Extremophiles (2010) 14:453–463
DOI 10.1007/s00792-010-0325-4

ORIGINAL PAPER

Transcription termination in the plasmid/virus hybrid pSSVx from *Sulfolobus islandicus*

Patrizia Contursi · Raffaele Cannio · Qunxin She

Received: 25 March 2010 / Accepted: 16 July 2010 / Published online: 25 August 2010

Abstract The pSSVx from *Sulfolobus islandicus,* strain REY15/4, is a hybrid between a plasmid and a fusellovirus. A systematic study previously performed revealed the presence of nine major transcripts, the expression of which was differentially and temporally regulated over the growth cycle of *S. islandicus*. In this study, two new transcripts were identified. Then, 3′ termini of all the RNAs were mapped using adaptor RT-PCR and RNase protection assays, and termination/arrest positions were identified for each transcript. The majority of the identified ending positions were located in the close vicinity of a T-rich sequence and this was consistent with termination signals identifiable for most of archaeal genes. Furthermore, termination also occurred at locations where a T-track sequence was absent but a stem-loop structure could be formed. We propose that an alternative mechanism based on secondary RNA structures and counter-transcripts might be responsible for the transcription termination at these T-track-minus loci in the closely spaced pSSVx genes.

Keywords Transcription termination · Stem-and-loop structure · Counter-transcripts · *Fuselloviridae*

Communicated by L. Huang.

Electronic supplementary material The online version of this article (doi:10.1007/s00792-010-0325-4) contains supplementary material, which is available to authorized users.

P. Contursi · Q. She
Archaeal Genetics Laboratory, Department of Molecular Biology, University of Copenhagen, Copenhagen Biocenter Ole Maaløes Vej 5, 2200 Copenhagen N, Denmark

P. Contursi (✉)
Dipartimento di Biologia Strutturale e Funzionale, Università degli Studi di Napoli Federico II, Complesso Universitario Monte S. Angelo, Via Cinthia, Naples, Italy
e-mail: contursi@unina.it

R. Cannio
Istituto di Biochimica delle Proteine, CNR, via Pietro Castellino 111, 80131 Naples, Italy

Introduction

Although the transcriptional machinery is conserved for all three domains of life, archaeal transcription apparatus exhibits striking similarities to that of Eucarya. Unlike the bacterial RNA polymerase that has only 4 different subunits, the archaeal RNA polymerase consists of 11–14 subunits and most of them have a counterpart in eukaryotic RNA polymerase, though the latter contains a few additional components (Korkhin et al. 2009; Kwapisz et al. 2008; Kusser et al. 2008). Archaeal general transcription factors, TBP, TFB and TFE are close homologues of eukaryotic TBP, TFIIB and TFIIE, both in terms of structure and of function. Therefore, the archaeal transcriptional machinery constitutes a simplified version of the eukaryotic RNA pol II (Bartlett 2005; Bell and Jackson 2001; Geiduschek and Ouhammouch 2005; Thomm 2007). Due largely to the relative ease of over-expression of their recombinant forms in, and their purification from, *E. coli,* hyperthermophilic proteins including those involved in the transcription process are more attractive targets for studying molecular mechanisms than their mesophilic counterparts. Consequently, hyperthermophilic archaea not only serve as useful model for the dissection of transcription initiation in Archaea, but also provide simplified/alternative systems for studying eukaryotic transcription mechanisms (Thomm 2007).

To date, relatively little has been conducted to study transcription termination in Archaea. As a consequence,

little is known about the mechanism of transcription termination and its regulation. Few studies have reported on the mapping of the 3′ termini of archaeal genes, most of which encode stable RNAs rather than mRNAs (Brenneis et al. 2007; Kjems and Garrett 1987; Kjems et al. 1987; Larsen et al. 1986; Müller et al. 1985; Wich et al. 1986). These studies have revealed that archaeal transcription terminators contain oligo-T stretches (Berkner and Lipps 2007; Larsen et al. 1986; Müller et al. 1985; Østergaard et al.1987) and it has been demonstrated that these signals are sufficient to direct termination by an archaeal RNA polymerase in vitro (Østergaard et al. 1987; Reiter et al. 1988b) and in vivo (Santangelo et al. 2009). Because oligo-T sequences are recognised by the eukaryotic RNA pol III as terminator signals (Braglia et al. 2005; Gunnery and Mathews 1995; Gunnery et al. 1999), it suggests that there is a close relationship between archaeal transcriptional termination and that by eukaryotic RNA pol III. However, Rho-independent bacterial terminators and the bacterial Rho-factor can also mediate termination of transcription in archaeal systems indicating clear similarities also between archaeal and bacterial termination mechanisms (Santangelo and Reeve 2006). Thus, it is interesting to systematically investigate, which mechanisms function in archaeal transcriptional termination.

Plasmids and viruses are essential tools for studying molecular mechanisms such as transcription, DNA replication, recombination and repair (Kornberg and Baker 1992), and are instrumental for genetic studies. A survey of the extrachromosomal elements in the hyperthermophilic crenarchaeon *Sulfolobus* (Zillig et al. 1996, 1998) has revealed the existence of many new viruses, with the *Fuselloviridae* being the most common family amongst the *Sulfolobales* (Prangishvili et al. 2001). *Sulfolobus* spindle-shaped virus 1 (SSV1) is the best-studied member of this family. DNA replication of SSV1 increases after induction by UV or other DNA-damaging agents and seems to be mediated by transcription at the promoter T_{ind} (Reiter et al. 1987; Schleper et al. 1992). The transcriptional analysis conducted on SSV1 has provided the basics for an early definition of the consensus sequences of archaeal promoters (Reiter et al. 1988a) and terminators (Reiter et al. 1988b) in both constitutive and UV-inducible transcripts.

One of the most striking findings in *Sulfolobus* virus studies has been the identification of the first archaeal helper and satellite virus system, SSV2 and pSSVx, in which SSV2 acts as an ordinary virus and helper virus to SSVx, whilst pSSVx is a satellite virus assumed to generate virus particles thanks to the packaging mechanisms of SSV2 (Arnold et al. 1999; Stedman et al. 2003).

SSV2 and pSSVx belong to the family *Fuselloviridae* and they do not induce cell lysis of their hosts in the entire life cycle, but impose strong inhibition of the growth of their host upon infection (Contursi et al. 2006). At the sequence level, the pSSVx genome contains two open reading frames (ORFs), which are conserved in the *Fuselloviridae* family (Arnold et al. 1999, Stedman et al. 2003); the remaining genome sequence is typically plasmidic, with the putative minimal replicon shared with members of the pRN plasmid family, such as pRN1 (Keeling et al. 1996), pRN2 (Keeling et al. 1998), pHEN7 from different *S. islandicus* species (Peng et al. 2000), pDL10 from *Acidianus ambivalens* (Kletzin et al. 1999), and several defective integrated plasmids occurring in *Sulfolobus* genomes (She et al. 2004).

