeal virus SSV1 in *Sulfolobus solfataricus*: construction and testing of viral shuttle vectors. *Genetics* 152:1397–1405.
- Cannio R, Contursi P, Rossi M, Bartolucci S. 1998. An autonomously replicating transforming vector for *Sulfolobus solfataricus*. *J. Bacteriol.* 180:3237–3240.
- Contursi P, Cannio R, Prato S, Fiorentino G, Rossi M, Bartolucci S. 2003. Development of a genetic system for hyperthermophilic *Archaea*: expression of a moderate thermophilic bacterial alcohol dehydrogenase gene in *Sulfolobus solfataricus*. *FEMS Microbiol. Lett.* 218:115–120.
- Jonuscheit M, Martusewitsch E, Stedman KM, Schleper C. 2003. A reporter gene system for the hyperthermophilic archaeon *Sulfolobus solfataricus* based on a selectable and integrative shuttle vector. *Mol. Microbiol.* 48:1241–1252.
- Reiter WD, Palm P, Henschen A, Lottspeich F, Zillig W, Grampp B. 1987. Identification and characterization of the genes encoding three structural proteins of the *Sulfolobus* virus-like particle SSV1. *Mol. Genet. Genomics* 206:144–153.
- Prangishvili D, Garrett RA, Koonin EV. 2006. Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. *Virus Res.* 117:52–67.
- Kraft P, Kummel D, Oeckinghaus A, Gauss GH, Wiedenheft B, Young M, Lawrence CM. 2004. Structure of D-63 from *Sulfolobus* spindle shaped virus 1: surface properties of the dimeric four-helix bundle suggest an adaptor protein function. *J. Virol.* 78:7438–7442.
- Helmer-Citterich M, Anceschi MM, Banner DW, Cesareni G. 1988. Control of ColE1 replication: low affinity specific binding of Rop (Rom) to RNAI and RNAII. *EMBO J.* 7:557–566.
- Kraft P, Oeckinghaus A, Kummel D, Gauss GH, Gilmore J, Wiedenheft B, Young M, Lawrence CM. 2004. Crystal structure of F-93 from *Sulfolobus* spindle-shaped virus 1, a winged-helix DNA-binding protein. *J. Virol.* 78:11544–11550.

13. Menon SK, Maaty WS, Corn GJ, Kwok SC, Eilers BJ, Kraft P, Gillitzer E, Young MJ, Bothner B, Lawrence CM. 2008. Cysteine usage in *Sulfolobus* spindle-shaped virus 1 and extension to hyperthermophilic viruses in general. *Virology* 376:270–278.
14. Wu RY, Zhang RG, Zagnitko O, Dementieva I, Maltzev N, Watson JD, Laskowski R, Gornicki P, Joachimiak A. 2003. Crystal structure of *Enterococcus faecalis* SlyA-like transcriptional factor. *J. Biol. Chem.* 278: 20240–20244.
15. Alekshun MN, Levy SB, Mealy TR, Seaton BA, Head JF. 2001. The crystal structure of MarR a regulator of multiple antibiotic resistance, at 2.3-Å resolution. *Nat. Struct. Biol.* 8:710–714.
16. Schlenker C, Goel A, Triplet BP, Menon S, Willi T, Dlakić M, Young MJ, Lawrence CM, Copié V. 2012. Structural studies of E73 from a hyperthermophilic archaeal virus identify the “RH3” domain, an elaborated ribbon-helix-helix motif involved in DNA recognition. *Biochemistry* 51: 2899–2910.
17. Schleper C, Kubo K, Zillig W. 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. U. S. A.* 89:7645–7649.
18. Ceballos RM, Marceau CD, Marceau JO, Morris S, Clore AJ, Stedman KM. 2012. Differential virus host-ranges of the *Fuselloviridae* of hyperthermophilic *Archaea*: implications for evolution in extreme environments. *Front. Microbiol.* 3:295 doi:10.3389/fmicb.2012.00295.
