A novel archaeal regulatory protein, Sta1, activates transcription from viral promoters.

Alexandra Kessler, Guennadi Sezonov, J. Inaki Guijarro, Nicole Desnoues, Thierry Rose, Muriel Delepierre, Stepphen D. Bell, David Prangishvili

Nucleic Acids Research Advance Access published September 14, 2006

Nucleic Acids Research, 2006, Vol. 00, No. 00 ***1–9***
doi:10.1093/nar/gkl502

A novel archaeal regulatory protein, Sta1, activates transcription from viral promoters

Alexandra Kessler[1,4], Guennadi Sezonov[1], J. Iñaki Guijarro[2], Nicole Desnoues[1], Thierry Rose[3], Muriel Delepierre[2], Stephen D. Bell[5] and David Prangishvili[1,*]

[1]Unité de Biologie Moléculaire du Gène chez les Extrêmophiles, [2]Unité de RMN des Biomolécules (CNRS URA 2185), [3]Unité d'Immunogénétique Cellulaire, Institut Pasteur, 25-28 rue du Dr Roux, 75724 Paris Cedex 15, France, [4]Department of Microbiology, University of Regensburg, Universitäts strasse, 31, D-93053 Regensburg, Germany and [5]MRC Cancer Cell Unit Hutchison/MRC Research Centre, Hills Road, Cambridge CB2 2XZ, UK

Received March 17, 2006; Revised June 30, 2006; Accepted July 3, 2006

ABSTRACT

While studying gene expression of the rudivirus SIRV1 in cells of its host, the hyperthermophilic crenarchaeon *Sulfolobus*, a novel archaeal transcriptional regulator was isolated. The 14 kDa protein, termed *Sulfolobus* transcription activator 1, Sta1, is encoded on the host chromosome. Its activating effect on transcription initiation from viral promoters was demonstrated in *in vitro* transcription experiments using a reconstituted host system containing the RNA polymerase, TATA-binding protein (TBP) and transcription factor B (TFB). Most pronounced activation was observed at low concentrations of either of the two transcription factors, TBP or TFB. Sta1 was able to bind viral promoters independently of any component of the host pre-initiation complex. Two binding sites were revealed by footprinting, one located in the core promoter region and the second ~30 bp upstream of it. Comparative modeling, NMR and circular dichroism of Sta1 indicated that the protein contained a winged helix–turn–helix motif, most probably involved in DNA binding. This strategy of the archaeal virus to co-opt a host cell regulator to promote transcription of its genes resembles eukaryal virus–host relationships.

INTRODUCTION

The mechanisms and regulation of gene expression in the Archaea have been studied during the past 25 years [reviewed in (1)]. However, our knowledge on them remains modest in comparison to what is known on transcription in the other two domains of life, the Eukarya and Bacteria. Initial studies revealed that the archaeal basal transcription machinery resembles the core components of the eukaryal RNA polymerase (RNA Pol) II apparatus (2–7). Through the establishment of *in vitro* transcription systems for some archaea (8–13), it became possible to identify the archaeal factors necessary for specific initiation of transcription. Consisting of only the TATA-binding protein (TBP), transcription factor B (TFB), homologous to the eukaryotic TFIIB, and the RNA polymerase, a multi-subunit enzyme, the minimal archaeal transcription pre-initiation complex appears to be a simplified version of the eukaryotic RNA Pol II system. With the ongoing genome sequencing projects many transcription regulators could be identified in archaeal genomes. Surprisingly, many of them were homologs to the members of the bacterial Lrp-like regulator family (14,15). How regulation of an eukaryotic-like system could occur using bacterial-like regulators remains an intriguing question, mainly from an evolutionary point of view. Some of these regulators have been studied in cell-free transcription systems. Except the transcription activators Ptr2 from *Methanocaldococcus jannaschii* (16), and the homologous Lrp protein Mth from *Methanothermococcus thermolithotrophicus* (17), these were exclusively repressors: MDR1-repressor of the ABC-transporter-gene from *Archaeoglobus fulgidus* (18), LrpA from *Pyrococcus furiosus* (19,20), the negatively autoregulated factor Lrs14 from *Sulfolobus solfataricus* (21,22), and Phr involved in the heat-shock response of *P.furiosus* (23). However, the physiological functions of most of these regulators are still unclear. It would appear that a majority of *trans*- and *cis*-acting regulatory transcription factors of the Archaea still remain unknown.

In a situation in which efficient genetic tools are not yet available, one possibility to study transcription regulation in hypethermophilic archaea is offered by diverse crenarchaeal virus–host systems. Although studies on transcription of the *Sulfolobus* virus SSV1, crucial for the identification of archaeal promoter sequences, were carried out about two decades ago, detailed analysis of transcription of viruses of hyperthermophilic crenarchaea over the replication cycle was performed only recently. *In vivo* transcription studies on the rudiviruses SIRV1 and SIRV2 infecting

*To whom correspondence should be addressed. Tel: +33 144 38 9119; Fax: +33 145 68 8834; Email: prangish@pasteur.fr

the hyperthermophilic crenarchaeon *Sulfolobus islandicus* demonstrated a rather simple and barely chronological pattern of transcription, with a few cases of temporal regulation (24). SIRV promoters, similar to the host promoters, contain a TATA-box and a TFB responsive element. However, most of them contain an additional virus-specific consensus element. These observations suggested a major role for the host transcription machinery in the transcription of viral genomes, as well as possible involvement of virus-specific transcription factors.

Here, we report on the isolation and characterization of a host-encoded transcription regulator Sta1 involved in the activation of transcription from promoters of the crenarchaeal virus SIRV1.

MATERIALS AND METHODS

Biotinylation of promoter DNAs

Biotinylated promoter DNA used for the magnetic DNA affinity purification experiments were generated by PCR using biotinylated primers. The promoter regions 56, 134 and 399 were amplified from genomic DNA using the following primer sets: GACTCTGTTCTTGAGTTTGCA and Biotin-ATTGAATTAGTTCCAAAGTCTATTAGCG for 56, GAAATTCTGTTGGGCAACAGGAGC and Biotin-AGCAGATATGACAATTTAATAGTT for 134, and TTAGACTTGAAACAAATAACGGATAAC and Biotin-TTCTCAACTAATTCTTAAACCAATATA for 399. Biotinylated T6 promoter was reamplified from the T6 promoter plasmid described previously (13,18) using the primer set TGCATCCAACGCGTTGGGAGCTCTC and Biotin-TAATACGACTCACTATAGGG.