We have used the virus/plasmid hybrid pSSVx as a model system to study archaeal transcription mechanisms and demonstrated that: (i) the replication of pSSVx is induced during the growth of the natural *S. islandicus* REY15/4 host (Contursi et al. 2006); (ii) the expression of all pSSVx genes/counter-transcripts is differentially regulated during the virus life cycle suggesting a close interdependence between gene expression and pSSVx replication (Contursi et al. 2007). In contrast, the investigation of the transcriptional activity of the cryptic plasmid pRN1, which, as for pSSVx, belongs to the pRN plasmid family, revealed a simple pattern of transcription for all the identified transcripts, the expression levels of which follow mainly the variation in the copy number of the pRN1 (Berkner and Lipps 2007).

Herein, the transcription termination in pSSVx was investigated by using the adaptor RT-PCR (Valk et al. 1997) and RNAse protection assay (Sambrook and Russell 2001) techniques. The results of this analysis provide further insights into the sequence requirement for transcription termination in *Sulfolobus* and predict the existence of an alternative mechanism relying on secondary stem-and-loop structures and counter-transcripts.

Materials and methods

Enzymes and chemicals

The MEGAscript® T7 kit, the RPAIII kit, the AMV reverse transcriptase and the Turbo™ DNase were purchased from Ambion. Synthetic oligonucleotides were supplied by TAG Copenhagen A/S. Radioactive materials were obtained from Perkin Elmer. The MinElute Reaction Cleanup Kit was from Qiagen.

Sulfolobus growth conditions and RNA extraction and manipulation

Sulfolobus islandicus REY15/4 (Zillig et al. 1998) was cultured with a rich medium under the growth conditions

previously described (Contursi et al. 2007). Cells were harvested at the early growth phase (26th hour) to obtain RNAs for determination of the end points of ctRNA1, ctRNA2, ctRNA3, ORFc68, ORF 76 and ORF60/91 transcripts, whereas cells harvested at the late growth phase (30th hour) were used for all the other mapping experiments (RNA4, ORF154/288, ORF60, ORF91/ORF892, ORF60/ORF91/ORF892). Total RNAs were extracted from cell samples by using the Trizol, Invitrogen reagent following the instructions of the manufacturer. DNA contaminants were removed from the RNA preparations by digestion with Turbo^{TM} DNase, Ambion and this procedure was repeated twice to ensure the complete removal of residual DNA.

Primer extension analysis

The precise determination of the transcriptional initiation site of RNA4 and ctRNA3 by oligonucleotide-primed reverse transcription of RNA was performed as described previously (Contursi et al. 2007), but using two- to three-fold more RNA template. RNA preparations from the early and late log phase were used as template for the primer extension experiments. The oligos used are shown in Table 1 of the supplemental material.

Adaptor RT-PCR

The termination regions of the pSSVx transcripts were determined by the adaptor RT-PCR method, which can effectively prevent the formation of false-positive amplicons derived from residual plasmid DNA in the RNA preparations (Berkner and Lipps 2007; Valk et al. 1997). As much as 0.5–1 μg of total RNA was reverse transcribed with Ambion M-MLV-reverse transcriptase (50 units). During reverse transcription, an adaptor primer, consisting of 20 nt complementary to the 3′ end of the target RNA and a 20-nt adaptor sequence was used to produce the cDNA molecules. After synthesis, the enzyme and the adaptor primer were removed from cDNAs using MinElute Reaction Cleanup Kit (Qiagen). To amplify these cDNAs, 5′ sequence-specific primers were used in conjunction with 3′ primers complementary to the adaptor sequence (the primers used are listed in Table 1 of the supplemented material). The transcript end is assumed to be located between the last primer that yielded a PCR product and the first primer that did not yield a product as described by Berkner and Lipps (2007).

RNAse protection assay

RNAse protection assay was employed to determine the 3′ ends of ORF60 using the RPAIII kit from Ambion. The pGEM-T Easy vector (Promega), bearing a sequence region partially encompassing both the ORF60 and ORF91 genes, was used to generate a uniformly labelled RNA probe (MEGAscript® T7 kit, Ambion). The labelled probe purified on preparative polyacrylamide gel as described by Reiter et al. (1988b) was mixed with 30 μg of REY 15/4 RNA. The digestion of the resulting hybrids with an RNAseA/RNAseT1 mix and analysis of the length of protected fragments on polyacrylamide sequencing gel was performed under the conditions suggested by the Ambion manufacturer.

Results

Compact organisation of pSSVx genes and identification of two new transcripts

Previously, nine transcripts had been identified for pSSVx in *S. islandicus* REY15/4 (Contursi et al. 2007). In this study, two new pSSVx transcripts both of which are non-coding RNAs were identified by primer extension (Fig. 2) and their sizes were determined by adaptor RT-PCR (see below). Therefore pSSVx has 11 genes, four of which are non-coding RNA genes. The complete transcriptional map of pSSVx is schematised in Fig. 1, in which all the identified transcripts as well as their location and relative orientation are indicated. The first newly identified transcript, designated counter-transcript 3 (ctRNA3), overlaps ORF76 and extends towards the opposite direction. The second one denoted RNA4 lies downstream of ORF288 and starts at the nucleotide immediately after ORF288 stop codon. Initiation nucleotides were determined for both RNAs. Transcription of the ctRNA3 gene starts from an adenine, whilst RNA4 initiates at a cytosine (Fig. 2). Although this does not fit well with the general pattern of gene transcription in *Sulfolobus* where RNAs usually start with either "A" or "G", transcripts and counter-transcripts of pSSVx, SIRV1 and pRN1 were also found to begin with "T" or "C" residues (Kessler et al. 2004; Berkner and Lipps 2007; Contursi et al. 2007), suggesting that the selection of transcription start sites is less stringent for transcripts produced from small compact genomes.

The promoter regions of RNA4 and ctRNA3 lie within the coding regions of ORF288 and ORF154, respectively. An A/T-rich motif centred at positions -23/-29 and comprising both the TATA box and a canonical BRE consensus sequence was found in the upstream region of the transcriptional start sites of the RNA4 transcript (Fig. 2). In contrast, the region upstream the start sites of ctRNA3 contains a canonical TATA box, but apparently no typical BRE motif. Several highly expressed viral/plasmidic genes/RNAs of the genetic elements pSSVx, SIRV1 and pRN1 lack a

