Experimental RNomics in *Aquifex aeolicus*: identification of small non-coding RNAs and the putative 6S RNA homolog

Dagmar K. Willkomm, Jens Minnerup\textsuperscript{1}, Alexander Hüttenhofer\textsuperscript{2} and Roland K. Hartmann* \\

Philipps-Universität Marburg, Institut für Pharmazeutische Chemie, Marbacher Weg 6, D-35037 Marburg, Germany, \textsuperscript{1}Institute for Experimental Pathology, ZMBE, University of Münster, Von-Esmarch-Strasse 56, D-48149 Münster, Germany and \textsuperscript{2}Institute for Molecular Biology, Division of Functional Genomics and RNomics, Innsbruck Medical University, A-6020 Innsbruck, Austria

Received December 15, 2004; Revised and Accepted March 16, 2005

ABSTRACT

By an experimental RNomics approach, we have generated a cDNA library from small RNAs expressed from the genome of the hyperthermophilic bacterium *Aquifex aeolicus*. The library included RNAs that were antisense to mRNAs and tRNAs as well as RNAs encoded in intergenic regions. Substantial steady-state levels in *A.eolicus* cells were confirmed for several of the cloned RNAs by northern blot analysis. The most abundant intergenic RNA of the library was identified as the 6S RNA homolog of *A.eolicus*. Although shorter in size (150 nt) than its \(\gamma\)-proteobacterial homologs (~185 nt), it is predicted to have the most stable structure among known 6S RNAs. As in the \(\gamma\)-proteobacteria, the *A.eolicus* 6S RNA gene (ssrS) is located immediately upstream of the *ygfA* gene encoding a widely conserved 5-formyltetrahydrofolate cyclo-ligase. We identified novel 6S RNA candidates within the \(\gamma\)-proteobacteria but were unable to identify reasonable 6S RNA candidates in other bacterial branches, utilizing mfold analyses of the region immediately upstream of *ygfA* combined with 6S RNA blastn searches. By RACE experiments, we mapped the major transcription initiation site of *A.eolicus* 6S RNA primary transcripts, located within the *pheT* gene preceding *ygfA*, as well as three processing sites.

INTRODUCTION

*Aquifex aeolicus* is a microaerobic, hyperthermophilic bacterium that grows at temperatures of up to 95°C. The genus *Aquifex* has been proposed to be the deepest branch of the bacterial phylogenetic tree on the basis of 16S rRNA as well elongation factor Tu and G phylogenies (1–3). However, a closer relationship of *Aquifex* to the \(\delta/\epsilon\) division of proteobacteria and the Chlamydia–Cytophaga group of bacteria is suggested by phylogenies based on RNA polymerase \(\beta, \beta’\) and \(\sigma^{70}\) subunits (4,5) as well as conservation analyses of small insertions and deletions in a variety of proteins (6).

The approach termed experimental RNomics (7) has laid the foundation for the boom-like discovery of novel non-messenger RNAs in very recent years [(8) and references therein]. Our motivation to apply this method to *A.eolicus* was fueled by unsuccessful attempts to identify a ribonuclease P (RNase P) enzyme in this bacterium. RNase P, an ubiquitous ribonucleoprotein enzyme that catalyzes tRNA \(5’\) end maturation in all kingdoms of life, is composed of a single protein and a catalytic RNA in bacteria, with no exceptions known so far. In *A.eolicus* however, neither have candidate genes for the protein (*rnpA*) and RNA subunit (*rnpB*) of RNase P been identified in the sequenced genome (9,10), nor have biochemical experiments revealed any RNase P (or RNase P-like) activity in cell lysates (11). From the idea that a putative *A.eolicus* RNase P RNA might have escaped detection due to genome sequencing mistakes or functional idiosyncrasies, we scrutinized a cDNA library generated from small RNAs in the size range of 100–450 nt (8,12). While no putative RNase P RNA candidate could be revealed, the presented experimental RNomics study of *A.eolicus* has identified a variety of expressed sequence tags, which comprise fragments of mRNA encoding protein genes of unknown function, novel non-messenger RNA candidates encoded in intergenic regions, and potential antisense RNAs including tRNA antisense transcripts. Strikingly, ~10% of the cDNA library clones—almost twice as many as those derived from tmRNA—represented an RNA originating from the intergenic region between the...
were grown as described (9). Together with similarities in sequence and structure to known bacterial 6S RNAs, this led us to conclude that we have identified the putative 6S RNA homolog of *A. aeolicus*. By 5′-RACE, we further identified its major transcription initiation as well as processing sites. Processing of the primary transcript generates a mature product predicted to form the rod-like secondary structure with a central bulge region typical of bacterial 6S RNAs.