Magnetic DNA affinity purification of Sta1

The magnetic affinity purification was carried out as described previously (25) with modifications. For preparation of the affinity beads, 4 mg of streptavidin-coated magnetic beads (Dynabeads; Dynal Biotech) were resuspended in 2 × B & W buffer (10 mM Tris–HCl, pH 7.5, 2 M NaCl and 1 mM EDTA) to a final concentration of 5 μg/μl, after washing them once in 500 μl 1× B & W buffer. For immobilization of the promoter DNA to the beads, 800 μl 1× B & W buffer and 10 μg of the biotinylated promoter DNA fragment were added and incubated for 30 min at room temperature. After magnetic separation the affinity beads were washed three times in 1 ml 1× B & W buffer and resuspended in 150 μl of TE buffer. For the affinity purification, the affinity beads were incubated with crude extracts prepared from infected and non-infected cells (∼1.4 mg total protein) for 5 min at 25°C in a total volume of 2 ml buffer A [20 mM Tris–HCl, pH 8.0, 10 mM EDTA, 80 mM $(NH_4)_2SO_4$, 15% glycerol and 0.05% NP-40]. The plasmid pUC18 (200 μg) was added as unspecific competitor. After magnetic separation the beads were washed twice with 250 μl buffer A_{75} (20 mM Tris–HCl, pH 8.0, 10 mM EDTA, 75 mM NaCl, 15% glycerol and 0.05% NP-40). The bound protein was eluted with 150 μl buffer A_{380} (20 mM Tris–HCl, pH 8.0, 10 mM EDTA, 380 mM NaCl, 15% glycerol and 0.05% NP-40) and 20 μl of the eluate were subjected to SDS gel analysis.

Preparation of crude extracts of *Sulfolobus* cells

S.islandicus REN2H1 cells were grown as described previously (26) to an OD_{600} of 0.4, pelleted and resuspended in TBS buffer (10 mM Tris–HCl, pH 8.0 and 150 mM NaCl). Cells were lysed through sonication and the soluble protein fraction was collected after centrifugation in a SORVALL SS34 rotor at 170 000 r.p.m. for 20 min at 4°C. For the preparation of crude extracts of virus-infected cells, a growing *S.islandicus* REN2H1 culture was infected at an OD_{600} of 0.2 with SIRV1/VIII with a multiplicity of infection of 1 and incubated to an OD_{600} of 0.4.

Purification of RNA polymerase, TBP and TFB

RNA polymerase, TBP and TFB were purified as described previously (13,27). Transcription assays and DNase I footprinting were performed using RNA polymerase, TBP and TFB as described previously (18,21).

Transcription assays and DNase I footprinting

For transcription assays, PCR products of promoters 56, 134 and 399 were generated from genomic SIRV1 DNA using standard conditions and oligonucleotides. 5547S1F: 5′-GACTCTGTTCTTGAGTTTGCA-3′ and 5679S1R: 5′-TGGAATTCCATTAGTTCCAAGTCTATT-3′ for promoter 56; 10964S1R: 5′-AGCAGAATATGACAATTTAATAGTT-3′ and 11276S1R: 5′-GAAATTCTGTTGGGCAACAGGAGC-3′ for promoter 134; and 5034S1F: 5′-TTAGACTTGAAACAAATAACGGATAAC-3′ and 5367S1R: 5′-TTCTCAACTAATTCTTAAACCAATATA-3′ for promoter 399. The PCR products were cloned directly into pDrive (Qiagen) by T/A cloning. A plasmid carrying T6 promoter was generated as described previously (13). *In vitro* transcription reactions were performed using 100 ng of the corresponding plasmid DNA, 0.2 mM NTPs, 10 μg *Sulfolobus* whole-cell extract or 20 ng of TBP and TFB (or as indicated in figure legends), 1 μg RNA polymerase and Sta1 in amounts indicated in the figure legends. The reactions were carried out for 20 min at 70°C in 50 μl transcription buffer (50 mM Tris–HCl, pH 8.0, 75 mM KCl, 25 mM $MgCl_2$ and 1 mM DTT). Reactions were stopped by adding 250 μl NEW buffer (10 mM Tris–HCl, pH 8.0, 750 mM NaCl, 10 mM EDTA, 0.5% SDS and 40 μg/ml glycogen). The *in vitro* synthesized RNA was isolated by phenol–chloroform extraction followed by ethanol precipitation. Transcription products were detected by primer extension using radiolabeled T7 primer in the case of the T6 promoter template or sequence-specific primers for viral promoter templates as described previously (24).

DNase I footprinting was performed using a 300 bp fragment generated by PCR using the radiolabeled oligonucleotides 10964S1F and 11276S1R (see above). The DNA template was incubated with Sta1 as indicated in the legend to Figure 5 in 50 μl transcription buffer for 10 min at 48°C. Samples were treated for 1 min with 0.1 U of DNase I (Roche). Reactions were stopped by adding 250 μl NEW buffer. DNA fragments were isolated by phenol–chloroform extraction followed by ethanol precipitation. Pellets were resuspended in 20 μl TE buffer. Twenty microliters of 50% formamide loading dye were added and 20 μl of

the denatured sample were analyzed on an 8% denaturing polyacrylamide gel.

Mass spectrometry

Protein identification was performed by the Mass spectrometry facility of the MRC Laboratory of Molecular Biology. After SDS–PAGE, protein was in-gel digested with trypsin. Peptides analyzed were recovered on a Voyager-DE STR Biospectromery Workstation (PerSeptive Biosystems). Peak analysis and database interrogation were performed using the Mascot software package.

Heterologous expression of SSO0048 and purification of the recombinant protein

The gene SSO0048 was amplified by PCR from *S.solfataricus* P2 genomic DNA using primers 5′SSO0048 (5′-GGAATTCCTATGTCTGAAACCCAATTAA-3′) and 3′SSO0048 (5′-GGATCCCTCGAGTTACAATGGCTTG-AATTCCT-3′). The PCR product was digested with NdeI and XhoI and ligated to NdeI–XhoI digested pET30a. The sequence of the cloned DNA fragment was shown to be identical to the original *Sulfolobus* sequence. The expression construct was transformed into *Rosetta (DE3)pLysS* cells. Overexpression of non-tagged Sta1 protein was induced during logarithmic growth of cells by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to 1 mM for 4 h. The cell pellet was resuspended in N100 buffer (50 mM Tris–HCl, pH 8.0, 100 mM NaCl and 10 mM β-mercaptoethanol) lysed by sonication and clarified by centrifugation. Sta1 was purified to apparent homogenity from the crude cell lysate after the removal of heat-denatured cellular proteins by chromatography on a Heparin–Sepharose column. Sta1 was eluted using a linear NaCl gradient. Peak fractions were verified by SDS–PAGE and Coomassie blue staining.

Alternatively, for NMR experiments on Sta1, the protein was produced with a C-terminal hexahistidine-tag (Sta1-h6). This construct contained an 8-residue tail (LEHHHHHH) and a modification of the wild-type protein at position 127 (M instead of K). The gene of Sta1-h6 was cloned in a pET30a vector, and the protein was expressed and purified like recombinant Sta1, with only the Heparin–Sepharose step being replaced by an affinity chromatography using an Ni-NTA column. As assessed by biochemical assays and circular dichroism (CD) in the far-UV region, the tag does not influence the structure or the activity of Sta1.