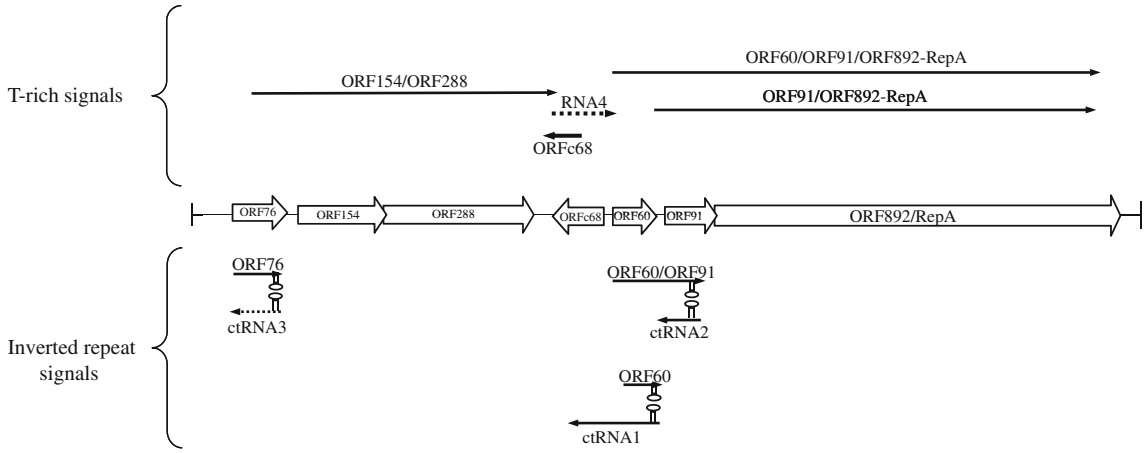


Fig. 1 Transcriptional map of pSSVx genes. The pSSVx transcripts previously mapped by Northern blot and primer extension (Contursi et al. 2007) are represented by *arrows* and are divided into two groups according to the termination signals identified. The newly identified RNA4 and ctRNA3 transcripts are represented by *dotted arrows*. Differential transcription termination is indicated by the relative length of the transcripts. The stem-and-loop structures as predicted by the Mfold programme 2.3 are shown. ctRNA1, ctRNA2 and ctRNA3 belong to the T-rich signal cluster. They are grouped together with the genes bearing the inverted repeat signals to show the putative "kissing complexes" (see text)

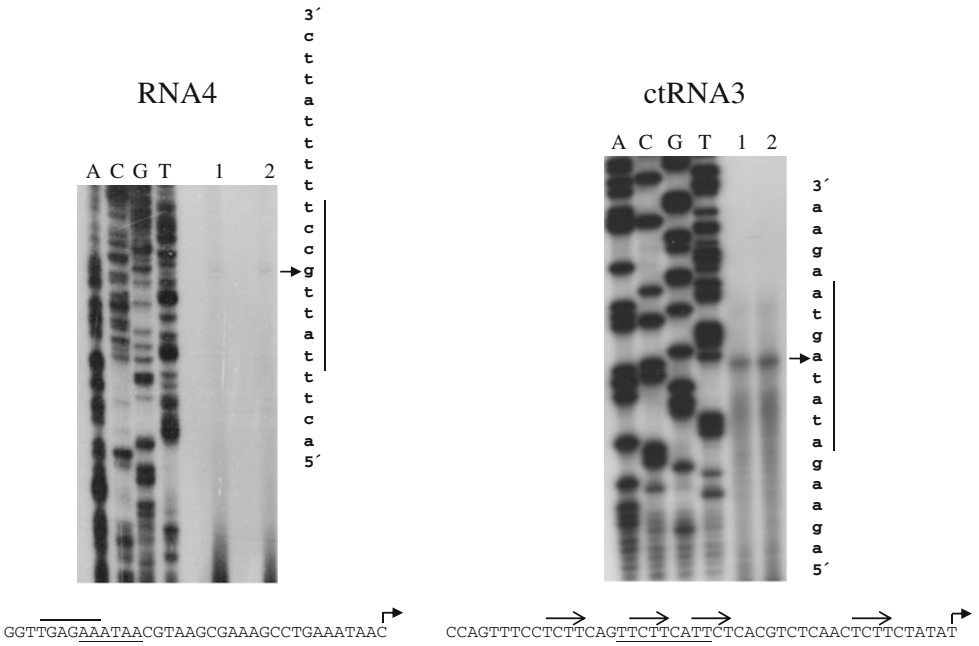


Fig. 2 Primer extension analysis of RNA4 and ctRNA3 transcripts**.** Total RNA isolated from REY15/4 cells was analysed by primer extension. The oligonucleotides used are RNA4prext and ctRNA3-prext (cf. Table 1 of supplemented material). The cDNA products were electrophoresed with a sequence ladder generated by a different primer on the pSSVx genome and the transcriptional start sites (indicated by the *arrows*) were deduced on the basis of the length of the correspondent sequence product. The putative TATA box and BRE are *under and overlined*, respectively. Direct repeats are highlighted by *dotted horizontal arrows*. Transcription initiation sites are indicated on the sequence by *bent arrows* above the corresponding nucleotides

typical BRE motif (Kessler et al. 2004; Berkner and Lipps 2007; Contursi et al. 2007), suggesting that motifs alternative to the conserved BRE elements can ensure efficient gene expression. Indeed, in vivo promoter analysis of the *araS* gene encoding an arabinose binding protein has revealed an alternative BRE in the promoter where both BRE and an upstream activation sequence element (*ara* box) are essential for a reporter gene expression (Peng et al. 2009). In this regard, the presence of four copies of the direct repeat TCTT in the promoter region of ctRNA3 (Fig. 2) mimics the *ara*-box element, suggesting that this RNA expression could be regulated as for the *araS* gene.

Apparently, a relatively short sequence in the pSSVx genome is responsible for high transcriptional activity. This sequence segment (924 bp) from the 3′ end of ORF288 to the 5′ end of ORF892 is very rich in compactly organised transcriptional signals, since it should enable transcription termination of several transcripts as well as allow transcription initiation of eight of the eleven RNAs identified (Fig. 1).

Identification of termination signals in pSSVx

Most of the in vitro and in vivo studies are consistent with the notion that archaeal termination is stimulated by oligo-T stretches even though in a context-dependent fashion (Reiter et al. 1988b; Santangelo and Reeve 2006; Santangelo et al. 2009). Thus, we searched for the presence of runs of 4 or more T residues within the first 200 bp of the 3′-flanking region: few oligo-T tracts, ranging from T4 to T12 were present at the 3′ end of most of the mono- and bi-cistronic units (i.e. ctRNA1, ctRNA2, ctRNA3, RNA4, ORFc68, ORF154/288 and ORF91/ORF892) and were unevenly interspersed within the first 150–180 bp of the predicted 3′ ends (Fig. S1).

On the other hand, the functionally clustered and/or closely spaced mono- and bi-cistronic units ORF 76, ORF60 and ORF60/ORF91 lacked the T-rich signal. Instead, they carried inverted repeats in the termination regions (Fig. S1). The "Mfold 2.3" (http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1-2.3.cgi) was used to predict the propensity of the inverted repeats to fold into stem-and-loop structures and to calculate the ΔG values at 75°C. These values can be used as a parameter for the stability of the stem-and-loop structures and hence for the likelihood of these structures causing pausing and/or transcription termination (Wilson and von Hippel 1995). Less negative values indicate weaker terminators that will probably result in an increased read through by the elongation complex, whereas more negative values point out to efficient terminators. For the inverted repeats present in the termination regions of the ORF 76, ORF60 and ORF60/ORF91 transcription, their ΔG values at 75°C vary from −0.59 to −3.07 kcal/mol (Fig. S2) and these values fall into the same order of magnitude of the stem-and-loop structures that have been identified as termination signals for the pRN1 transcripts (Berkner and Lipps 2007). This suggests that the identified stem-and-loop structures should function as transcriptional terminators in pSSVx.

To verify whether the identified DNA elements including the T-rich and the inverted repeat sequences functioned as termination signals in vivo, the adaptor RT-PCR method (Berkner and Lipps 2007) and the RNAse protection assay were used to map the 3′ termini of the transcripts. The adaptor RT-PCR was employed to determine the termination sites for all the pSSVx genes carrying the T-rich signal and for two (ORF76 and ORF60/91) of the three transcription units bearing the inverted repeat. The mapped 3′ termini of all the pSSVx transcripts together with their transcription start sites are listed in Table 1.