**MATERIALS AND METHODS**

**Bacterial strain**

*A. aeolicus* cells, kindly provided by R. Huber and K.O. Stetter, were grown as described (9). *A. aeolicus* liquid cell cultures were harvested in the early stationary growth phase (0.5–1.0 × 10^10^ cells per ml; Robert Huber, personal communication) to compensate for the low yields of cell mass obtained from laboratory cultures of *A. aeolicus*.

**Construction of a cDNA library encoding small stable RNAs of *A. aeolicus***

A cDNA library was constructed from *A. aeolicus* total RNA essentially as described previously (14). The size range of RNAs excised from a preparative polyacrylamide gel for library construction covered 100–450 nt, excluding the 5S rRNA band (~120 nt).

**Analysis of the cDNA library**

Initially, 42 clones were sequenced, which allowed us to identify cDNAs encoding abundant RNA species, that is fragments of 5S, 16S and 23S rRNA, tmRNA, of pre-mature tandem tRNA^Ile–tRNA^Ala transcripts encoded in the 16S–23S spacer of the two rRNA operons, and fragments from the intergenic region between *pheT* (β-subunit of phenylalanyl-tRNA synthetase) and the open reading frame *aq_1731* (13). Together with similarities in sequence and structure to known bacterial 6S RNAs, this led us to conclude that we have identified the putative 6S RNA homolog of *A. aeolicus*. By 5′-RACE, we further identified its major transcription initiation as well as processing sites. Processing of the primary transcript generates a mature product predicted to form the rod-like secondary structure with a central bulge region typical of bacterial 6S RNAs.

**Plasmid preparation**

For the dot blot analysis, minipreparations of plasmid DNA harboring cDNA inserts were performed by the alkaline lysis method according to (15), whereas a commercial kit (Concert Rapid Plasmid Miniprep System, Gibco-BRL) was employed to prepare high purity plasmid DNA for sequencing purposes.

**Dot blot experiments**

Of each plasmid DNA, ~300 ng were spotted onto positively charged nylon membranes (Roche) for hybridization with digoxigenin-labeled oligonucleotide or PCR probes according to the ‘DIG Application Manual for Filter Hybridization’ (Roche). The following oligonucleotides, 5′ end-labeled with digoxigenin, were used as probes:

**DIG 5S rRNA**

*(Aeolicius) 5S rRNA, positions 92–120):

5′-GGCACGGGAAAGTAGTGTCGTGCGAGGGG

**DIG 16S rRNA**

*(Aeolicius) 16S rRNA, positions 1521–1550):

5′-CCGGCGACTGGGGGGAAGTCTGTAACAAGG

**DIG 23S rRNA**

*(Aeolicius) 23S rRNA, positions 2921–2952):

5′-CCGAGCGGTACTAATCGCCGTCCAACTCC-AGG

**DIG *pheT*-aq1731**

*(Aeolicius) genome nt 1219836–1219866):

5′-AAAGCTCTGAGGCCACGCACTTCC-TGC

**DIG tmRNA**

*(Aeolicius) genome nt 1153699–1153724):

5′-ACCCGCAAACCTACCAGGGCAGC

**DIG rRNA^he-^RNA^Ala**

*(Aeolicius) genome nt 1193812–1193844 and 571042–571010):

5′-GGTTCAGGTCCTGGG-GAGGCACATATTAGGGG

Furthermore, two PCR probes were generated by addition of 0.03 mM digoxigenin-dUTP (Roche) to a standard Taq DNA polymerase (Stratagene) PCR reaction using *A. aeolicus* genomic DNA as template. These probes covered 16S rRNA positions 53–776 (primers: 5′-ACACATGCAAGTCTG-GGCC and 5′-GGACAGCCCAGCAGG) and 23S rRNA positions 25–787 (primers: 5′-TGGATGCTCGTCTC-CCC and 5′-GCCCATACCAGGGCA).

Spotted DNAs were cross-linked to the membranes by exposure to UV light (312 nm, 0.25 J/cm²), followed by pre-hybridization of membranes in 5× SSC, 0.1% sodium lauroyl sarcosinate, 0.02% SDS, 1.5% blocking reagent (Roche) for 1 h at 60–70°C, and hybridization for 12 h in the same buffer with 10 pmol of probe per ml at 60–65°C (for the oligonucleotides) or 70°C (for the PCR probes). Washing was performed twice for 5 min with 2× SSC/0.1% SDS at room temperature, followed by two additional washes for 15 min with 0.5× SSC/0.1% SDS (for the oligonucleotides) or 0.1% SSC (for the PCR probes) at the respective hybridization temperature. Membranes were then rinsed with buffer B1 (0.1 M maleic acid, pH 7.5, 0.15 M NaCl, 0.06% Tween-20) and equilibrated for 30 min in buffer B2 (0.1 M maleic acid, pH 7.5, 0.15 M NaCl, 0.15% Tween-20, 1% blocking reagent). After addition of anti-Digoxigenin-AP Fab fragments (Roche) at a dilution of 1:10 000 and another 60 min of incubation, membranes were washed twice for 15 min with buffer B1 and equilibrated for 10 min in substrate buffer (100 mM Tris·HCl, pH 9.5, 100 mM NaCl and 50 mM MgCl₂). For detection, NBT (nitro blue tetrazolium in 70% dimethylformamide, Promega) and BCIP (5-bromo-4-chloro-3-indolyl-phosphate in 100% dimethylformamide, Promega) were added to the substrate buffer to a final concentration of 175 and 88 μg/ml, respectively. The detection reaction was stopped by rinsing the membrane with water.