Circular dichroism

CD in the far-UV region was performed on an Aviv 215 spectropolarimeter (Aviv Biomedical Inc., Lakewood, NJ). The concentration of Sta1 prepared in 10 mM sodium acetate, pH 5.5 (buffer NA), ranged between 20 and 100 μM. It was determined from the molar extinction coefficient of the protein calculated as described previously (28) CD spectra were recorded at 20°C between 180 and 260 nm with a step of 0.5 nm, a bandwidth of 1 nm and an optical path of 0.02 cm. The integration time was 4 or 1 s for points between 180 and 200 nm or 200.5 and 260 nm, respectively. Three scans were averaged. The CD spectrum of the protein was deconvoluted in terms of secondary structure content using the CONTIN (29) algorithm implemented in CDPro (30).

Analytical centrifugation

Sedimentation/diffusion equilibrium experiments were run at 20°C on an XL-I or XL-A ultracentrifuge (Beckman Coulter Inc.) of the Plateforme de Biophysique (Institut Pasteur). The ultracentrifuges were equipped with an AN-60 ti four hole rotor. Homodimerization of Sta1 was analyzed using rates between 14 and 22 kr.p.m. with samples of Sta1 (4.2, 8.4 or 42 μM) prepared in buffer NA supplemented with 150 mM NaCl.

Binding of Sta1 to a 30 bp DNA oligonucleotide, called Reg2, was followed with spinning rates of 12–20 kr.p.m. using samples obtained in a buffer containing 20 mM Tris–HCl, 150 mM NaCl and 1 mM Na EDTA (pH 8.0). HPLC-purified single-stranded oligonucleotides (5′-AATT-TATTAATTTAAAGAATAAAATTGATA-3′ and its complementary strand) were purchased from Proligo (Sigma-Aldrich). Oligonucleotides were mixed at an equimolar ratio in running buffer and annealed by incubation at 75°C for 10 min followed by a slow (2 h) return to room temperature. Experiments were run with 20 μM of Sta1 (protein only experiment), 5 μM DNA (DNA only experiment) or 15 μM DNA/30 μM Sta1 (binding experiment).

NMR

NMR experiments were acquired on an Inova 600 (Varian Inc., Palo Alto, CA) spectrometer with a 14.1 Tesla magnetic field. The spectrometer was equipped with a cryoprobe. Spectra were recorded, processed and analyzed using Vnmr 6.1C (Varian), NMRPipe (31) and NMRView 5.2 (32). Purified Sta1-h6 was dialyzed against 20 mM NH_4HCO_3 and freeze-dried. The lyophilized protein was dissolved in 20 mM CD_3COONa, pH 5.5 (uncorrected meter reading) prepared with 15 or 100% D_2O, for experiments in H_2O or D_2O, respectively. Experiments were performed at 37°C with a protein concentration of 0.3 mM. Homonuclear ^1H NOESY (nuclear Overhauser effect spectroscopy) spectra (33) were acquired with a 100 (H_2O) or 80 (D_2O) ms mixing time. The spectral width was 11 p.p.m., with 32 or 64 accumulations per free-induction decay and 400 (H_2O) or 256 (D_2O) complex data points in the indirect dimension.

Comparative modeling

A BLAST (34) search of the PDB with the sequence of Sta1 produced a single hit with a low *E*-value (0.007). The hit corresponded to the protein Mj233 from *M.jannaschii* (PDB code 1KU9). The structure of Mj233 was used as a template to obtain a model of Sta1 using Modeller v6.2 (35). The geometrical quality of the model was assessed using Procheck 3.5.4 (36).

RESULTS

Purification of DNA-binding proteins

SIRV1 promoters have been previously identified experimentally (24). In initial *in vitro* transcription studies on them, using a reconstituted system of *S.solfataricus*, in many cases only weak transcription initiation could be observed in comparison to strong transcription from the T6 promoter from the *Sulfolobus shibatae* virus 1, SSV1, known to be efficiently transcribed by the *Sulfolobus* transcription machinery. For example, transcription initiation from the SIRV1 56 gene promoter could be improved using whole-cell extracts of non-infected cells of *S.islandicus* (Figure 1). This observation suggested that viral promoters may need additional factors, present in the whole-cell extract, to turn on transcription of their genes.

Promoters of three genes of SIRV1, 56 and 399 (unknown function) and 134 (encoding for the structural protein), were chosen for a search of proteins involved in the regulation of transcription of the viral genome, as the previously obtained transcription map of SIRV1 had indicated that the expression of the three genes could be under transcriptional control (24).

The promoters were amplified from the viral DNA by PCR using biotinylated primers and applied in magnetic DNA affinity purification experiments (see Materials and Methods), using whole-cell extracts prepared from non-infected as well as virus-infected host cells. An unspecific competitor, pUC18 DNA, was added in high excess. The T6 promoter of SSV1, which was previously shown to be efficiently transcribed in a reconstituted transcription system of *Sulfolobus* (1,13,27), served as a control. In the conditions of the experiment, no protein was observed to bind to the T6 promoter (Figure 2). In contrast, a 14 kDa protein was bound specifically to all three SIRV promoters (Figure 2). The same result was obtained using crude extracts prepared from both virus-infected and non-infected cells, suggesting that the 14 kDa protein was encoded by the *Sulfolobus* host. Through elution from the beads, the protein could be purified to homogeneity, as judged by SDS–PAGE (Supplementary Figure S1).

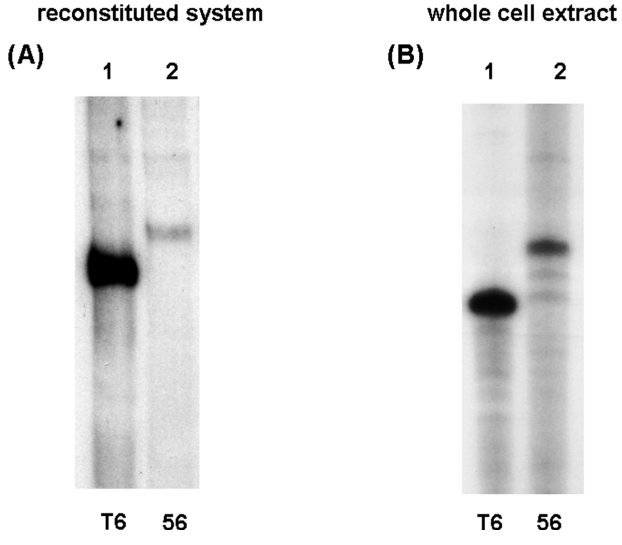

Figure 1. Transcription activation in the presence of whole-cell extracts from *S.islandicus* REN 2H1. *In vitro* transcription from SSV promoter T6 and the SIRV promoter 56 using either (**A**) a reconstituted transcription system consisting of 20 ng recombinant TFB and TBP and 200 ng of the RNA polymerase from *S.solfataricus* or (**B**) whole-cell extracts from *S.islandicus* REN2H1. (A) Lane 1, T6 promoter; lane 2, promoter 56 of SIRV1. (B) Ten micrograms of whole-cell extracts from *S.islandicus* REN2H1 were added per reaction. Lane 1, T6 promoter; lane 2, promoter 56.