Transcripts ending at T-rich termination sequences

This set of transcripts includes the following mono- or bi-tri cistronic transcripts (ctRNA1, ctRNA2, ctRNA3, ORFc68, RNA4, ORF154/ORF288, ORF91/ORF892-RepA and ORF60/ORF91/ORF892-RepA). The adaptor RT-PCR experiments are exemplified with the analysis of the transcripts containing the ORF892-RepA gene encoding a putative polymerase/primase (Fig. 3, panel A) and with that of the ORF154/288 operon encoding proteins necessary for specific pSSVx DNA recognition and pre-packaging process (Fig. 3, panel B). Only the primers located upstream of each T-rich signal yielded a PCR product, whilst those located downstream of the signals failed to support any PCR activity (see Fig. S1 for the locations of the primers and of the T-rich sequences).

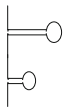
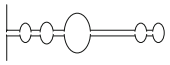
Comparing the sequence regions around the 3' termini of ctRNA1, ctRNA2, ctRNA3, RNA4, ORFc68, ORF154/288 and ORF91/ORF892, revealed the following findings: (i) the A+T content is skewed around the termination sites ($\cong$70% vs. the average value of 61.3%); (ii) these transcripts ended in the vicinity of a T-rich motif constituted by 8–12 thymidine residues that are either arranged as a continuous stretch or separated by few nucleotides (Table 1). In most cases, this main T-rich motif is preceded by 2–4 tri–tetra–penta–esa T-stretches that are unevenly distributed over a 150–180 bp region (Fig. S1). Possibly, these T-stretches function in concert to increase efficiency in transcriptional termination.

Transcripts ending at inverted repeats

This set includes functionally clustered and/or closely spaced genes (ORF76, ORF60 and ORF 60/91) for which the sequence analysis does not reveal any T-rich signals at their 3′ ends. ORF60 and ORF91, together with ORFc68 and ORF892/RepA, constitute the so-called *rep* locus of pSSVx. A complex transcription pattern involving premature termination and anti-termination as well as significant variations in the temporal expression of these genes have been previously observed and related to the control of replication and copy number of pSSVx in its natural host *S. islandicus* (Contursi et al. 2007).

ORF76 encodes a leucine zipper DNA-binding protein and is the most conserved gene in crenarchaeal conjugative and cryptic plasmids (Lipps et al. 2001). Northern blot analysis had demonstrated that ORF76 was always

Table 1 Adaptor RT-PCR and RNAse protection results

Transcript	Start position	Ending positions	Termination signals
ORF76	1095	1325/1405	
ORF154/288	1367	3091/3150	**ttttttcattttt**
ORFc68	3027	2676/2710	**ttatttttccgttattt**
ORF60	3150	3343	
ORF91/ORF892	3338	669/709	**tttttttctttt**
ctRNA1	3341	2710/2676	**ttatttttccgttattt**
ctRNA2	3557	3370/3324	**tttttactccatttt**
ctRNA3	1383	2023/1056	**tttcgcctttt**
RNA4	2686	3091/3150	**ttttttcattttt**
ORF60/91	3150	3550/3612	

Transcripts start sites mapped previously (Contursi et al. 2007) are shown as well as the termination ends determined by RT-adaptor PCR and RNAse protection assays. The putative transcription termination signals are indicated: the T-rich motifs are constituted by 8–12 thymidine residues and are arranged either on a continuous stretch or are separated by few nucleotides and the stem-and-loop structures found downstream the closely spaced and/or functionally clustered genes

expressed as a monocistronic messenger (Contursi et al. 2007). Hence, the 41-bp sequence interposed between the stop codon of ORF76 and the transcription start nucleotide of the operon ORF154/288 (Fig. S1) is able to arrest precisely the transcription of ORF76, as well as to drive the ORF154/ORF288 transcription. Interestingly, the adaptor RT-PCR revealed that the ORF76 transcript terminated in proximity of the transcription initiation site of ORF154/288 and in the same region where the oppositely oriented ctRNA3 also initiates (Fig. S1 and Fig. S2). The inverted repeat was the only identifiable sequence in the 41-bp intergenic region with potential to function as a termination signal because of the high propensity to fold into a stem-and-loop structure (Fig. S1).

ORF91 is situated in a tandem array partially overlapping with the ORF892-RepA (Arnold et al. 1999). The ORF91 protein contains a putative Zn motif similar to that found in nucleic acid (single and double strand RNA and/or DNA) binding proteins from plant RNA viruses (carlaviruses) and transcriptional activators of late genes in both coliphage and relative satellites (Contursi et al. 2007). Our previous results showed that ORF91 is transcribed either as bicistronic (ORF60/ORF91 and ORF91/ORF892) or tricistronic (ORF60/ORF91/ORF892) mRNAs (Fig. 1). The end position of the ORF60/91 transcript could be determined only on RNAs prepared from cells harvested at an early stage of growth. In fact at this growth phase, the read-through activity and the formation of the ORF60/ORF91/ORF892 transcript were negligible and hence did not interfere with PCR identification. It was found that the ORF60/ORF91 transcript terminated downstream of the predicted stem-and-loop structure and of the start site of the oppositely transcribed ctRNA2 (Fig. S1 and Fig. S2).

The ORF60 encodes a protein exhibiting significant sequence identity to CopG proteins, namely transcriptional regulators in the copy number control of several broad host-range bacterial plasmids (Arnold et al. 1999; del Solar et al. 2002). Transcriptional analysis had revealed that this gene was transcribed as mono-(ORF60), bi-(ORF60/ORF91) or tri-(ORF60/ORF91/ORF892) cistronic mRNAs (Fig. 1) (Contursi et al. 2007). In this case, the read-through at the putative termination region of the ORF60 transcript was always active throughout all the stages of the pSSVx cycle, precluding the application of the adaptor RT-PCR methodology. Thus, the RNase protection assay was alternatively employed to identify the termination site at