**Northern blotting**

Northern blotting was performed essentially as described in (14), with minor modifications: hybridization temperatures ranged from 50 to 58°C, all wash steps were performed at
room temperature, and signals were detected after 3 h to 5 days of phosphoimaging. Oligonucleotide probes were designed to match cDNA clones representing antisense or intergenic transcripts; those yielding prominent signals are given below. Experiments were repeated at least twice, for longer target sequences with two different probes.

Probe 32a: 5' -CGTGTGTAATAGCCGAGGGCGGAAAG Probe 32b: 5' -TGCAGAGGCTTGTTCCCCGAGGAA Probe 24a: 5' -CTACCCGGCTATCTCCCAAGAAGACGG Probe 24b: 5' -CCACGGGGAGTGATCACGAAATGACG Probe 27a: 5' -CAGCCCGCGCCTGGTGAATCC Probe 27b: 5' -GGAGGGTGGCCGAGCCGAGCAAG Probe Phet-aq1731a: 5' -CTGGCGCAATGCAGGAAGTGGCCT Probe 36a: 5' -CGCTAACCACAGTTCTGCAAGC Probe 65a: 5' -GGAGTTTCCCAACAATAGAAAGTCTCCCTT

Sequencing and data analysis
cDNA clones of the library were sequenced on an ABI Prism 3700 (Perkin Elmer) sequenator using the BigDye terminator cycle sequencing reaction kit and the M13 rsP reverse primer (5' -CCCAGCTTTTACATTATGCTTCGCCGCTC). The LaserGene SeqMan II program package was employed to identify clones with identical or overlapping cDNA sequences, and sequence identities were identified by blastn searches on the NCBI web page. Sequence alignments were performed with ClustalW available at http://www.ebi.ac.uk/clustalw/#. The most prevalent group of non-ribosomal RNAs then predominantly originating from the pSPORT 1 vector. Of the remaining 914 clones, the majority (68%) corresponded to fragments of ribosomal RNAs (5S, 16S and 23S rRNA). The most prevalent group of non-ribosomal RNAs then were intergenic RNAs, with 90 out of the 100 clones derived from the intergenic region between reading frames phet and aq1731. The group of intergenic RNAs was followed in number by tRNAs and their precursors (68), mRNAs (66), tmRNA (43) and, as the least abundant group of A. aeolicus RNAs within the cDNA library, 18 RNAs with antisense orientation to annotated reading frames. For a complete list of cDNA sequences, excluding rRNAs and tRNA, see Table S1 of the Supplementary Material. Library members that represented intergenic and antisense RNAs were categorized according to their genetic context (Figure 1B), their genomic locations are illustrated in Figure 2A and B, and further details are listed in Table 1.

RESULTS
RNA classes
Among the 1042 cDNA library clones, 994 yielded either dot blot signals and/or analyzable sequence information (Figure 1A): 80 out of the 994 clones represented either very short sequence stretches or non-Aquifex sequences, predominantly originating from the pSPORT 1 vector. Of the remaining 914 clones, the majority (68%) corresponded to fragments of ribosomal RNAs (5S, 16S and 23S rRNA). The most prevalent group of non-ribosomal RNAs then were intergenic RNAs, with 90 out of the 100 clones derived from the intergenic region between reading frames phet and aq1731. The group of intergenic RNAs was followed in number by tRNAs and their precursors (68), mRNAs (66), tmRNA (43) and, as the least abundant group of A. aeolicus RNAs within the cDNA library, 18 RNAs with antisense orientation to annotated reading frames. For a complete list of cDNA sequences, excluding rRNAs and tRNA, see Table S1 of the Supplementary Material. Library members that represented intergenic and antisense RNAs were categorized according to their genetic context (Figure 1B), their genomic locations are illustrated in Figure 2A and B, and further details are listed in Table 1.

