Research paper

tRNA-like elements in *Haloferax volcanii*

Annette Hölzle\textsuperscript{a}, Britta Stoll\textsuperscript{a}, Thomas Schnattinger\textsuperscript{b}, Uwe Schöning\textsuperscript{b}, Brian Tjaden\textsuperscript{c}, Anita Marchfelder\textsuperscript{a,\*}

\textsuperscript{a} Biology II, Ulm University, Albert-Einstein-Allee 11, 89069 Ulm, Germany
\textsuperscript{b} Informatics, Ulm University, 89069 Ulm, Germany
\textsuperscript{c} Computer Science Department, Wellesley College, Wellesley, MA 02481, USA

\section*{1. Introduction}

All functional RNAs are generated from precursor molecules by a plethora of processing steps. The generation of mature RNA molecules by processing is an important layer of gene expression regulation catalysed by ribonucleases. Here, we analysed 5S rRNA processing in the halophilic Archaeon *Haloferax volcanii*. Earlier experiments showed that the 5S rRNA is cleaved at its 5' end by the endonuclease tRNase Z. Interestingly, a tRNA-like structure was identified upstream of the 5S rRNA that might be used as a processing signal. Here, we show that this tRNA-like element is indeed recognised as a processing signal by tRNase Z. Substrates containing mutations in the tRNA-like sequence are no longer processed, whereas a substrate containing a deletion in the 5S rRNA sequence is still cleaved. Therefore, an intact 5S rRNA structure is not required for processing. Further, we used bioinformatics analyses to identify additional sequences in *Haloferax* containing tRNA-like structures. This search resulted in the identification of all tRNAs, the tRNA-like structure upstream of the 5S RNA and 47 new tRNA-like structural elements. However, the \textit{in vitro} processing of selected examples showed no cleavage of these newly identified elements. Thus, tRNA-like elements are not a general processing signal in *Haloferax*.

\*Corresponding author. Tel.: + 49 731 5022658; fax: + 49 731 50 12 22658. E-mail address: anita.marchfelder@uni-ulm.de (A. Marchfelder).
molecules [10]. In addition, it seems that tRNAs are often duplicated and sometimes separated at the gene level into several pieces that are ligated together on the RNA level. Trans-splicing of tRNA halves has been shown, as well as the ligation of tRNA thirds [11–13].

Recently, a tRNA-like structure which might also be used as cleavage signal was identified upstream of a 5S rRNA gene in the halophilic archaeon Haloferax volcanii [14]. Whether or not more tRNA-like structures which could serve as processing signals are present in archaeal genomes is not known. The tRNA 5′ and 3′ processing enzymes have both been identified in archaea, tRNA 5′ ends are generated by the ubiquitous RNase P, and the tRNA 3′ end processing is catalysed by the endonuclease tRNase Z [15]. Besides RNase P and tRNase Z and the intron-splicing endonuclease [16–19], very few ribonucleases have been identified so far in archaea. Sequence comparisons with bacterial ribonucleases identified few homologues in archaea and, particularly in Haloarchaea, where only three more ribonucleases (RNase HII, RNase J and RNase R) were identified [20–24]. Biochemical approaches identified proteins which were not annotated as ribonucleases but which obviously moonlight as ribonucleases [25]. Thus, either Haloarchaea indeed encode only a few ribonucleases that have a broad substrate spectrum and are responsible for all processing events, or they have additional ribonucleases that are very specific for Haloarchaea and do not show any sequence conservation with known ribonucleases.

Except for the tRNA structure being a processing signal for RNase P and tRNase Z and the bulge-helix-bulge motif for the tRNA splicing endonuclease, no processing signals for RNA molecules have yet been identified in Archaea.

Here, we investigated whether the tRNA-like structure upstream of the 5S rRNA is recognised as a processing element. In addition, we explored the role of tRNA-like structures as general processing element in H. volcanii.