Effect of the 14 kDa DNA-binding protein on viral transcription

In order to get insights into its function, the highly purified 14 kDa protein was studied using a cell-free transcription system of the host. The system consisted of recombinant TBP and TFB as well as highly purified RNA polymerase from *Sulfolobus* (13). As DNA templates, we used the same viral promoters that were used for affinity purification. The T6 promoter again served as a control. Although the 14 kDa protein had no effect on transcription initiation on the T6 promoter, a stimulation of transcription was observed for the viral gene promoters 56 and 134 (Figure 3A). Owing to its origin and activating effect, we term the protein *Sulfolobus* transcription activator, Sta1.

Identification and heterologous expression of the *sta1* gene, activating effect of the recombinant protein

For identification of the gene encoding Sta1, the protein was identified by using MALDI–TOF mass spectrometry. It was identified as a *S.islandicus* homolog of the gene SSO0048 of *S.solfataricus*, a species closely related to *S.islandicus*. The putative protein encoded by this gene, owing to its predicted helix–turn–helix motif in the annotation of *S.solfataricus* genome sequence (37), was presumed to be a transcription factor with homology to the *S.solfataricus* Lrs14 transcription regulator (18,22).

The *SSO0048* gene of *S.solfataricus* was cloned and expressed in *Escherichia coli* in native form, as well as with a C-terminal His-tag. The recombinant protein in both forms, Sta1 and Sta1-h6, was purified to apparent homogeneity (Supplementary Figure S2).

The activity of the recombinant protein was inspected by *in vitro* transcription experiments using the SIRV1 gene promoter 134 as a template. To ensure that the *in vitro* transcription start site was identical to the one *in vivo*, the primer extension product of the *in vitro* transcription reaction was

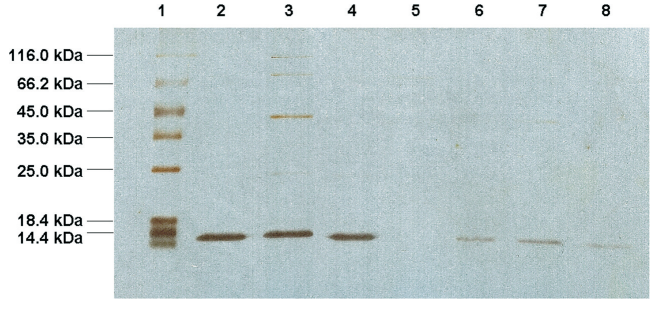

Figure 2. DNA affinity purification of the 14 kDa DNA-binding protein from whole-cell extracts of *S.islandicus* REN2H1. Lane 1, size markers; lanes 2, 3 and 4, proteins purified by their binding to promoters 56, 134 and 399, respectively, from non-infected cell extracts; lanes 6, 7 and 8, the same as lanes 2, 3 and 4, correspondingly, but from SIRV1-infected cell extracts; and lane 5, control experiment with T6 promoter and non-infected cell extract. Proteins were silver stained.

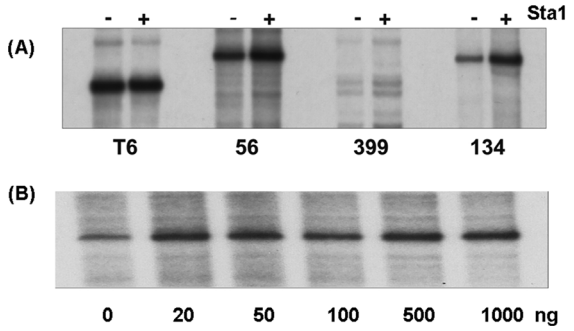

Figure 3. Transcription activation from SIRV1 promoters by the native *S.islandicus* 14 kDa protein, Sta1, and its recombinant form. (**A**) *In vitro* transcriptions were carried out by the native protein on the indicated promoters in the assay containing 20 ng TBP, 20 ng TFB, 250 ng RNA polymerase, either with or without 20 ng of the 14 kDa protein. (**B**) *In vitro* transcriptions were carried out by different amount of the recombinant protein on promoter 134 in the assays (50 µl) containing 20 ng TBP, 20 ng TFB, 250 ng RNA polymerase and either 20, 50, 100, 500 or 1000 ng of the recombinant Sta1.

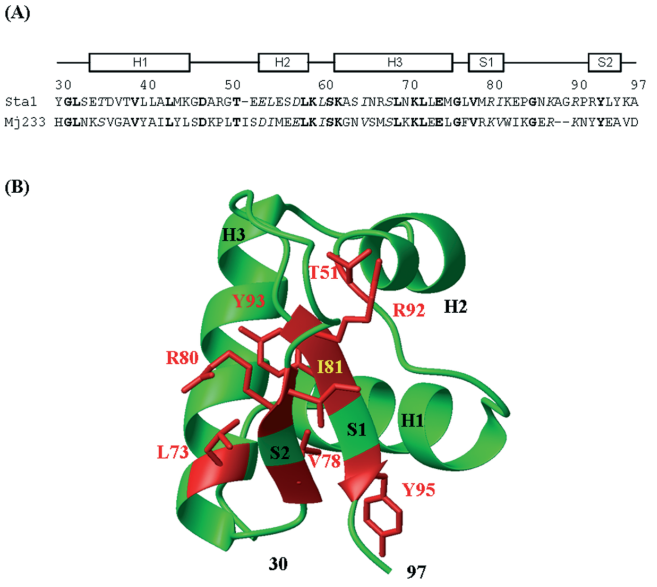

Figure 4. (**A**) Sequence alignment of Sta1 (30–97) and Mj233 (22–88) used to obtain the model of Sta1. Numbering and secondary structure correspond to Sta1. Identical residues are shown in boldface and similar residues in italics. The sequences of Sta1 and Mj233 display 28% of identical and 43% of similar residues. (**B**) Ribbon diagram of the model of Sta1 (wHTH region). The backbone and side chains of the residues for which at least one long-range NOE predicted by the model (distance <5 Å) was unambiguously identified in the NMR NOESY spectra of Sta1 are indicated in red.

analyzed together with its sequence reaction. The recombinant Sta1 was further tested in different amounts in the reconstituted transcription system with the promoter 134 as DNA template. The results shown in Figure 3B demonstrate that the recombinant protein at a concentration of 0.4 µg/ml has the same stimulating effect on transcription as the one purified from *Sulfolobus* cells. Increasing the concentration of the recombinant Sta1 > 0.4 µg/ml no stronger stimulation was observed (Figure 3B).

Structural analysis of Sta1

A model of the putative helix–turn–helix region of Sta1 was obtained by comparative modeling and validated experimentally using NMR and CD (Figure 4). The protein Mj233 from *M.jannaschii*, which is the closest homolog of Sta1 with known structure, served as a template (Materials and Methods). Mj233 forms a homodimer in which each monomer contains a winged helix–turn–helix (wHTH) motif in its N-terminal region and two C-terminal α-helices involved in the dimerization interface (38). The wHTH module consists of a two- or three-stranded antiparallel β-sheet and three α-helices.