Authenticity of cDNA 5' and 3' ends
A basic question was whether the cDNA clones reflected the authentic 5' and 3' ends of their cellular RNA templates. Since the 5' and 3' ends of many stable RNAs in E. coli are known, we looked at E. coli cDNA libraries constructed in the same way as the A. aeolicus library (8). Inspection of 61 and 27 individual sequences, respectively, from contigs representing the small E. coli RNAs RyeB (104 nt) and RybB (81 nt) revealed that all individual sequences matched, within a few nucleotides, the experimentally verified natural 3' end (data not shown), whereas not a single sequence included the genuine 5'-terminus. Similarly, at least one-third of the individual clones represented the natural 3' end situation in the case of
A total of 66 clones corresponded to internal mRNA fragments of protein coding regions annotated in the *A. aeolicus* genome (section A in Table S1 of Supplementary Material). Some of these included part of the flanking 5′- or 3′-non-translated region (NTR) in addition to the coding mRNA segment, and three clones comprised a 3′-segment of one cistron, the intergenic region and a 5′-segment of the downstream cistron (e.g. *nuG*-rplK). Two cDNA clones (Aae-24 and Aae-117) mapped onto the junction of two protein genes with opposite orientation whose 3′ ends overlap by 10 and 4 nt, respectively (Figure 2A). These two RNAs were therefore sense to one and antisense to the other mRNA, as further discussed below in the context of antisense RNAs.

**Figure 1.** Composition of the *A. aeolicus* cDNA library (clones identified by sequencing and/or dot blot hybridization). (A) Number of clones in the different categories. Unclear assignment: sequences with several or no matches to the *A. aeolicus* genome, mostly due to shortness of the respective sequence. (B) Number of clones in the different categories. Unclear assignment: sequences with several or no matches to the *A. aeolicus* genome, mostly due to shortness of the respective sequence. (B) Antisense and intergenic sequences categorized in more detail by genetic context.

**E. coli** RNase P RNA (*rnpB*, 21 clones, 377 nt in its mature form), 6S RNA (*ssrS*, 45 clones, 184–186 nt in its mature form) and tmRNA (*ssrA*, 24 clones, 362 nt in its mature form). For all five small RNAs, 10–18 5′-terminal nucleotides were still missing in the most 5′-extended individual contig sequences (data not shown). In conclusion, genuine 3′ ends of biochemically uncharacterized RNAs can be inferred from the 3′ end of cDNA clones in the library when contigs comprise several individual sequences. This is due to the library construction strategy based on C-tailing of RNA molecules followed by reverse transcription and second strand synthesis (see Materials and Methods), which ensures isolation of native 3′ ends, while the genuine 5′ ends of RNAs need to be determined by additional experiments (see below).

**mRNAs**

A total of 68 tRNA-derived clones were identified within the library (Table S1, D), namely 35 covering portions of the tRNA<sup>ile</sup>–tRNA<sup>Aaa</sup> tandem precursor encoded in the spacer region of the two tRNA operons, seven fragments from the cluster of the four consecutive tRNAs Thr–Tyr–Gly–Thr, region of the two rRNA operons, seven fragments from the cluster of the four consecutive tRNAs Thr–Tyr–Gly–Thr, two sequences from tRNA<sup>Glu</sup>(UUU) that is part of a dicistronic transcript with tRNA<sup>Aaa</sup>(GGA), and three of tRNA<sup>Val</sup>(UAC) cotranscribed with tRNA<sup>Glu</sup>(UUC). These tRNA clones possessed 3′ extensions or truncations, and none of them included a post-transcriptionally added CCA terminus. Furthermore, several of the tRNA clones derived from di- or tetracistronic tRNA precursors partly or fully covered two consecutive tRNA domains.

**tRNAs and their precursors**

A total of 68 tRNA-derived clones were identified within the library (Table S1, D), namely 35 covering portions of the tRNA<sup>ile</sup>–tRNA<sup>Aaa</sup> tandem precursor encoded in the spacer region of the two tRNA operons, seven fragments from the cluster of the four consecutive tRNAs Thr–Tyr–Gly–Thr, two sequences from tRNA<sup>Glu</sup>(UUU) that is part of a dicistronic transcript with tRNA<sup>Aaa</sup>(GGA), and three of tRNA<sup>Val</sup>(UAC) cotranscribed with tRNA<sup>Glu</sup>(UUC). These tRNA clones possessed 3′ extensions or truncations, and none of them included a post-transcriptionally added CCA terminus. Furthermore, several of the tRNA clones derived from di- or tetracistronic tRNA precursors partly or fully covered two consecutive tRNA domains.