2. Results and discussion

Earlier experiments showed that the 5S rRNA 5′ end is generated by the tRNA 3′ processing enzyme tRNase Z in H. volcanii [14]. The tRNase Z enzyme is known to process tRNA precursors in order to generate a tRNA 3′ end ready for the addition of the terminal CCA.
sequence [26]. The tRNase Z has been shown to generally recognise the conserved tRNA three-dimensional structure but substrates containing only the acceptor stem and the T arm are also cleaved, albeit with a lower processing efficiency [27–30]. The 5' leader sequence upstream of the 5S rRNA can be folded into a tRNA-like structure that resembles a tRNA missing the anticodon stem and with a D replacement loop instead of a D arm (Fig. 1). In order to analyse whether or not the tRNA-like structure in the 5' leader sequence is recognised by the tRNase Z enzyme, we investigated several variants of the 5S rRNA precursor for processing by tRNase Z.

2.1. Deletion of the 5S rRNA structure is tolerated

First, we generated a variant in which the 3' part of the 5S rRNA gene was deleted (the last 77 bp of the 117 bp 5S rRNA gene) and which therefore could no longer fold into the 5S rRNA structure (Fig. 2A). This substrate was otherwise identical to the one that was shown to be processed by tRNase Z. Incubation of this 5S deletion variant with an S100 fraction showed that this substrate is efficiently processed into both the 5' leader and deleted-5S rRNA products (Fig. 2B). In addition, the variant was incubated with the recombinant tRNase Z from Haloferax, again showing that the variant was efficiently processed (Fig. 2C). Thus, deletion of part of the 5S rRNA sequence did not interfere with substrate processing. A complete 5S rRNA sequence that can fold into the typical 5S rRNA structure was not required, showing that the full 5S rRNA structure was not recognised by the processing enzyme.

2.2. Mutation of the 5' leader sequence results in a loss of processing

Next, we replaced the entire 5' leader sequence with a vector sequence (pUC18) that could not fold into a tRNA-like structure (Fig. 3A). This substrate was incubated with the S100 extract, but no processing products were visible (Fig. 3B). In addition, this

---

Fig. 3. The 5' leader sequence is necessary for processing. The 5' leader sequence was replaced by pUC18 vector sequence and incubated with an S100 protein extract or recombinant tRNase Z. A. Structure of the substrate, the sequence upstream of the 5S rRNA was replaced by pUC18 vector sequence. B. Incubation of the normal 5S rRNA substrate and the 5' leader mutant substrate with a protein extract from H. volcanii (S100) or recombinant tRNase Z. Whereas the normal 5S substrate (lanes 4 and 5, 5S) was efficiently processed, the 5' leader variant (lanes 1 and 2, 5leadvar) was not cleaved. The sizes of DNA markers in nucleotides are shown on the left. Precursor and products are shown schematically at the sides. Lanes 3 and 7 (c): control reactions without proteins, lane 6 (m): DNA size marker, lanes 2 and 4 (S): incubation with the S100 extract, lanes 1 and 5 (Z): incubation with recombinant tRNase Z.
substrate was also incubated with tRNase Z (Fig. 3B), but again no processing was observed, confirming that the 5S rRNA structure, which was still present in this substrate, was not recognised by the processing enzyme and indicating that the original 5’ leader sequence was required for processing.