1D spectra (data not shown) and 2D NOESY ^{1}H NMR experiments of Sta1-h6 showed several characteristics indicating that the protein was rich in α-helices and contained β-sheets. In order to test the model, we assigned several signals of the NOESY spectra of Sta1-h6 acquired in D_2O and H_2O (Supplementary Figure S3). As β-sheets produce well-resolved downfield-shifted NH and Hα signals, and aromatic proton signals generally show very good dispersion, we focused on the antiparallel β-sheet predicted by the model, which contained two tyrosine residues (Y93 and Y95) in the second strand. We identified two tyrosine spin systems with downfield-shifted NH and Hα resonances, which indicated that the corresponding aromatic residues were located in a β-sheet. Remarkably, we found several NOEs implicating these tyrosine residues that were in accordance with the topology of the antiparallel β-sheet of the model. A careful analysis of the NOESY spectra allowed us to unambiguously assign several long-range NOEs in agreement with the model between residues 51 and 95 (Figure 4B). Hence, our results showed that Sta1 had indeed a wHTH motif that resembled to that of Mj233. Of note, analysis of the NOESY spectra indicated that Sta1 did not contain any extra β-sheet.

Sedimentation–diffusion equilibrium experiments performed by centrifugation indicated that, similar to Mj233 and generally, prokaryotic transcription regulators Sta1 showed a dimer–monomer equilibrium. The dissociation constant (K_D) of this equilibrium was 5 µM at 20°C. Far-UV CD spectra of Sta1 were recorded at concentrations higher than the K_D to observe the dimer's CD. Once normalized, the latter was concentration independent between 20 and 100 µM and was dominated by contributions of α-helical structures (Supplementary Figure S4). As estimated by deconvolution of the CD spectrum, Sta1 contained 39% of α-helices and 16% of β-sheets. Taking into account that the content of α-helices of the wHTH motif (residues 30–97) was ~25%, the rest of the molecule (residues 1–29 and 98–129) should contain ~20 residues in α-helices.

Functional analysis of Sta1

For functional studies of Sta1, the 134 promoter was chosen, as it revealed the strongest level of activation (Figure 3A). In order to identify Sta1-binding sites, we employed DNase I footprinting assays on the promoter 134. Two non-sensitive regions were clearly visible (Figure 5). One binding site was shown to be located in the core promoter region, the second ~30 nt upstream of it. About 20 bp were protected in both regions at low concentrations of the protein (Figure 5). Increasing the protein concentration resulted in extension of

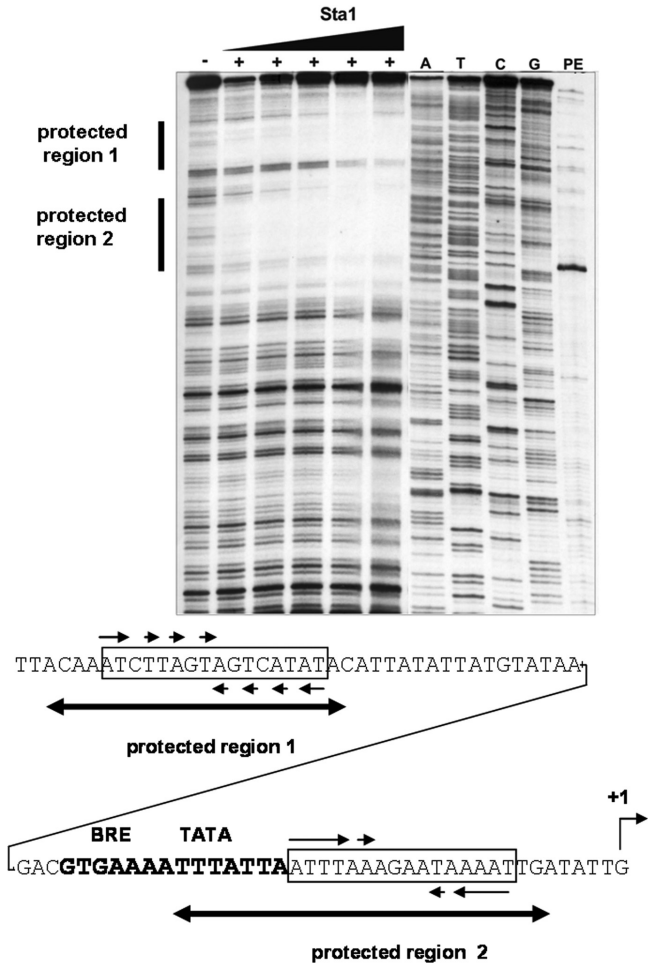

Figure 5. Binding sites of Sta1 in the SIRV1 promoter 134. DNase I footprinting either with 50, 100, 250, 500 and 1000 ng of Sta1, or without it (first lane), was subjected to electrophoresis together with the corresponding sequencing reactions (A, T, C and G) and the primer extension product of the SIRV1 promoter 134. Bold lines on the left point to non-sensitive regions (protected regions 1 and 2). The sequence of the SIRV1 promoter 134 region is shown below the autoradiograph. The transcription start site, the TATA-box and the BRE element are specified. The proposed 16 bp Sta1-binding sites are highlighted by boxes and putative imperfect inverted repeats are indicated by arrows.

the protected areas, suggesting multiple binding. The results of the footprinting confirmed the initial observation that Sta1 was able to bind to viral promoters in the absence of other components of the transcription pre-initiation complex (Figure 2).

Sta1 showed a monomer–dimer equilibrium. To assess, which species binds DNA, we performed sedimentation/equilibrium experiments of Sta1 at 20°C in the presence or absence of a 30 bp oligonucleotide (Reg2), corresponding to a fragment of the protected region 2 in the footprinting experiment (Figure 5). Experiments were performed at protein concentrations above the Sta1 dimerization K_D. Sedimentation/equilibrium profiles (Supplementary Figure S5) of Reg2 alone were well-fitted by a single species with a mass corresponding to double-stranded DNA (18.4 kDa). Fitting of the Sta1 profiles showed an equilibrium of monomeric and dimeric forms of the protein, with only the dimeric form being substantially populated at 20 μM, the concentration used in the experiment. Importantly, the binding experiment was well described by the association of a dimer of protein with double-stranded DNA, with a dissociation constant of $10 \pm 2\ 10^{-6}$ M. Hence, these results indicate that Sta1 binds to DNA as a dimer.

Considering that Sta1 binds DNA as a homodimer, its binding sites are expected to comprise inverted repeat sequences, as is the case for many transcription regulators. Imperfect palindromes detected in both the 20-bp-long protected regions are highlighted by arrows in Figure 5.

The apparent proximity of one of the binding sites of Sta1 with the TATA-box and the BRE element suggested that the activating effect of Sta1 might influence binding by the general transcription factors, TBP and/or TFB. Thus, the effect of Sta1 was studied in *in vitro* transcription experiments with varying concentrations of TBP and TFB. As expected, in absence of TBP, no transcription initiation could occur (Figure 6A). In the presence of low amounts of TBP, however, the activating effect of Sta1 was observed. Quantification of band intensities by phosphorimaging revealed that the strongest activation effect was observed using low amounts of TBP. Indeed, in the presence of 1 ng of TBP, transcription of promoter 134 is about five times stronger if Sta1 is added and the effect is even more dramatic (nearly 10-fold difference) in the presence of 5 ng of TBP (Figure 6A). The activating effect of Sta1 was also observed in analogous titration experiments in which the concentration of TFB was varied (Figure 6B).