**tmRNA**

The gene encoding tmRNA had been annotated in the *A. aeolicus* genome (9), and tmRNA was represented by 43 clones in our cDNA library. This RNA, originally termed 10Sa RNA and of similar size as bacterial RNase P RNA (17,18), formerly denoted 10Sb, was of particular interest to us in order to assess the probability for a potential RNase P RNA to be included in our library. In *E. coli* cells, the tmRNA copy number was determined to be 10% that of ribosomes (17). This ratio was reported to be lower (1–2%) for RNase P RNA under a variety of growth conditions (19). Thus, a 5- to 10-fold molar excess of tmRNA over RNase P RNA, in line with previous approximations (20), seems to be the best estimate available at present. Accordingly, 4–9 RNase P RNA clones were to be expected in our cDNA library, provided that such an RNA would have been as good a substrate as tmRNA in all enzymatic reactions involved in library construction. In a related study, cDNA libraries generated in the same way as the *A. aeolicus* library investigated here were constructed from *E. coli* cells harvested in the lag, exponential or stationary phase (8). Each growth phase-specific library consisted of ~1000 clones after exclusion of clones representing tRNA and tRNA fragments: 24 clones represented tmRNA (11 in the lag phase, 9 in the exponential and 4 in the stationary phase library) and 21 RNase P RNA (12 in the lag phase, 9 in the exponential and none in the stationary phase library; (8), Supplementary Table S1 therein). Based on the absence of an RNase P RNA clone in the stationary phase *E. coli* library, it remains a formal possibility that we excluded potential *A. aeolicus* RNase P RNA clones because *A. aeolicus*
cells were harvested in early stationary phase. However, the stationary phase *E. coli* library was derived from cells harvested at a cell density of about $10^9$/ml [1.5 OD$_{600}$; (8)] whereas *A. aeolicus* cells had a density of $10^8$/ml or less at their harvest. Likewise, we found 42 tmRNA clones in our *A. aeolicus* library, whereas only 4 tmRNA clones were represented in the *E. coli* stationary phase library of comparable size. Thus, the *E. coli* and *A. aeolicus* libraries are hardly comparable, and we consider it rather unlikely that we failed to identify a hypothetical *A. aeolicus* RNase P RNA or fragment thereof simply for statistical reasons.

Antisense tRNAs

Remarkably, the library included antisense RNAs complementary to tRNAs (Figure 2A, Table 1 and Table S1, Section C II c). In each case, full complementarity was confined to one tRNA isoacceptor only, indicating that the antisense transcript was derived from the opposite strand of this particular tRNA transcription unit.

The *A. aeolicus* genome encodes four serine-specific tRNA isoacceptors (anticodons UGA, CGA, GCU and GGA). One clone (*Aae-27*) represented a 101-nt fragment complementary to tRNA$^{38\text{Ser}}$(UGA) over almost its entire length except for the

---

**Figure 2.** (A and B) Potential small stable RNAs in the *A. aeolicus* cDNA library and their genetic context [(A) antisense RNAs; (B) intergenic RNAs]. Light gray, reading frames; dark gray, tRNA genes; black, expressed small RNAs identified in the library with the length of sequence (in nt) matching the *A. aeolicus* genome indicated. The direction of the arrows corresponds to gene orientation. Contigs are grouped according to subclasses of Figure 1B. Note that RNAs smaller than 40 nt are not drawn to scale. (C) Expression of selected library RNAs analyzed by northern blotting.
3’-terminal 8 nt (the discriminator defined as the 3’-terminal nucleotide) and also covering 19 nt of the tRNA 5’-flanking sequences. In addition, antisense RNA Aae-27 almost perfectly matches tRNA\textsuperscript{Ser}(CGA), with the region of complementarity from tRNA nucleotide +4 to 9 nt upstream of the discriminator nucleotide. Complementarity to the remaining two isoacceptors (GCU, GGA) is interrupted by numerous mismatches, suggesting that a putative antisense interaction might be essentially restricted to the UGA and CGA isoacceptors. As a consequence, recognition of only two (UCA and UCG) of the six serine codons would be affected. Noteworthy, these two Ser codons are the rarest Ser codons in \textit{A.aeolicus} (Genome Atlas Database; http://www.cbs.dtu.dk/services/GenomeAtlas), utilized at a frequency of 0.16% (UCA) and 0.04% (UCG) compared with 0.75% (UCU), 3.38% (UCC) and 0.78% (AGU and AGC each). We also analyzed expression of the Aae-27 RNA by northern blotting (Figure 2C). With two different oligonucleotide probes (27a and b, see Materials and Methods), a common signal at ~70 nt was detected, indicating that this antisense tRNA reaches substantial expression levels in \textit{A.aeolicus} cells. The smaller size of 70 nt compared with the 101 nt of clone Aae-27 suggests that the latter stems from a precursor molecule to the 70-nt species.

Another antisense tRNA fragment, represented by two clones in the library (Aae-13), was fully complementary to tRNA\textsuperscript{Leu}(CAA), covering the tRNA except for the first 5 and last 9 nt. Complementarity is reduced but still substantial for the other four Leu isoacceptors, with the second best match to the UAA isoacceptor (comprising a stretch of 41 nt that included 3 mismatches). The UUG codon recognized by the CAA isoacceptor is one of the two rarely used Leu codons in \textit{A.aeolicus} (the second is CUA), but UUG should also be recognized by the UAA isoacceptor as in other bacteria.