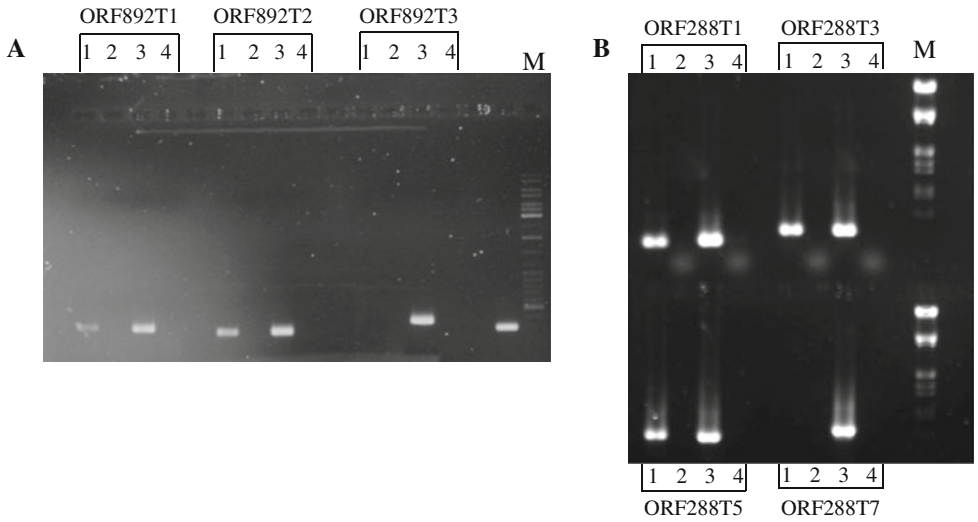


Fig. 3 Adaptor RT-PCR for the mapping of the 3′-end of ORF91/ORF892 and ORF154/ORF288 transcripts. **a** The RNA isolated from REY15/4 cells was reversely transcribed with three different adaptor primer, ORF892T1, ORF892T2 and ORF892T3, as described in the text. *Lane M* markers. *Lane 1* adaptor RT-PCR product, *lane 2* control without the adaptor pimer, *lane 3* PCR product as positive control, *lane 4* control without RT. The 3′-end of ORF91/ORF892 lies in the sequence region between the ORF892T2 and ORF892T3 primers (669–709 nucleotides of the linear map of pSSVx). **b** The same RNA used in the experiment of *panel A* was subjected to adaptor RT-PCR by using a set of 7 primers. Results of the analysis are shown only for four of the primers used. The end position of ORF154/ORF288 lies between nucleotides 3091 and 3150 of the linear map of pSSVx

the 3' end of ORF60. Total RNAs were prepared from *S. islandicus* REY15/4 and hybridised to an RNA probe uniformly labelled with ^{32}P-UTP. The 436-nt long riboprobe contained 96 bases of ORF60 3′-coding sequence, 243 bases of ORF91 5′-coding sequence, 31 bases of the intergenic sequence and 66 nucleotides matching the polylinker of the pGEM-T Easy vector (Fig. 4).

The three protected fragments (370, 243 and 134 nucleotides) were generated upon protection by the four transcripts ORF60, ORF60/ORF91, ORF91/ORF892 and ORF60/ORF91/ORF892 produced at the *rep* locus. Signal assessment was based on the mapping of the 5′ termini of all transcripts (Contursi et al. 2007). In fact, both the bicistronic ORF60/ORF91 and the tricistronic ORF60/ORF91/ORF892 mRNAs contributed to generate the 370-nt fragment, whereas the 243-nt signal corresponded to the portion of the riboprobe protected by the ORF91/ORF892 transcript. The shorter RNA fragment of 134 nt, matched the size of the ORF60 singular transcript ending at a specific termination site, i.e. 38 nt downstream of ORF 60 stop codon. This result is in agreement with the length of the ORF60 transcript previously estimated by Northern blot (ca. 150 nt). The variation of the relative abundance of the protected fragments in the different growth phases (Fig. 4) mirrors the differential temporal regulation of transcription of all RNA species at the *rep* locus (Contursi et al. 2007). The end point of ORF60 transcript mapped immediately downstream of two potential stem-and-loop structures and close to the initiation transcription sites of the ORF91/ORF892 transcript and of the reversely transcribed ctRNA1 (Fig. S1 and Fig. S2). The ladder of short fragments in the size range between 40 and 80 nt are likely due to the presence of these secondary structures that were also detected in our previous primer extension analysis (Contursi et al. 2007) and might act like termination intermediaries (see below).

In conclusion, ORF76, ORF 60 and ORF60/ORF91 showed three common features, i.e. (i) an inverted repeat, (ii) a counter-transcript starting in the vicinity of the 3′ ends, containing the inverted repeat and extending over most of the coding sequence, (iii) a purine–uridine-rich motif surrounding and including the stem-and-loop structure (Fig. S2). These elements suggest that transcription termination of these genes must take place through a mechanism different from that operating at the T-rich termination sequences.

ORF154/288 transcript carries a long untranslated tail

The partially overlapping ORF154 and ORF288 exhibited high sequence similarity to a tandem of ORFs present in the genomes of SSV viruses with identity scores of 63–75% for the homologues of ORF154 and 31–49% for those of ORF288. They have been earlier indicated as proteins necessary for specific pSSVx DNA recognition and pre-packaging process (Arnold et al. 1999; Stedman et al. 2003).

In contrast to all terminations described above, transcription termination was not found to occur in the region up