2.3. Point mutations preventing folding into a tRNA-like structure

In order to show that not only the primary but also the secondary and tertiary structures of the 5’ leader sequence are required for efficient processing, we introduced point mutations into the leader sequence predicted to interfere with folding into a tRNA-like element (Fig. 4A). The first variant had two point mutations in the central part of the acceptor stem predicted to prevent the RNA from folding into the tRNA-like structure (Supplementary Fig. 1A). *In vitro* processing with this variant showed that it is processed with a drastically reduced efficiency, as the processing products were barely visible (Fig. 4B, lane v1). For the second variant, point mutations were introduced into the T stem (Fig. 4A), predicted to prevent formation of the T stem and to result in a different structure (Supplementary Fig. 1B). Processing assays with this variant showed that it also was processed inefficiently (Fig. 4B, lane v2). In a third variant, mutations were introduced into the D replacement loop (Fig. 4A). These mutations did not interfere with the formation of the two-dimensional structure (Supplementary Fig. 1C), but since the two G nucleotides in the D loop had been removed a tertiary interaction between these conserved nucleotides and the nucleotides U and C in the T loop (Supplementary Fig. 2) would no longer be possible; therefore, the typical 3D tRNA structure should not be formed anymore. Again, processing assays showed that this substrate was indeed unprocessed (Fig. 4B, lane v3). Moreover, none of the three variants was cleaved upon incubation with an extract of S100 (data not shown). Thus, taken together, these data suggest that folding of the 5’ leader sequence into a tRNA-like structure is required for recognition by tRNase Z. Earlier studies have shown, that tRNase Z enzymes require the conserved tRNA three-dimensional structure for processing but substrates containing only the acceptor stem and the T arm are also cleaved, albeit with a lower processing efficiency [27–30].

2.4. Are tRNA-like elements general processing signals in *Haloferax volcanii*

The tRNAs and tRNA-like structures are known, from mitochondria, to be processing signals for converting primary transcripts and precursor RNAs into functional molecules. After detecting the tRNA-like element upstream of the 5S rRNA and showing that the tRNA structure is used as a processing signal, we were interested in analysing whether tRNA-like structures are used as general processing signals in *Haloferax*. To our knowledge a systematic search for tRNA-like elements has not been done to date in any archaeal organism. Therefore, using two bioinformatics approaches we searched the *H. volcanii* genome for tRNA-like structures.

The first approach screened for hairpin structures consisting of an unpaired loop of seven nucleotides corresponding to the sequence GTPGCNNNC, a G:C base pair closing the loop, and a stem of at least four consecutive stacking base pairs (Supplementary Fig. 3A). This search identified 85 elements (Supplementary Table 2A); 36 of these were parts of tRNA genes, confirming that the program was able to identify sequences that contained the elements requested. Two hits were located upstream of the two 5S rRNA genes, identifying the tRNA-like structure upstream of the two 5S rRNA genes, leaving 47 new elements. These elements were analysed regarding their location with respect to protein and RNA coding genes. If they were indeed used as processing signals they would have been expected to be located near the upstream or downstream boundaries of genes. Furthermore, the sequences were analysed using the Vienna RNA Secondary Structure Folding program (rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) to investigate whether or not they could fold into tRNA-like structures. Secondary structure predictions for most of the predicted elements showed that the T arm (T stem of five base pairs and T arm containing the conserved nucleotides, see Fig. 1) folded correctly (compare with Supplementary Fig. 4), but none of the predicted t-elements folded into a complete tRNA structure or a tRNA-like structure. Since our computational prediction of secondary structures cannot take into account tertiary interactions, we decided to test in *in vitro*
processing assays the abilities of some of the predicted elements to act as processing sites. From the 47 sequences eight were selected, based on their location to be investigated by in vitro processing assays (Table 1, Supplementary Figure 4). Element F15 was used for in vitro processing since it overlaps by 23 nucleotides Hvo_1148 (coding for the ribosomal protein S15) such that processing at the 3' end of F15 would generate a 5' truncated mRNA, where the second ATG (which is located nine nucleotides downstream of the first ATG) would have to be used to start translation. Element F22 was located in an antisense fashion to the 3' end of the mRNA for a hypothetical protein (Hvo_0896). Element F28 was located in an antisense fashion to the 3' end of Hvo_0896 (acyl-CoA ligase) and Hvo_0525 (hypothetical protein). Element F79 was located sense to 3' end of Hvo_0366 (hypothetical protein). Element F81 was located upstream of the ISH51 transposase mRNA (Hvo_B0152). Processing at this element could generate a monocistronic ISH51 mRNA out of a potentially longer precursor mRNA. Element R11 overlapped with the 3' end of the mRNA for the signal-transducing histidine kinase homologue (Hvo_1414). If processing occurred at this element, the mRNA would be processed 29 nucleotides downstream from the stop codon. Element R56 was located in the 3' UTR of the sporulation regulator homologue (Hvo_2804). Element R78 was located in the 5' third of the mRNA for the hydrolytic enzyme Ip1D (Hvo_B0343). Processing at this site would generate an mRNA truncated by 260 nucleotides at the 5' end. The genes for the eight selected elements including the up- and downstream sequences were cloned and transcribed into substrates for the processing assays.