The third antisense tRNA (Aae-H8) is complementary to \textit{5S RNAs} of ~60 nt at the 3’-portion (including CCA encoded in the gene) of \textit{tRNA\textsuperscript{Lys}(CUU)} and 55 nt of its 3’-flanking sequences.

### Other antisense RNAs

In addition to the antisense tRNA clones discussed above, eleven RNA contigs with antisense orientation to annotated reading frames were identified in the cDNA library (Figure 2A and Table 1). Eight of those were complementary to internal stretches of protein genes. For clones Aae-32 and Aae-134, RNAs of ~60 nt were identified in northern blots (Figure 2C). Clone Aae-140 RNA is noteworthy, since it is fully complementary over its 21 nt to transcripts of two annotated protein genes (\textit{aq_512} and \textit{aq_501} = \textit{pmu}, phosphoglucomutase/phospho-mannomutase), separated by ~8 kb, which are almost identical over ~660 nt of their N-terminal coding sequences, indicating that they are paralogs (Figure 2A). A second category is represented by the aforementioned clones Aae-24 and Aae-117, which fall into regions where two ORFs of opposite orientation partially overlap. Here, the antisense RNA is partially in sense with one mRNA and antisense with respect to the other transcribed in opposite direction to the first. To account for this situation, clones Aae-24 and Aae-117 have been listed with antisense as well as ‘sense with mRNA’ contigs (Table S1, A and C). Such a constellation could suggest interdependent expression of the two overlapping genes regulated by an antisense mechanism. Probes specific to clone Aae-24 identified an RNA of 180 nt in

---

**Table 1. Summary of data obtained for \textit{A.aeolicus} small non-coding RNA candidates (intergenic and antisense)**

| Contig | No. of clones | Length of sequence (nt) | GC content | Northern signal (nt) | Category | Comments | Accession no. |
|--------|--------------|-------------------------|------------|---------------------|----------|----------|--------------|
| Aae-8  | 90           | 170                     | 60         | 120                 | I        | Putative 6S RNA 5’-RACE (ca.: 120, 160, 210, 370 nt) | A888820   |
| Aae-13 | 2            | 71                      | 66         | n.d.                | A        | Antisense to tRNA\textsuperscript{Lys}      | A888797   |
| Aae-17 | 1            | 183                     | 38         | 70/110/260          | I        |                                     | A887999   |
| Aae-18 | 1            | 212                     | 49         | Ambiguous, multiple | I        |                                     | A887999   |
| Aae-24 | 4            | 145                     | 48         | 180                 | A/S      | Maps to junction of opposing reading frames | A888800   |
| Aae-27 | 1            | 101                     | 60         | 70                  | A        | Antisense to tRNA\textsuperscript{Ser}      | A888801   |
| Aae-32 | 1            | 38                      | 68         | 60                  | A        |                                     | A888802   |
| Aae-36 | 1            | 30                      | 53         | 280/360/460         | I        |                                     | A888803   |
| Aae-43 | 1            | 15                      | 27         | —                   | A        |                                     | A888804   |
| Aae-65 | 2            | 87                      | 56         | 80/100              | I        |                                     | A888805   |
| Aae-89 | 1            | 53                      | 43         | —                   | A        |                                     | A888806   |
| Aae-96 | 1            | 86                      | 49         | —                   | A        |                                     | A888807   |
| Aae-98 | 1            | 89                      | 62         | Ambiguous, multiple | I        |                                     | A888808   |
| Aae-106| 1            | 19                      | 58         | Multiple 240-320    | I        |                                     | A888809   |
| Aae-108| 1            | 22                      | 45         | —                   | A        |                                     | A888810   |
| Aae-109| 1            | 18                      | 44         | Multiple 80-1000    | A        |                                     | A888811   |
| Aae-114| 1            | 21                      | 43         | —                   | A        |                                     | A888812   |
| Aae-117| 1            | 134                     | 38         | —                   | A/S      | Maps to junction of opposing reading frames | A888813   |
| Aae-118| 1            | 146                     | 47         | —                   | I        | Maps close to putative gene for 4.5S RNA homolog | A888814   |
| Aae-134| 1            | 59                      | 51         | 60                  | A        |                                     | A888815   |
| Aae-139| 1            | 22                      | 36         | —                   | A        | Complementary to \textit{amtB} coding region + 5’-NTR | A888816   |
| Aae-140| 1            | 21                      | 67         | n.d.                | A        | Full complementarity to two gene loci       | A888817   |
| Aae-202| 1            | 40                      | 65         | Multiple 30-1000    | I        |                                     | A888818   |
| Aae-H8 | 1            | 85                      | 41         | n.d.                | A        | Antisense to tRNA\textsuperscript{Lys}      | A888819   |