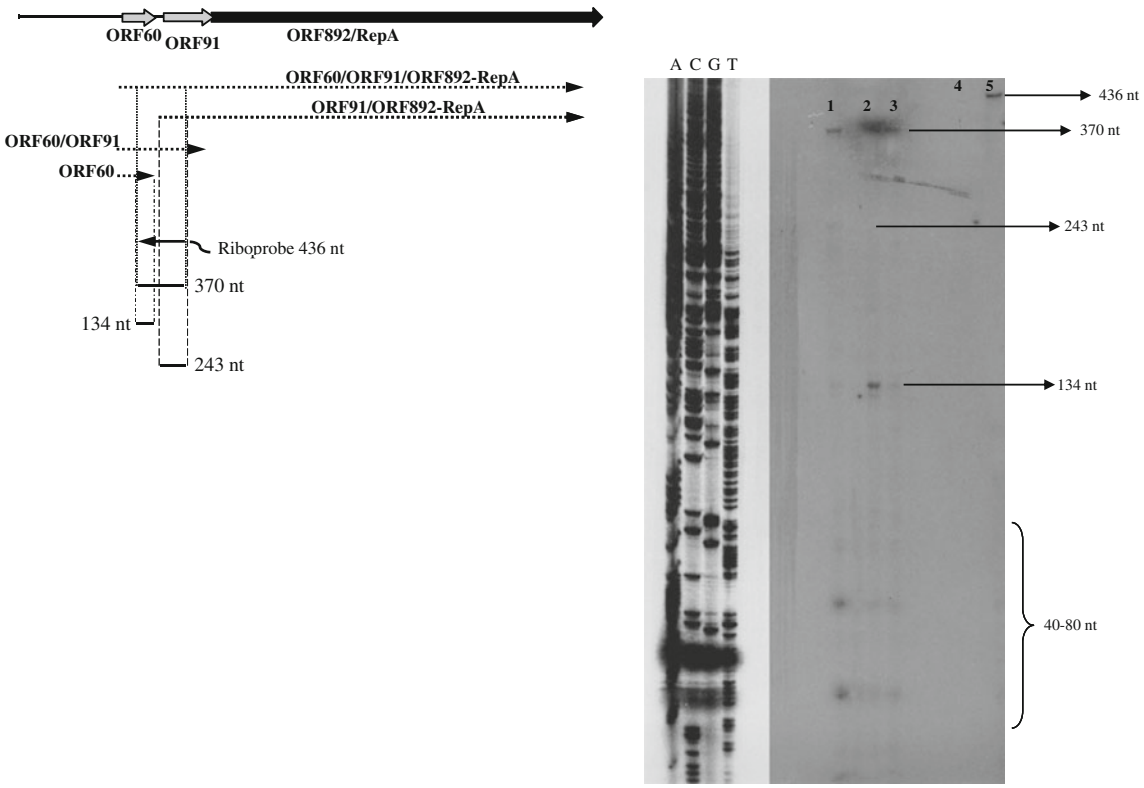


Fig. 4 Mapping of the 3′-end of ORF 60 by RNAse protection assay. Total RNAs prepared from *S. islandicus* REY15/4 before (26th incubation hour) and after (30th incubation hour) the pSSVx induction of replication sets were hybridised to an RNA probe uniformly labelled with ^{32}P-UTP. The location of the riboprobe is shown on the left. Lane 1 shows RNAse-generated fragments when the RNA from the 30th incubation hour was used. *Lanes 2–3* show RNAse-generated fragments when two different amounts of RNA from the 26th incubation hour were used. *Lanes 4 and 5* show controls lacking *Sulfolobus* RNA and undigested RNA probes, respectively. The size of the fragments (in nucleotides) were determined from DNA size markers run in parallel. The 436-nt long riboprobe contained 96 bases of ORF60 3′-coding sequence, 243 bases of ORF91 5′-coding sequence, 31 bases of the intergenic sequence of the two genes and 66 nucleotides matching the polylinker of the pGEM-T Easy vector. As indicated in the schematics, both the bicistronic ORF60/ORF91 and the tricistronic ORF60/ORF91/ORF892 mRNAs contributed to the 370-nt fragment, whereas the 243-nt signal corresponded to the portion of the ORF91/ORF892 transcript. The shorter RNA fragment of 134 nt matched the size of the ORF60 singular transcript ending at a discrete termination site, i.e. 38 nt downstream of ORF 60 stop codon

to 300 nt downstream of the ORF288 stop codon. More primers were then designed for the adaptor RT-PCR analysis (Table S1 of supplemented material) and this revealed that the ORF154/ORF288 transcript terminated between 406 and 466 nucleotides downstream of its stop codon and within a T-rich sequence correspondent to the promoter sequence of ORF60 (Fig. 1). Further, we assumed that the same termination point was also used for the RNA4. The region encompassing RNA4 sequence not only produces this autonomous transcript, but it also represents an extended 3′ tail of the ORF154/288 operon. This long 3′-untranslated region (approximately of 420 nt) is the only 3′UTR extending for more than 180 nucleotides beyond the stop codon amongst those of the pSSVx transcripts.

However, an oligo-T putative termination signal was also identified in a sequence region closer to the end of ORF288 (between 31 and 38 nt downstream of the stop codon). Therefore, the RNAase protection assay was employed to verify whether the transcription of ORF154–ORF288 stops also at this site. Unfortunately, the RNAase protection experiment resulted in a smear of multiple signals probably due to degradation activity affecting the ORF154/ORF288 transcript (Contursi et al. 2007), which did not allow to assess whether this site was a real termination site or not.

Discussion

Based on the sequence analysis data and the experimental evidence, pSSVx genes can be classified into two groups with regard to their transcriptional termination signals at the 3′ ends: genes showing alternatively: (i) T-rich stretches of more than 7 nucleotides or (ii) inverted repeats with high propensity to form stable stem-and-loop structures.

Only the first category of termination signals is consistent with previous reports describing that Archaea use the

intrinsic mechanism of transcription termination (Brenneis et al. 2007; Santangelo and Reeve 2006; Spitalny and Thomm 2008; Thomm et al. 1994).

In bacteria, an intrinsic termination signal appears as a hairpin followed by approximatively eight uridines at the 3′ terminus. The hairpin loop and the U-trail have distinct and complementary roles: the T-stretches induce pausing of the RNA polymerase near the termination point because of the unusual instability of the A:U DNA–RNA hybrid. Pausing of RNA pol at the T-trail then promotes the formation of the hairpin, resulting in unwinding of 4–5 base of the A:U hybrid adjacent to the hairpin stem. Consequently, the shortened hybrid A:U sequence destabilises the transcription elongation (TEC), leading to the complex disassembly and thereby transcriptional termination (Greenblatt 2008; Gusarov and Nudler 1999; Nudler and Gottesman 2002; Wilson and von Hippel 1995; Yarnell and Roberts 1999).

The transcription termination of the first group of pSSVx genes recalls the bacterial intrinsic mechanism, although only the presence of one of the two elements, i.e. the T-trail is sufficient for termination. In fact, Archaea share termination features more similar to transcription termination driven by eukaryotic Pol III, which pauses and subsequently releases the TEC when it simply encounters dT(n) tracts (Braglia et al. 2005; Gunnery et al. 1999; Hamada et al.2000). This mechanism is consistent with the chemical nature of nucleic acids because a U-rich RNA:DNA hybrid is significantly less stable than structurally similar dA:dT duplexes. As a result, a T-stretch sequence functions as a slippery sequence to facilitate transcript release (Martin and Tinoco 1980).

The pSSVx transcripts showing the T-rich motif at their 3′ termini are ctRNA1, ctRNA2, ctRNA3, RNA4, ORFc68, ORF154/288 and ORF91/ORF892, as determined by RT-PCR mapping. The T-rich sequence is preceded and/or followed by additional shorter T-stretches distributed over 150–180 bp at the 3′ terminus. Since reduced elongation rates have been correlated with increased termination efficiency for both the bacterial RNA pol and the eukaryotic pol III (Artsimovitch and Landick 2000; Lee et al. 1990, Powell and Reines 1996), these additional oligo d(T)-tracts reasonably act as "speed humps" by slowing down progressively the rate of the RNA pol, thus increasing the dwell time of RNA pol at the termination signal.

The second group of pSSVx genes, which includes the functionally clustered and/or closely spaced genes ORF76, ORF60 and ORF60/91, is distinguished by inverted repeats at their 3′ ends with potential to fold into a secondary structure and by a purine/uridine rich motif partially overlapping and/or immediately downstream of the inverted repeats (Fig. S1). Moreover, this motif comprises also the transcription start site of a counter-transcript.

The ORF76 shows all these features but constitutes a unique case, since it is always transcribed as a monocistronic RNA with a precise and invariant termination site. Homologues of ORF76 on the crenarchaeal plasmids pRN1, pRN2, pDL10, pHEN7 and pNOB show longer intergenic region separating their 3′ ends from the contiguous genes. Moreover, in contrast to ORF76, they all end with a T-stretch. Therefore, a horizontal gene transfer event occurred to insert the ORF154/288 operon downstream of ORF 76 and provided a new termination signal residing in the promoter of ORF154/288 operon. A spatial relationship between promoter regions and termination signals has been already pointed for SSV1 transcripts (Reiter et al. 1988b), confirming that initiation of transcription at a defined and canonical promoter of the downstream neighbourhood gene is effective for termination. In contrast to ORF76, downstream of ORF 60 and ORF 60/91 at the *Rep* locus, no canonical promoters (directing the expression of the tandem-arranged genes ORF91 and ORF892, respectively; see Fig. 1) exist. The absence also of other defined termination signals explains the balance between transcription termination and read-through previously demonstrated for these ORFs (Contursi et al. 2007).