Table 1
Eight t-elements identified in the first bioinformatics approach were chosen for in vitro processing assays. The first bioinformatics approach identified 85 elements in the genome of Haloferax (all 85 sequences are listed in Supplementary Table 2). Eight of these (F15, 22, 28, 79, 81 and R11, 56 and 78) were selected to generate DNA templates for generating the substrates for in vitro processing.

<table>
<thead>
<tr>
<th>t-Element</th>
<th>Gene upstream</th>
<th>Distance to upstream gene (bp)</th>
<th>Gene downstream</th>
<th>Distance to downstream gene (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F15</td>
<td>tR10-AlaGGC</td>
<td>app. 200</td>
<td>Hvo_1148 (ribosomal protein S15)</td>
<td>overlaps with 5' end of orf</td>
</tr>
<tr>
<td>F22</td>
<td>located antisense to HVO_0896 (acyl-CoA ligase)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F28</td>
<td>located antisense to HVO_0525 (hypothetical protein)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F79</td>
<td>located sense to 3' end of HVO_0366 (hypothetical protein)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F81</td>
<td>HVO_B0153 (Mur ligase family Capβ protein)</td>
<td>app. 500</td>
<td>HVO_B0152 (transposase ISH51)</td>
<td>app. 130</td>
</tr>
<tr>
<td>R11</td>
<td>located sense to 3' end of HVO_1414 (signal-transducing histidine kinase homologue)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R56</td>
<td>HVO_2804 (sporulation regulator homologue)</td>
<td>app. 40</td>
<td>HVO_2805 (hypothetical protein)</td>
<td>app. 20</td>
</tr>
<tr>
<td>R78</td>
<td>located antisense to HVO_B0343 (hydrolytic enzyme Ip1D)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2
Eight t-elements identified in the second bioinformatics approach. A second, more stringent, bioinformatics approach identified three elements.

<table>
<thead>
<tr>
<th>t-Element</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>tele1</td>
<td>3' part of tRNA_{Pro}^{GGG} (GGG)</td>
</tr>
<tr>
<td>tele2</td>
<td>part of mRNA for the htr like protein (Hvo_2870)</td>
</tr>
<tr>
<td>tele3</td>
<td>part of the mRNA for a hypothetical protein (Hvo_A0333)</td>
</tr>
</tbody>
</table>

![Fig. 5. t-element structures. Three sequences which match a given tRNA-like structure were identified by a stringent search using the RNAMotif tool (A,C,E). These sequences were then folded using RNAfold from the Vienna RNA Package (B,D,F).](image-url)
substrates were not processed by either of the protein fractions (data not shown).

Since this rather broad search for t-elements using only the T arm structure as a search criterion was not successful, we next tried a more stringent search. Therefore, the second bioinformatics approach searched for regions that not only matched the T arm, but also contained an adjacent bulge that we did not further specify (just its maximal length of 20 nucleotides; the so-called D replacement loop), and an enclosing stem consisting of seven base pairs (Suppl. Fig. 3B).