n.d., not determined; —, no Northern signal although assayed; ambiguous, substantially different signals with different probes; multiple, five and more different bands; I, intergenic; A, antisense; S, in sense with mRNA. For precise location within the genome and complete sequence, see Table S1 and contig list of the Supplementary Material.
northern blot experiments (Figure 2C). Clone Aae-139, 22 nt in length, defined another category (Table S1, C II b). It is complementary to 17 nt of the 5′-UTR and to the first 5 nt of the coding region of _aq 112_ encoding an ammonium transporter (amtB). AmtB appears to be part of a polycistronic operon involved in glutamine synthesis, consisting of the genes _glkB, glmA, amtB_ and _aq 113_ (9). _GlnA_ and _amtB_ are separated by an intergenic region of 62 nt. A possible role of the clone Aae-139 antisense RNA could be blockage of ribosomal access to the AUG start codon of _amtB_, thereby attenuating its translation. A precedent for such a mechanism involving a cis-encoded antisense RNA is translational inhibition of transposase mRNA encoded by the _E. coli_ insertion sequence IS10. A single clone that also belongs to this category of potential cis-encoded antisense RNAs directed against protein genes was identified in a similar cDNA library of non-coding RNAs from the crenarchaeon _Sulfolobus solfataricus_ [RNA Sso-203; (21)].

Finally, for several of the antisense RNA candidates, e.g. Aae-43, 109 and others, it needs to be kept in mind that annotation of reading frames on the opposing strand is hypothetic or putative and that open reading frames, albeit short, in sense with the putative antisense RNAs do exist. Final assignment of these RNAs as non-coding antisense RNAs thus must await further experimental evidence.

### Intergenic RNAs

A total of 93% of the _A. aeolicus_ genome sequence was reported to represent protein-coding regions, 0.8%stable RNA genes and only 6.2% intergenic sequences (9). After exclusion of sequences derived from intergenic sequences within the two tRNA operons, 100 intergenic cDNA clones remained, 90 of them derived from the region separating _pheT_ and _aq 1731_.

The genomic location of the intergenic transcripts is illustrated in Figure 2B and their positional details are given in Table S1, section B. Intergenic regions are of particular interest, because they may harbor genes encoding other non-messenger RNAs in addition to rRNAs, tRNAs, tmRNA and the 4.5S RNA homolog (Signal Recognition Particle Database; http://psyc.he.uth.edu/dbs/SRPDB/SRPDB.html). Notably, the intergenic RNA Aae-118 originated from a genomic location very close to the putative gene of the 4.5S RNA homolog (between _aq 1854_ and _aq 1855_). An interesting aspect of _A. aeolicus_ intergenic RNAs is their G+C-content, considering that stable RNAs in this hyperthermophilic bacterium have an increased proportion of G+C (tRNAs: 68.5%; tmRNA: 66.9%; rRNA: 65%), while the average G+C-content of the genome is only 43.4% and relatively constant (10). The _pheT-aq 1731_ intergenic transcripts as well as a few other clones, such as Aae-202 and Aae-98, indeed conformed to the criterion of a substantially elevated G+C content (Figure 2B and Table 1). Clone Aae-202 RNA was derived from in between the reading frames _aq 1666_ and _thrS_ (threonyl-tRNA synthetase) to which also an antisense transcript, clone Aae-89 RNA, was found in the library. Clone Aae-98 RNA, encoded by the extrachromosomal element (ecel1) of _A. aeolicus_, was transcribed from the unusually large intergenic region of 2 kb between ORFs _aa 18_ and _aq 19_, but interestingly also displayed extensive similarity to a stretch of 89 nt within an open reading frame of unknown identity (aa05) in ecel1.

The _pheT-aq 1731_ intergenic RNA represented the vast majority of intergenic transcript clones in our library. The genetic context provided the first evidence for it being the 6S RNA homolog (_ssrS_) of _A. aeolicus_. In _E. coli_, 6S RNA is the 5′-processing product released from a dicistronic primary transcript with _ygfA_, encoding a 5-formyltetrahydrofolate cyclo- ligase, and _aq 1731_ is identified as the _A. aeolicus_ _ygfA_ homolog in NCBI blastp searches. In view of the low similarities of _E. coli_ and _A. aeolicus_ in terms of genetic organization, we considered it highly unlikely that the genetic co-localization of a stable RNA and _ygfA_ in both bacteria is fortuitous rather than witnessing a functional commonality. Also, clear structural similarities to other verified and putative 6S RNAs are evident, as outlined below.