19. Zillig W, Stetter KO, Wunderl S, Schulz W, Priess H, Scholz I. 1980. The *Sulfolobus*-“*Caldariella*” group: taxonomy on the basis of the structure of DNA-dependent RNA polymerases. *Arch. Microbiol.* 125:259–269.
20. Oppenheim AB, Kobiler O, Stavans J, Court DL, Adhya S. 2005. Switches in bacteriophage lambda development. *Annu. Rev. Genet.* 39: 409–429.
21. Contursi P, Jensen J, Aucelli T, Rossi M, Bartolucci S, She Q. 2006. Characterization of the *Sulfolobus* host-SSV2 virus interaction. *Extremophiles* 10:615–627.
22. Contursi P, Cannio R, Prato S, She Q, Rossi M, Bartolucci S. 2007. Transcriptional analysis of the genetic element pSSVx: differential and temporal regulation of gene expression reveals correlation between transcription and replication. *J. Bacteriol.* 17:6339–6350.
23. Contursi P, Cannio R, She Q. 2010. Transcription termination in the plasmid/virus hybrid pSSVx from *Sulfolobus islandicus*. *Extremophiles* 14: 453–463.
24. She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ, Chan-Weihere CCY, Clausen IG, Curtis BA, De Moors A, Erauso G, Fletcher C, Gordon PMK, Heikamp-de Jong I, Jeffries AC, Kozera CJ, Medina N, Peng X, Thi-Ngoc HP, Redder P, Schenk ME, Theriault C, Tolstrup N, Charlebois RL, Doolittle WF, Duguet M, Gaasterland T, Garrett RA, Ragan MA, Sensen CW, Van der Oost J. 2001. The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. U. S. A.* 98:7835–7840.
25. She Q, Brugger K, Chen L. 2002. Archaeal integrative elements and their impact on evolution. *Res. Microbiol.* 153:325–332.
26. Reiter WD, Palm P, Yeats S, Zillig W. 1987. Gene expression in archaeobacteria: physical mapping of constitutive and UV-inducible transcripts from the *Sulfolobus* virus-like particle SSV1. *Mol. Gen. Genet.* 209:270–275.
27. Fröls S, Gordon PM, Panlilio MA, Schleper C, Sensen CW. 2007. Elucidating the transcription cycle of the UV-inducible hyperthermophilic archaeal virus SSV1 by DNA microarrays. *Virology* 365:48–59.
28. Reiter WD, Palm P, Zillig W. 1988. Analysis of transcription in the archaeobacterium *Sulfolobus* indicates that archaeobacterial promoters are homologous to eukaryotic pol II promoters. *Nucleic Acids Res.* 16:1–19.
29. Reiter WD, Palm P, Zillig W. 1988. Transcription termination in the archaeobacterium *Sulfolobus*: signal structures and linkage to transcription initiation. *Nucleic Acids Res.* 16:2445–2459.
30. Gudbergdottir S, Deng L, Chen Z, Jensen JV, Jensen LR, She Q, Garrett RA. 2011. Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.* 79:35–49.
31. Lee C, Kim J, Shin SG, Hwang S. 2006. Absolute and relative QPCR quantification of plasmid copy number in *Escherichia coli*. *J. Biotechnol.* 123:273–280.
32. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191.
- 32a. Prangishvili D, Forterre P, Garrett RA. 2006. Viruses of the *Archaea*: a unifying view. *Nat. Rev. Microb.* 4:837–848.
33. Palm P, Schleper C, Grampp B, Yeats S, McWilliam P, Reiter WD, Zillig W. 1991. Complete nucleotide sequence of the virus SSV1 of the archaeobacterium *Sulfolobus shibatae*. *Virology* 185:242–250.