The activating effect of Sta1 was also observed in analogous titration experiments in which the concentration of TFB was varied (Figure 6B). The effect of Sta1 was shown to be specific for SIRV1 promoters: transcription efficiency of the *Sulfolobus* promoter T6 was not affected in experiments with varying concentrations of either TBP or TFB (Figure 6B). These results are in line with the failure to pull-down Sta1 using the immobilized T6 promoter.

DISCUSSION

In order to isolate proteins involved in transcription of genes of the *Sulfolobus* virus SIRV1, DNA affinity purification experiments were conducted. With the help of three different viral promoters immobilized on magnetic beads, we isolated a 14 kDa DNA-binding protein from crude extracts of host cells. All three promoters bound to the same protein, which turned out to be encoded on the chromosome of the *Sulfolobus* host. When we used crude extracts from virus-infected host cells in the purification experiments, the same protein was bound to all three promoters. The protein was identified as the product of the *S.islandicus* homolog of the gene SSO0048 from *S.solfataricus*, and was named Sta1.

Analysis of Sta1, either purified from host cell extracts or recombinant, showed an activating effect on transcription from two viral promoters in *in vitro* transcription experiments.

Sta1 is the first archaeal transcription regulator isolated by pull-down assays (25) with archaeal promoters directly from cells. All other archaeal transcription regulators have been identified by *in silico* analysis. Our results open possibilities

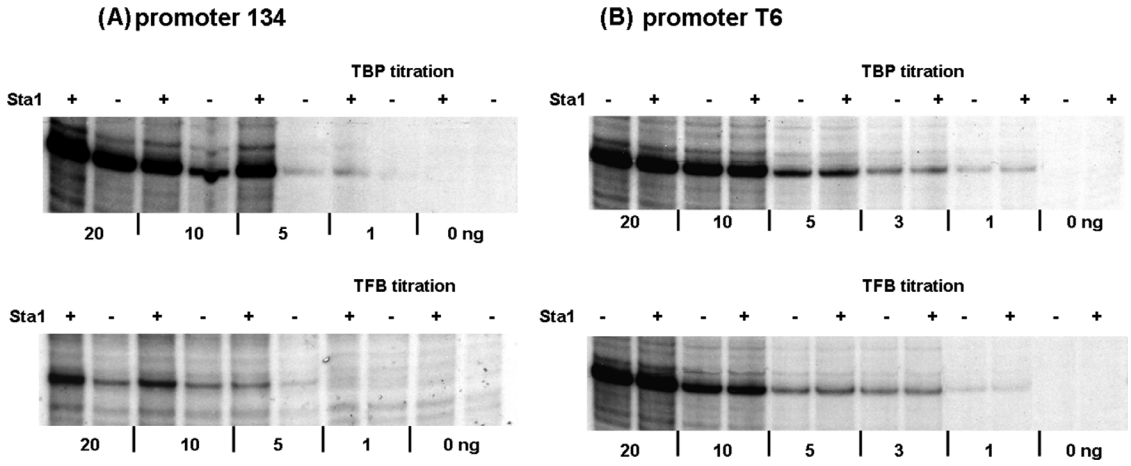

Figure 6. Activating effect of Sta1 in the presence of varying concentrations of TBP and TFB. (**A**) The assays (50 μl) on promoter 134 contained either 20 ng TFB, 250 ng RNA polymerase, 20 ng Sta1, and either 20, 10, 5, 1 ng or no TBP (upper panel), or 20 ng TBP, 250 ng RNA polymerase, 20 ng Sta1, and either 20, 10, 5, 1 ng or no TFB (lower panel). (**B**) The assays (50 μl) on T6 promoter contained either 20 ng TFB, 250 ng RNA polymerase, 20 ng Sta1, and either 20, 10, 5, 3, 1 ng or no TBP (upper panel), or 20 ng TBP, 250 ng RNA polymerase, 20 ng Sta1, and either 20, 10, 5, 3, 1 ng or no TFB (lower panel).

for the identification of unknown proteins involved in the regulation of archaeal gene expression.

Binding sites for Sta1 were determined by DNase I footprinting. Two distinct protected regions of ~20 bp could be clearly identified. One of them was located in the core promoter region and the other was ~30 nt upstream of it. The analysis of the protected sequences allowed us to identify the imperfect 16 bp inverted repeat which is also present in the promoter region of the SIRV1 gene 56 (Figure 7). The alignment of the identified and putative Sta1-binding sites let us to design the consensus site as ATNT-N_8-A/TNAT (Figure 7). Location of a binding site in immediate proximity of the TATA-box and the BRE element suggested that the activating effect of Sta1 could be associated with the TBP and/or TFB, for example, by enhancing their recruitment or by stabilization of their binding. To confirm the possibility of a TBP/TFB-dependent effect, we performed *in vitro* transcription experiments under suboptimal concentrations of TBP and TFB. The results demonstrated that in the presence of low amounts of one of the two factors, either TBP or TFB, Sta1 is necessary for transcription initiation from the assayed viral promoter, especially in the case of low amounts of TBP.

In general, mechanisms of transcription activation in archaea are poorly understood. Sta1 is the first transcription activator from the Crenarchaeota for which an activating effect has been directly demonstrated in *in vitro* studies. The current knowledge on the molecular basis of transcription activation in the entire archaeal domain of life is limited to results gleaned from analysis of the recently described factor Ptr2 from *M.janaschii* and its Lrp ortholog from *M.thermolithotrophicus* (16,17). Activation by Ptr2 is generated by recruitment of the TATA-binding protein to the promoter, and conveys its stimulatory effect, in contrast to Sta1, from two upstream-located binding sites (16,17). Significantly, Sta1 is only distantly related to Ptr2 and appears to operate by distinct mechanisms from Ptr2, facilitating transcription at limiting TBP and TFB, suggesting that it belongs to a novel class of archaeal transcriptional activators.

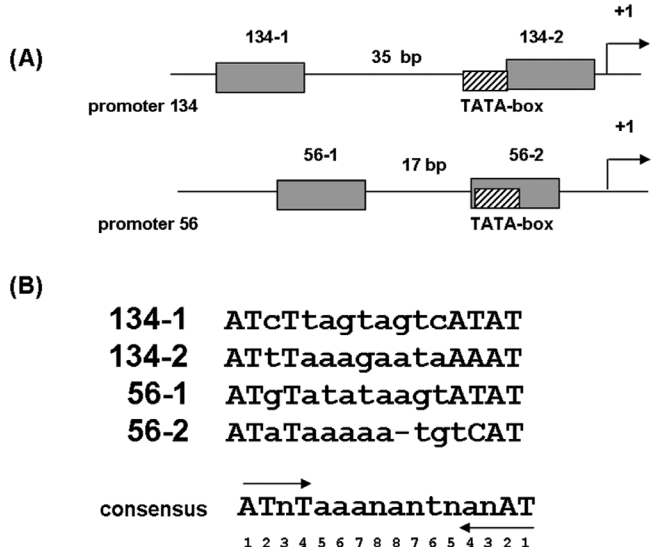

Figure 7. The putative 16 bp consensus Sta1-binding site. (**A**) The positions of the identified (promoter 134) and putative (promoter 56) Sta1-binding site in the promoter regions of the corresponding SIRV1 genes are indicated by solid boxes. The striped boxes correspond to the TATA-box. (**B**) The alignment of the Sta1-binding sites and the proposed consensus site. The imperfect inverted repeat is indicated by arrows. In the consensus sequence the upper case letters indicate highly conserved nucleotides, the lower case letter indicates a nucleotide with one mismatch and 'n' indicates the non-conserved nucleotides.