The transcription termination and read-through take place in restricted regions and in close proximity of two distinct transcription initiation points: one relative to the contiguous collinear gene, and the other, inversely oriented, relative to a counter-transcript (ctRNA1 or ctRNA2, Fig. 1). The transcription interference (TI) resulting from convergent transcription (Crampton et al. 2006; Osato et al. 2007) can be evoked to explain this tuneable shift of termination towards anti-termination occurring at the *Rep* locus. In fact, according to the TI model, a significant proportion of RNAPs elongating ORF60 and ORF60/91 RNAs clash into the open promoter complexes of the counter-transcripts ctRNA1 and ctRNA2, respectively, thus stalling or pausing. Stalling of RNAP would favour the formation of a stem and loop in a fashion similar (and hence alternative) to the T-trail stretches. This secondary structure in turn exerts a drastic destabilising effect that renders the paused RNAP prone to termination and to disengagement from the template (Gusarov and Nudler 1999; Wilson and von Hippel 1995). At this point, RNAP molecules on the opposite strands can complete the elongation of the counter-transcripts from the stall site after this initial delaying. Conversely, the read-through of the ORF60 and ORF60/91 would prevail on their transcription termination when the rate of the transcription initiation of the relative counter-transcript is low. In agreement with this hypothesis, our previous result demonstrated that when the expression of the antisense ctRNA1 and ctRNA2 is switched off, the termination at the 3′ end of ORF60 and ORF60/ORF91 is less efficient allowing the termination read-through.

Nevertheless, the hairpins on the nascent mRNA have a low calculated intrinsic thermodynamic stability at the growth temperature of *Sulfolobus* (75°C) (Mitra et al. 2009) (Fig. S2) to fulfil the proposed function. Hybridisation between the complementary loops formed on the nascent mRNAs and their counter-transcripts might increase the stability of these hairpins through the so-called kissing complexes (Brantl 2002, 2007).

To sum up, the proposed mechanism is based on three main factors: formation of the stem and loop, transcription of ctRNAs and pausing of RNAP. Efficient termination occurs only when defined promoter sequences are located downstream of the stem-and-loop/purine–uridine-rich motif, as for ORF76. In the absence of such canonical promoters, transcription termination is tuneable and balanced by read-through, possibly as a consequence of transcriptional interference.

Because of the peculiar transcriptional pattern and regulation mechanisms occurring at the *rep* locus, pSSVx represents an interesting model system for the study of the mechanisms involved in the regulation of termination–anti-termination in Archaea. The identification of effectors, including possible protein terminator and/or anti-terminator factors, and investigation of their individual contributions are necessary to define the molecular basis of these mechanisms. The development of a shuttle vector based on the pSSVx (Aucelli et al. 2006) could help to understand whether single ORFs and regulatory sequences are relevant to the balance between termination and anti-termination and contribute to the pSSVx/SSV2 interplay.

Acknowledgments We are grateful to Simonetta Bartolucci for critically reading the manuscript. This work was aided by a grant from "Università Federico II di Napoli" in the framework of a programme named "short-term mobility" for researchers and by the "Danish Council for Independent Research: Technology and Production Sciences (grant number 274-07-0116 to Q.S.)".

References

Arnold HP, She Q, Phan H, Stedman K, Prangishvili D, Holz I, Kristjansson JK, Garrett R, Zillig W (1999) The genetic element pSSVx of the extremely thermophilic crenarchaeon *Sulfolobus* is a hybrid between a plasmid and a virus. Mol Microbiol 34:217–226

Artsimovitch I, Landick R (2000) Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. Proc Natl Acad Sci USA 97:7090–7095

Aucelli T, Contursi P, Girfoglio M, Rossi M, Cannio R (2006) A spreadable, non-integrative and high copy number shuttle vector for *Sulfolobus solfataricus* based on the genetic element pSSVx from *Sulfolobus islandicus*. Nucleic Acids Res 34:e114

Bartlett MS (2005) Determinants of transcription initiation by archaeal RNA polymerase. Curr Opin Microbiol 8:677–684

Bell SD, Jackson SP (2001) Mechanism and regulation of transcription in archaea. Curr Opin Microbiol 4:208–213

Berkner S, Lipps G (2007) Characterization of the transcriptional activity of the cryptic plasmid pRN1 from *Sulfolobus islandicus* REN1H1 and regulation of its replication operon. J Bacteriol 189:1711–1721

Braglia P, Percudiani R, Dieci G (2005) Sequence context effects on oligo(dT) termination signal recognition by *Saccharomyces cerevisiae* RNA polymerase III. J Biol Chem 280:19551–19562

Brantl S (2002) Antisense-RNA regulation and RNA interference. Biochim Biophys Acta 3:15–25

Brantl S (2007) Regulatory mechanisms employed by cis-encoded antisense RNAs. Curr Opin Microbiol 10:102–109

Brenneis M, Hering O, Lange C, Soppa J (2007) Experimental characterization of Cis-acting elements important for translation and transcription in halophilic archaea. PLoS Genet 3:e229

Contursi P, Jensen S, Aucelli T, Rossi M, Bartolucci S, She Q (2006) Characterization of the Sulfolobus host-SSV2 virus interaction. Extremophiles 10:615–627

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 189:6339–6350

Crampton N, Bonass WA, Kirkham J, Rivetti C, Thomson NH (2006) Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. Nucleic Acids Res 34:5416–5425

del Solar G, Hernandez-Arriaga AM, Gomis-Ruth FX, Coll M, Espinosa M (2002) A genetically economical family of plasmid-encoded transcriptional repressors involved in control of plasmid copy number. J Bacteriol 184:4943–4951

Geiduschek EP, Ouhammouch M (2005) Archaeal transcription and its regulators. Mol Microbiol 56:1397–1407

Greenblatt JF (2008) Transcription termination: pulling out all the stops. Cell 132:971–982

Gunnery S, Mathews MB (1995) Functional mRNA can be generated by RNA polymerase III. Mol Cell Biol 15:3597–3607

Gunnery S, Ma Y, Mathews MB (1999) Termination sequence requirements vary among genes transcribed by RNA polymerase III. J Mol Biol 286:745–757

Gusarov I, Nudler E (1999) The mechanism of intrinsic transcription termination. Mol Cell 3:495–504

Hamada M, Sakulich AL, Koduru SB, Maraia RJ (2000) Transcription termination by RNA polymerase III in fission yeast. J Biol Chem 275:29076–29081

Keeling PJ, Klenk HP, Singh RK, Feeley O, Schleper C, Zillig W, Doolittle WF, Sensen CW (1996) Complete nucleotide sequence of the *Sulfolobus islandicus* multicopy plasmid pRN1. Plasmid 35:141–144

Keeling PJ, Klenk H-P, Singh RK, Schenk ME, Sensen CW, Zillig W (1998) *Sulfolobus islandicus* plasmids pRN1 and pRN2 share distant but common evolutionary ancestry. Extremophiles 2:391–393

Kessler A, Brinkman AB, van der Oost J, Prangishvilli D (2004) Transcription of the rod-shaped viruses SIRV1 and SIRV2 of the hyperthermophilic archaeon sulfolobus. J Bacteriol 186:7745–7753

Kjems J, Garrett RA (1987) Novel expression of the ribosomal RNA genes in the extreme thermophile and archaebacterium *Desulfurococcus mobilis*. EMBO J 6:3521–3530