This approach identified three potential t-elements (Table 2, Supplementary Table 2B). The first element was located on the chromosomal plasmid pHV4 and was identical to the 3′ part of the tRNAAsp (GGG). Processing at this element would generate the normal tRNA 3′ end (Supplementary Fig. 5A). The prediction program folded the sequence into the tRNA-like structure (Fig. 5A), whereas folding of this sequence with the Vienna RNA folding program yielded a different structure (Fig. 5B). The second sequence identification was located on the main chromosome and was part of the mRNA for the htr-like protein. The prediction program folded the sequence into the tRNA-like structure (Fig. 5C), but folding of this sequence with the Vienna RNA folding program yielded a different structure (Fig. 5D). Processing at this element would cleave the mRNA (Supplementary Fig. 5B). The third element was also located on the main chromosome and was part of the mRNA for a hypothetical protein. Both structure folding programs folded the sequence into the tRNA-like structure (Fig. 5E and F). Processing at this element would cleave the mRNA in its middle (Supplementary Fig. 5C). The genes for these t-elements were isolated and transcribed into precursor RNAs, which were subsequently incubated with a Haloferax S100 and recombinant tRNase Z. In vitro processing assays with S100 extracts and recombinant tRNase Z under different conditions showed that these substrates were not processed by either of the protein fractions (data not shown).

3. Conclusions

Processing of the 5′ end of the 5S rRNA requires a tRNA-like structure in the upstream leader sequence, which is recognised by the tRNase Z protein. Mutations predicted to inhibit folding into this structure prevent processing. Thus, in this case, a tRNA-like structure is used as a processing signal to generate the mature 5S rRNA 5′ end. However, our experiments showed that tRNA-like elements are not used as general processing signals in Haloferax. Sequences that had conserved tRNA characteristics were identified in the genome, but they did not seem to fold in tRNA-like elements and were not processed in vitro. The sequence upstream of the 5S rRNA might be a remnant tRNA gene that was reduced from a full tRNA into a tRNA-like structure that retained enough tRNA-like structural elements to still be recognised by the tRNase Z and act as a processing signal.

4. Materials and methods

Strains and culture conditions and isolation of chromosomal DNA. H. volcanii strain H119 (ΔpyrE2, ΔtrpA, ΔleuB) [31] was aerobically grown at 45 °C in Hv-YPC or Hv-Min medium [31]. Chromosomal DNA from H. volcanii was isolated using the alternative rapid chromosomal isolation method as published in the Halohandbook (www.haloarchaea.com/resources/halohandbook/Hallohandbook_2008_v7.pdf).

Substrate preparation. Templates for 5′ extended and 3′ matured 5S rRNA precursors from H. volcanii were synthesized by PCR from clone pUC18-5SI [14] using the HvoSS177 and HvoSS4rev primers (for primer sequences see Supplementary Table 1). The resulting template pHvSS-5 contained the T7 promoter, the 55 rRNA gene (117 bp) and the 5′ leader (66 bp). The deletion variant of the 55 rRNA gene was generated by PCR on the pUC18-5SI template using HvoSS177 and HvoSSp primers, generating a DNA fragment containing the coding sequence for the 5′ leader (66 bp) and the 5′ part of the 55 rRNA gene (40 bp). For mutation of the 5′ leader sequence, the sequence was replaced by the pUC18 sequence by first deleting the upstream sequence of the 55 rRNA in clone pUC18-5SI using inverse PCR on clone pUC18-5SI with the primers SS3 and repvUC18SS. The resulting plasmid pUC18-5SI contained the 55 rRNA gene and downstream sequence. This plasmid was used as a template in a PCR reaction with primers T7gpUC18SS (which contained the T7 promoter sequence for in vitro transcription) and 5SS4. The resulting DNA was used for in vitro transcription, which yielded an RNA with a 66 nucleotide upstream sequence (pUC18 sequence) and 55 rRNA.

The templates for the substrates containing point mutations of the 5′ leader sequence (variant1, variant2 and variant3) were generated using the Stratagene QuickChange site-directed mutagenesis kit, according to the manufacturer’s instructions, and the clone pUC18-5SI (see above [14]) was used as a DNA template. The primers used for the mutations are listed in Supplementary Table 2. The resulting plasmids 5S-variant1, 5S-variant2 and 5S-variant3 were used as templates in the PCR reactions with the primers HvoSS177 and HvoSS4rev. The resulting PCR products were used for in vitro transcription.