Genomic analysis implies that Sta1 could represent a group of archaeal-specific transcription regulators. Conserved domains search (39) identified an archaea-specific domain in Sta1, annotated as predicted transcription regulator (CDD 12 688). Clear homologs of Sta1, containing 120–130 amino acids, are present on all three sequenced genomes of *Sulfolobus* species, at least in five copies in each of them, based on an *E*-value threshold of 0.01. However, given the small size of the protein, additional homologs may be present but not identified as significant. Applying the same threshold,

Sta1 homologs of 120–130 amino acids were also identified in many other archaeal genomes. In contrast, BLAST searches against bacterial proteins reveal only a weak similarity to proteins of the MarR family. No evident homologs of Sta1 were found in eukaryotes.

The protein encoded by gene *SSO0048* was initially annotated as a putative transcription regulator with a similarity to the *S.solfataricus* transcriptional regulator Lrs14 (37) of the Lrp family. However, we showed here using NMR and homology modeling that Sta1, in contrast to the classical regulators of the Lrp family that display a HTH module (15,19,22,40), comprises a wHTH motif with an antiparallel β-sheet, which is absent in Lrp-like proteins. Our results, together with a BLAST search and a DALI (41) search of homologous protein structures, indicated that Sta1 contained a wHTH domain similar to that of Mj233 (38) and multiple antibiotic resistance proteins such as MarR, which is implicated in stress response (38,42). The electrostatic potential of the model of the wHTH region of Sta1 suggested that helix H1 may be involved in hydrophobic interactions within the protein, and that the region of the molecule containing the loop between strands B1 and B2 (wing), rich in positively charged residues, the β-sheet as well as the so-called recognition helix (H3), are good candidates to participate in the interaction with DNA.

A strategy of the archaeal virus to implicate a host transcription regulator to promote transcription of viral genes resembles eukaryal virus–host relationships and is in line with pronounced similarities in transcription machineries of archaea and eukarya. Numerous cases of such regulations are known in eukaryal virus–host systems. Some examples include requirement of the host-encoded ribonucleoproteins A2/B1 and RBM3 in transcription activation of the vaccinia virus late genes (43); exploitation of the cellular transcription factor USF by herpes viruses for the regulation of their promoters (44); a presence of binding motifs for ubiquitous cellular transcriptional enhancer factors TEF-1 and TEF-2 and nuclear factor NF-1 upstream of regulatory regions of papillomaviruses (45,46); and in the case of baculoviruses transcription from most early genes appears to be activated by the interaction of both, host and viral transcription factors with viral regulatory sequence elements (47).

Known examples of recruitment by viruses of essential cellular regulators of different nature for the modulation of gene expression determine a considerable interest in the identification of cellular targets of Sta1. In this light, studies of gene regulation in archaeal viruses, in addition to contributing to the knowledge on diversity and evolution of molecular mechanisms of gene regulation, may provide a tool for identifying key regulators of cellular events.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

We thank Patrick Forterre and Simonetta Gribaldo for stimulating discussions, Sew Yeu Peak-Chew of the MRC Laboratory of Molecular Biology for Mass Spectrometry. The work was supported by grants from the Deutsche Forschungsgemeinschaft (PR 663/1–2), and the EMBO. The 600 MHz spectrometer was funded by the Région Ile de France and the Institut Pasteur. Funding to pay the Open Access publication charges for this article was provided by the Institut Pasteur, Paris, France.

Conflict of interest statement. None declared.

REFERENCES

1. Bell,S.D. and Jackson,S.P. (2001) Mechanism and regulation of transcription in archaea. *Curr. Opin. Microbiol.*, **4**, 208–213.
2. Langer,D., Hain,J., Thuriaux,P. and Zillig,W. (1995) Transcription in archaea: similarity to that in eucarya. *Proc. Natl Acad. Sci. USA*, **92**, 5768–5772.
3. Littlefield,O., Korkhin,Y. and Sigler,P.B. (1999) The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc. Natl Acad. Sci. USA*, **96**, 13668–13673.
4. Prangishvilli,D., Zillig,W., Gierl,A., Biesert,L. and Holz,I. (1982) DNA-dependent RNA polymerase of thermoacidophilic archaebacteria. *Eur. J. Biochem.*, **122**, 471–477.
5. Qureshi,S.A., Baumann,P., Rowlands,T., Khoo,B. and Jackson,S.P. (1995) Cloning and functional analysis of the TATA binding protein from *Sulfolobus shibatae*. *Nucleic Acids Res.*, **23**, 1775–1781.
6. Zillig,W., Palm,P., Klenk,H.P., Langer,D., Hüdepohl,U., Hain,J., Lanzendörfer,M. and Holz,I. (1993) *Transcription in Archaea*. Elsevier Science Publishers B.V, Amsterdam, The Netherlands.
7. Zillig,W., Stetter,K.O. and Janekovic,D. (1979) DNA-dependent RNA polymerase from the archaebacterium *Sulfolobus acidocaldarius*. *Eur. J. Biochem.*, **96**, 597–604.
8. Darcy,T.J., Hausner,W., Awery,D.E., Edwards,A.M., Thomm,M. and Reeve,J.N. (1999) *Methanobacterium thermoautotrophicum* RNA polymerase and transcription *in vitro*. *J. Bacteriol.*, **181**, 4424–4429.
9. Frey,G., Thomm,M., Brudigam,B., Gohl,H.P. and Hausner,W. (1990) An archaebacterial cell-free transcription system. The expression of tRNA genes from *Methanococcus vannielii* is mediated by a transcription factor. *Nucleic Acids Res.*, **18**, 1361–1367.
10. Hausner,W., Frey,G. and Thomm,M. (1991) Control regions of an archaeal gene. A TATA box and an initiator element promote cell-free transcription of the tRNA(Val) gene of *Methanococcus vannielii*. *J. Mol. Biol.*, **222**, 495–508.
11. Hethke,C., Geerling,A.C., Hausner,W., de Vos,W.M. and Thomm,M. (1996) A cell-free transcription system for the hyperthermophilic archaeon *Pyrococcus furiosus*. *Nucleic Acids Res.*, **24**, 2369–2376.
12. Hudepohl,U., Reiter,W.D. and Zillig,W. (1990) *In vitro* transcription of two rRNA genes of the archaebacterium *Sulfolobus* sp. B12 indicates a factor requirement for specific initiation. *Proc. Natl Acad. Sci. USA*, **87**, 5851–5855.
13. Qureshi,S.A., Bell,S.D. and Jackson,S.P. (1997) Factor requirements for transcription in the Archaeon *Sulfolobus shibatae*. *EMBO J.*, **16**, 2927–2936.
14. Aravind,L. and Koonin,E.V. (1999) DNA-binding proteins and evolution of transcription regulation in the archaea. *Nucleic Acids Res.*, **27**, 4658–4670.
15. Brinkman,A.B., Ettema,T.J., de Vos,W.M. and van der Oost,J. (2003) The Lrp family of transcriptional regulators. *Mol. Microbiol.*, **48**, 287–294.
16. Ouhammouch,M., Dewhurst,R.E., Hausner,W., Thomm,M. and Geiduschek,E.P. (2003) Activation of archaeal transcription by recruitment of the TATA-binding protein. *Proc. Natl Acad. Sci. USA*, **100**, 5097–5102.
17. Ouhammouch,M. and Geiduschek,E.P. (2005) An expanding family of archaeal transcriptional activators. *Proc. Natl Acad. Sci. USA*, **102**, 15423–15428.
18. Bell,S.D., Cairns,S.S., Robson,R.L. and Jackson,S.P. (1999) Transcriptional regulation of an archaeal operon *in vivo* and *in vitro*. *Mol. Cell*, **4**, 971–982.
19. Brinkman,A.B., Dahlke,I., Tuininga,J.E., Lammers,T., Dumay,V., de Heus,E., Lebbink,J.H., Thomm,M., de Vos,W.M. and van Der Oost,J. (2000) An Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus* is negatively autoregulated. *J. Biol. Chem.*, **275**, 38160–38169.