34. Chivers PT, Sauer RT. 2000. Regulation of high-affinity nickel uptake in bacteria. *J. Biol. Chem.* 275:19735–19741.
35. Schreiter ER, Drennan CL. 2007. Ribbon-helix-helix transcription factors: variations on a theme. *Nat. Rev. Microbiol.* 5:710–720.
36. Contursi P, D’Ambrosio K, Pirone L, Pedone E, Aucelli T, She Q, De Simone G, Bartolucci S. 2011. C68 from the *Sulfolobus islandicus* plasmid-virus pSSVx is a novel member of the AbrB-like transcription factor family. *Biochem. J.* 435:157–166.
37. Stacey SN, Dahlquist FW, Reich NO. 2007. The role of high-affinity nonspecific DNA binding by Lrp in transcriptional regulation and DNA organization. *J. Mol. Biol.* 369:1307–1317.
38. Guillière F, Peixeiro N, Kessler A, Raynal B, Desnoves N, Keller J, Delepierre M, Prangishvili D, Sezonov G, Guisjarro JI. 2009. Structure, function, and targets of the transcriptional regulator SvtR from the hyperthermophilic archaeal virus SIRV1. *J. Biol. Chem.* 284:22222–22237.
39. Qureshi SA. 2007. Protein-DNA interactions at the *Sulfolobus* spindle-shaped virus-1 (SSV1) T5 and T6 gene promoters. *Can. J. Microbiol.* 53: 1076–1083.
40. Lipps G, Stegert M, Krauss G. 2001. Thermostable and site-specific DNA binding of the gene product ORF56 from the *Sulfolobus islandicus* plasmid pRNI1, a putative archaeal plasmid copy control protein. *Nucleic Acids Res.* 29:904–913.
41. Peixeiro N, Keller J, Collinet B, Leulliot N, Campanacci V, Cortez D, Cambillau C, Nitta KR, Vincentelli R, Forterre P, Prangishvili D, Sezonov G, van Tilbeurgh H. 2013. Structure and function of AvtR, a novel transcriptional regulator from a hyperthermophilic archaeal lipothrixvirus. *J. Virol.* 87:124–136.
42. Zampini M, Hayes F. 2012. Combinatorial targeting of ribbon-helix-helix artificial transcription factors to chimeric recognition sites. *Nucleic Acids Res.* 40:6673–6682.
43. Di Fiore A, Fiorentino G, Vitale RM, Ronca R, Amodeo P, Pedone C, Bartolucci S, De Simone G. 2009. Structural analysis of BldR from *Sulfolobus solfataricus* provides insights into the molecular basis of transcriptional activation in *Archaea* by MarR family proteins. *J. Mol. Biol.* 388: 559–569.
44. Fiorentino G, Cannio R, Rossi M, Bartolucci S. 2003. Transcriptional regulation of the gene encoding an alcohol dehydrogenase in the archaeon *Sulfolobus solfataricus* involves multiple factors and control elements. *J. Bacteriol.* 185:3926–3934.
45. Bell SD, Cairns SS, Robson RL, Jackson SP. 1999. Transcriptional regulation of an archaeal operon in vivo and in vitro. *Mol. Cell* 4:971–982.
46. Brinkman AB, Dahlke I, Tuininga JE, Lammers T, Dumay V, de Heus E, Lebbink JH, Thomm M, de Vos WM, van Der Oost J. 2000. An Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus* is negatively autoregulated. *J. Biol. Chem.* 275:38160–38169.
47. Bell SD, Jackson SP. 2000. Mechanism of autoregulation by an archaeal transcriptional repressor. *J. Biol. Chem.* 275:31624–31629.
48. Lee SJ, Engelmann A, Horlacher R, Qu Q, Vierke G, Hebbeln C, Thomm M, Boos W. 2003. TrmB, a sugar-specific transcriptional regulator of the trehalose/maltose ABC transporter from the hyperthermophilic archaeon *Thermococcus litoralis*. *J. Biol. Chem.* 278:983–990.