The templates for the t-element substrates were generated by PCR using H. volcanii genomic DNA and two specific primers for each t-element (for primer sequences see Supplementary Table 2), and the resulting PCR products were cloned into pUC18, which was digested with SmaI. For in vitro transcription, templates were generated by PCR using two primers, where the 5′ primer contained a T7 RNA polymerase sequence. The templates contained the t-element and additional up- and downstream sequences (lengths of the up- and downstream sequences are shown in Supplementary Table 3).

In vitro transcription and purification of the transcripts were performed as previously described [32].

Isolation of S100 protein extract. H. volcanii cells were grown to an OD600 of 0.6–0.8 in 1 l medium. The cells were collected by centrifugation at 7,500 × g for 15 min at 4 °C. After washing of the cell pellet with buffer C (50 mM Tris–HCl pH 7.5, 5 mM KCl, 5 mM MgCl2), the cells were disrupted by sonication and a high-speed supernatant (S100) was obtained by ultracentrifugation at 100,000 × g for 60 min at 4 °C. The S100 fraction was dialysed against buffer E (50 mM Tris–HCl pH 7.5, 5 mM MgCl2).

Expression of HvoTrz in E. coli. The gene for the Haloferax tRNase Z enzyme was previously isolated and cloned into the expression vector pET29a (Novagen). The PET29a-hvoz was transformed into the strain Rosetta (DE3) pLYS (Novagen) and expressed and purified according to the manufacturer’s protocol using S-protein agarose (Novagen).

In vitro processing assays. The processing assay was performed with 100 ng HvoTrz or 10 μg S100 extract in 100 μl hvoz buffer (40 mM MES pH 5.5, 5 mM KCl, 5 mM MgCl2 and 2 mM DTT) in a volume of 100 μl at 37 °C for 30 min. The processing products were precipitated and separated on 8% polyacrylamide gels.

4.1. Bioinformatics prediction: first approach

A combination of primary sequence and secondary structure information was used to search the genome of H. volcanii for motifs indicative of tRNA-like processing signals. Candidate tRNA-like processing signals were identified throughout the genome by searching for hairpin loop structures (Supplementary Fig. 3A)
consisting of seven unpaired nucleotides in the loop corresponding to the primary sequence pattern of 5GTCNANC, a G:C base pair closing the loop, and at least four stacking base pairs in the stem of the hairpin loop. The five base pairs, four in the stem in addition to the G:C base pair closing the loop, were required to be uninterrupted by bulging or interior loops. The genome-wide screen yielded 85 candidates, 36 of which corresponded to tRNAs and two of which corresponded to the upstream region of the 5S rRNA (Supplementary Table 2A).

4.2. Bioinformatics prediction: second approach

For the in silico prediction of further processing sites we used the RNAMotif tool [33], which allows to search the genome for sequences matching a given secondary structure. Therefore we specified the secondary structure of the t-elements to consist of the T arm, an adjacent bulge of up to 20 nucleotides, and an enclosing stem of seven base pairs. For future work, we developed a computer program that guides the researcher throughout the whole process of identifying new RNA elements. It integrates several available tools for the abovementioned search, minimum-free-energy folding of the sequences found, automatic annotation of the secondary structure, and protection of mRNA in Crenarchaeota, RNA 17 (2011) 99–107.

Acknowledgements

We would like to thank Elli Bruckbauer for expert technical assistance. Furthermore, we thank the members of the SPP 1395 Priority Program “inKomBio” for helpful discussions. In addition we appreciate the time the reviewers took for reviewing the manuscript and for their constructive criticism. This work was supported by the Deutsche Forschungsgemeinschaft through a grant to A.M. (MA1538/12-1) and the SPP 1395 Priority Program “inKomBio” grants to A.M. (MA1538/14-1) and U.S. (Scho 302/8-1).

Appendix. Supplementary material


References