20. Dahlke,I. and Thomm,M. (2002) A *Pyrococcus* homolog of the leucine-responsive regulatory protein, LrpA, inhibits transcription by abrogating RNA polymerase recruitment. *Nucleic Acids Res.*, **30**, 701–710.
21. Bell,S.D. and Jackson,S.P. (2000) Mechanism of autoregulation by an archaeal transcriptional repressor. *J. Biol. Chem.*, **275**, 31624–31629.
22. Napoli,A., van der Oost,J., Sensen,C.W., Charlebois,R.L., Rossi,M. and Ciaramella,M. (1999) An Lrp-like protein of the hyperthermophilic archaeon *Sulfolobus solfataricus* which binds to its own promoter. *J. Bacteriol.*, **181**, 1474–1480.
23. Vierke,G., Engelmann,A., Hebbeln,C. and Thomm,M. (2003) A novel archaeal transcriptional regulator of heat shock response. *J. Biol. Chem.*, **278**, 18–26.
24. Kessler,A., Brinkman,A.B., van der Oost,J. and Prangishvili,D. (2004) Transcription of the rod-shaped viruses SIRV1 and SIRV2 of the hyperthermophilic archaeon *Sulfolobus*. *J. Bacteriol.*, **186**, 7745–7753.
25. Gabrielsen,O.S., Hornes,E., Korsnes,L., Ruet,A. and Oyen,T.B. (1989) Magnetic DNA affinity purification of yeast transcription factor tau—a new purification principle for the ultrarapid isolation of near homogeneous factor. *Nucleic Acids Res.*, **17**, 6253–6267.
26. Zillig,W., Kletzin,A., Schleper,C., Holz,I., Janekovic,D., Hain,J., Lanzendörfer,M. and Kristjansson,J.K. (1994) Screening for *Sulfolobales*, their plasmids and their viruses in Icelandic solfataras. *Syst. Appl. Microbiol.*, **16**, 609–628.
27. Bell,S.D. and Jackson,S.P. (2000) The role of transcription factor B in transcription initiation and promoter clearance in the archaeon *Sulfolobus acidocaldarius*. *J. Biol. Chem.*, **275**, 12934–12940.
28. Pace,C.N., Vajdos,F., Fee,L., Grimsley,G. and Gray,T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.*, **4**, 2411–2423.
29. Provencher,S.W. and Glockner,J. (1981) Estimation of globula protein secondary structure from circular dichroism. *Biochemistry*, **20**, 33–37.
30. Sreerama,N. and Woody,R.W. (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.*, **287**, 252–260.
31. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
32. Johnson,B.A. and Blevins,R.A. (1994) NMRView: a computer program for the visualisation and analysis of NMR data. *J. Biomol. NMR*, **4**, 604–614.
33. States,D.J., Haberkorn,R.A. and Ruben,D.J. (1982) A two dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants. *J. Magn. Reson.*, **48**, 286–292.
34. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
35. Sali,A., Overington,J.P., Johnson,M.S. and Blundell,T.L. (1990) From comparisons of protein sequences and structures to protein modelling and design. *Trends Biochem. Sci.*, **15**, 235–240.
36. Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
37. She,Q., Singh,R.K., Confalonieri,F., Zivanovic,Y., Allard,G., Awayez,M.J., Chan-Weiher,C.C., Clausen,I.G., Curtis,B.A., De Moors,A. *et al.* (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl Acad. Sci. USA*, **98**, 7835–7840.
38. Ray,S.S., Bonanno,J.B., Chen,H., de Lencastre,H., Wu,S., Tomasz,A. and Burley,S.K. (2003) X-ray structure of an *M. jannaschii* DNA-binding protein: implications for antibiotic resistance in *S. aureus*. *Proteins*, **50**, 170–173.
39. Marchler-Bauer,A. and Bryant,S.H. (2004) CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.*, **32**, W327–W331.
40. Leonard,P.M., Smits,S.H., Sedelnikova,S.E., Brinkman,A.B., de Vos,W.M., van der Oost,J., Rice,D.W. and Rafferty,J.B. (2001) Crystal structure of the Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus*. *EMBO J.*, **20**, 990–997.
41. Holm,L. and Sander,C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.
42. Alekshun,M.N., Levy,S.B., Mealy,T.R., Seaton,B.A. and Head,J.F. (2001) The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 A resolution. *Nature Struct. Biol.*, **8**, 710–714.
43. Wright,C.F., Oswald,B.W. and Dellis,S. (2001) Vaccinia virus late transcription is activated *in vitro* by cellular heterogeneous nuclear ribonucleoproteins. *J. Biol. Chem.*, **276**, 40680–40686.
44. Meier,J.L., Luo,X., Sawadogo,M. and Straus,S.E. (1994) The cellular transcription factor USF cooperates with varicella-zoster virus immediate-early protein 62 to symmetrically activate a bidirectional viral promoter. *Mol. Cell. Biol.*, **14**, 6896–6906.
45. Chong,T., Apt,D., Gloss,B., Isa,M. and Bernard,H.U. (1991) The enhancer of human papillomavirus type 16: binding sites for the ubiquitous transcription factors oct-1, NFA, TEF-2, NF1, and AP-1 participate in epithelial cell-specific transcription. *J. Virol.*, **65**, 5933–5943.
46. Chong,T., Chan,W.K. and Bernard,H.U. (1990) Transcriptional activation of human papillomavirus 16 by nuclear factor I, AP1, steroid receptors and a possibly novel transcription factor, PVF: a model for the composition of genital papillomavirus enhancers. *Nucleic Acids Res.*, **18**, 465–470.
47. Friesen,P.D. (1997) *Regulation of Baculovirus Early Gene Expression*. Plenum, NY.