Kjems J, Leffers H, Garrett RA, Wich G, Leinfelder W, Böck A (1987) Gene organization, transcription signals and processing of the single ribosomal RNA operon of the archaebacterium *Thermoproteus tenax*. Nucleic Acids Res 15:4821–4835

Kletzin A, Lieke A, Urich T, Charlebois RL, Sensen CW (1999) Molecolar analysis of pDL10 from *Acidianus ambivalens* reveals a family of related plasmids from extremely thermophilic and acidophilic Archaea. Genetics 152:1307–1314

Korkhin Y, Unligil MU, Littlefield O, Nelson PJ, Stuart DI, Sigler PB, Bell SD, Abrescia NGA (2009) Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure. PLoS Biol 7:e1000102

Kornberg A, Baker T (1992) DNA replication. Freeman and Company, New York

Kusser AG, Bertero MG, Naji S, Becker T, Thomm M, Beckmann R, Cramer P (2008) Structure of an archaeal RNA polymerase. J Mol Biol 376:303–307

Kwapisz M, Beckouët F, Thriaux P (2008) Early evolution of eukaryotic DNA-dependent RNA polymerases. Trends Genet 24:211–215

Larsen N, Leffers H, Kjems J, Garrett RA (1986) Evolutionary divergence between the ribosomal RNA operons of *Halococcus morrhuae* and *Desulfurococcus mobilis*. Syst Appl Microbiol 7:49–57

Lee DN, Phung L, Stewart J, Landick R (1990) Transcription pausing by *Escherichia coli* RNA polymerase is modulated by downstream DNA sequences. J Biol Chem 265:15145–15153

Lipps G, Ibanez P, Stroessenreuther T, Hekimian K, Krauss G (2001) The protein ORF80 from the acidophilic and thermophilic archaeon *Sulfolobus islandicus* binds highly site-specifically to double-stranded DNA and represents a novel type of basic leucine zipper protein. Nucleic Acids Res 29:4973–4982

Martin FH, Tinoco I Jr (1980) DNA–RNA hybrid duplexes containing oligo(dA:rU) sequences are exceptionally unstable and may facilitate termination of transcription. Nucleic Acids Res 8:2295–2299

Mitra A, Angamuthu K, Jayashree HV, Nagaraja V (2009) Occurrence divergence and evolution of intrinsic terminators across Eubacteria. Genomics 94:110–116

Müller B, Allmansberger R, Klein A (1985) Termination of a transcription unit comprising highly expressed genes in the archaebacterium *Methanococcus voltae*. Nucleic Acids Res 13:6439–6445

Nudler E, Gottesman ME (2002) Transcription termination and antitermination in *E. coli*. Genes Cells 7:755–768

Osato N, Suzuki Y, Ikeo K, Gojobori T (2007) Transcriptional interferences in *cis* natural antisense transcripts of humans and mice. Genetics 176:1299–1306

Østergaard L, Larsen N, Lieffers H, Kjems J, Garrett R (1987) A ribosomal RNA operon and its flanking region from the Archebacterium *Methanobacterium thermoautotrophicum*, Marburg strain: transcription signals, RNA structure and evoultionary implications. Syst Appl Microbiol 9:199–209

Peng X, Holz I, Zillig W, Garrett RA, She Q (2000) Evolution of the family of pRN plasmids and their integrase-mediated insertion into the chromosome of the crenarchaeon *Sulfolobus solfataricus*. J Mol Biol 303:449–454

Peng N, Xia Q, Chen Q, Liang YX, She Q (2009) An upstream activation element exerting differential transcriptional activation on an archaeal promoter. Mol Microbiol 74:928–939

Powell W, Reines D (1996) Mutations in the second largest subunit of RNA polymerase II cause 6-azauracil sensitivity in yeast and increased transcriptional arrest in vitro. J Biol Chem 271:6866–6873

Prangishvili D, Stedman KM, Zillig W (2001) Viruses of the extremely thermophilic archaeon Sulfolobus. Trends Microbiol 9:39–42

Reiter WD, Palm P, Yeats S, Zillig W (1987) Gene expression in archaebacteria: physical mapping of constitutive and UV-inducible transcripts from the Sulfolobus virus-like particle SSV1. Mol Gen Genet 209:270–275

Reiter WD, Palm P, Zillig W (1988a) Analysis of transcription in the archaebacterium Sulfolobus indicates that archaebacterial promoters are homologous to eukaryotic polII promoters. Nucleic Acids Res 16:1–19

Reiter WD, Palm P, Zillig W (1988b) Transcription termination in the archaebacterium Sulfolobus: signal structures and linkage to transcription initiation. Nucleic Acids Res 16:2445–2459

Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Santangelo TJ, Reeve JN (2006) Archaeal RNA polymerase is sensitive to intrinsic termination directed by transcribed and remote sequences. J Mol Biol 35:196–210

Santangelo TJ, Cubonova L, Skinner KM, Reeve JN (2009) Archaeal intrinsic transcription termination in vivo. J Bacteriol 191:7102–7108

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 USA 89:7645–7649

She Q, Shen B, Chen L (2004) Archaeal integrases and mechanisms of gene capture. Biochem Soc Trans 22:222–226

Spitalny P, Thomm M (2008) A polymerase III-like reinitiation mechanism is operating in regulation of histone expression in archaea. Mol Microbiol 67:958–970

Stedman KM, She Q, Phan H, Arnold HP, Holz I, Garrett RA, Zillig W (2003) Relationships between fuselloviruses infecting the extremely thermophilic archaeon *Sulfolobus*: SSV1 and SSV2. Res Microbiol 154:295–302

Thomm M (2007) Transcription: mechanism and regulation. In: Cavicchioli R (ed) Archaea molecular and cellular biology, Chap 6. American Society for Microbiology Press, Washington, DC, pp 139–157

Thomm M, Hausner W, Hethke C (1994) Transcription factors and termination of transcription in methanococcus. Syst Appl Microbiol 16:648–655

Valk PJM, Joosten M, Vankan Y, Löwenberg B, Delwel R (1997) A rapid RT-PCR based method to isolate complementary DNA fragments flanking retrovirus integration sites. Nucleic Acids Res 25:4419–4421

Wich G, Hummel H, Jarsch M, Bär U, Böck A (1986) Transcription signals for stable RNA genes in methanococcus. Nucleic Acids Res 14:2459–2479

Wilson KS, von Hippel PH (1995) Transcription termination at intrinsic terminators: the role of the RNA hairpin. Proc Natl Acad Sci USA 92:8793–8797

Yarnell WS, Roberts JW (1999) Mechanism of intrinsic transcription termination and antitermination. Science 284:611–615

Zillig W, Prangishvili D, Schleper C, Elferink M, Holz I, Albers S, Janekovic D, Gotz D (1996) Viruses, plasmids and other genetic elements of thermophilic and hyperthermophilic archaea. FEMS Microbiol Rev 18:225-236

Zillig W, Arnold HP, Holz I, Prangishvili D, Schweier A, Stedman K, She Q, Phan H, Garrett RA, Kristjansson JK (1998) Genetic elements in the extremely thermophilic archaeon *Sulfolobus*. Extremophiles 2:131–140