### 5′-RACE analysis of putative 6S RNA transcripts

The intergenic region between _pheT_ and _aq 1731_ comprises 195 nt. Among 28 sequenced library clones, 24 had a 3′ end located 39–44 nt upstream of the _aq 1731_ start codon (Figure 3). Since the procedure employed for cDNA library construction preserves the authentic 3′-terminus, we concluded that the majority of RNA molecules have their genuine 3′ end at nucleotide ~39 or ~41 with respect to _aq 1731_. Such 3′ end microheterogeneity is expected to reflect the natural situation, as recent 3′-RACE experiments for _E. coli_ 6S RNA have detected length variants of 184, 185 and 186 nt (22). On the 5′-side, several clones had their 5′ end 5–6 nt downstream of the _pheT_ stop codon (Figure 3). Assuming that the mature 5′ end of 6S RNA roughly coincides with the _pheT_ stop codon, we considered the mature 6S RNA homolog to have a length of ~150 nt, thus somewhat shorter than the size of 6S RNAs from _E. coli_ and _Pseudomonas aeruginosa_ (~185 nt (22–24)). Despite this length difference, mfold analysis predicted a rod-shaped secondary structure for the _A. aeolicus_ RNA, with astounding similarity to the proposed structure of _E. coli_ 6S RNA (Figure 4A; for details see Discussion). Remarkably, according to mfold the _A. aeolicus_ RNA is the most stable among all bona fide 6S RNAs (aligned in Figure 4B).

One of the sequenced _pheT-aq 1731_ clones extended into the _pheT_ cistron, but had its 3′ end at position ~41 (Figure 3), thus probably representing a 3′-mature but 5′-premature RNA. We interpreted this to suggest that the putative 6S RNA homolog is post-transcriptionally released from _pheT-aq 1731_ primary transcripts or, alternatively, that its transcription may be driven from a promoter within _pheT_. To map the 5′ end(s), we performed a 5′-RACE experiment. By utilizing this approach, we also wanted to clarify a seemingly contradictory finding: although the putative mature 6S RNA was predicted to comprise at least ~150 nt in order to be able to adopt the rod-shaped structure shown in Figure 4A, our northern blot analysis (Figure 2C) revealed a 120-nt signal. The results of the 5′-RACE analysis and their interpretation are shown in Figure 5. Among the five prominent bands (numbered I–V in panel A), one (II) only appeared when the TAP treatment (for details, see Materials and Methods) was included, indicating that this band reflects a primary transcript with a 5′-triphosphate end. The largest product I, whose 5′ end was not analyzed here, may represent a processing product released from longer _pheT-6S_ RNA cotranscripts, suggesting that 6S RNA is also produced by transcription initiating at sites further
Stop pheT

| A. aeolicus | CGAAGACTTTGCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
|------------|---------------------------------------------------------------|
| 8.1        | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.2        | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.3        | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.4        | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.5        | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.6        | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.7        | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.8        | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.9        | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.10       | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.11       | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.12       | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.13       | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.14       | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.15       | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.16       | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.17       | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.18       | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.19       | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.20       | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.21       | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.22       | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.23       | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.24       | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.25       | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.26       | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.27       | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |
| 8.28       | GAGGGACTTCCCACCTTTAAGACTTGAGGCCCAACGCTTTCTGAGACCTCGACCTCGAGCGGCGGACATCTTGGTTTCCCTCTCTAAGGACGAGGCGGAAGAAttgc |

Figure 3. Alignment of sequenced putative 6S RNA clones (contig Aae-8) from the A. aeolicus cDNA library. Top line: A. aeolicus genomic DNA sequence with the intergenic region in capital letters, protein coding regions in small letters and the stop codon of the upstream reading frame pheT (phenylalanyl tRNA synthetase β subunit) and the start codon of the downstream reading frame ag_1731 (5-formyltetrahydrofolate cyclo-ligase homolog) indicated by a gray box and oval, respectively. Lines below the first one: single sequences from individual library clones.

upstream. Several individual cDNA clones for bands II–V were sequenced and their 5' ends mapped (panel B, vertical arrows). Transcription of the primary transcript (product II) is initiated within the pheT gene. Notably, putative −10 and −35 promoter boxes, similar in sequence and spacing to the E.coli promoter consensus, could be easily identified at the expected location upstream of the transcriptional start site (Figure 5B). Product III seems to represent a processing intermediate, while the 5' end of product IV exactly maps to the site anticipated for the 6S RNA. The latter has been observed for several bacterial organisms, e.g. genes for amino acid biosynthesis, are dispersed throughout the A. aeolicus genome, and subunits of the same enzyme (e.g. glutamate synthase, gltB and

in view of extensive base-pairing in the terminal stem region (Figure 4A), we consider it very unlikely that this discontinuity in the RNA chain will cause significant disassembly of the RNA.

**DISCUSSION**

**Search for RNase P RNA in A. aeolicus**

This study was initiated as part of our efforts to identify an RNase P RNA in A. aeolicus. Yet, we could not identify a reasonable candidate in our cDNA library. The spectrum of potential reasons is diverse, including (i) there is no RNase P RNA in this organism because A. aeolicus possesses a protein-alone enzyme; (ii) the RNA may resist the reverse transcription and amplification procedure; (iii) RNase P RNA was not included in our library because of low copy number or (iv) RNase P RNA may assemble from fragments encoded by split genes. The latter has been observed for several bacterial tRNAs (25,26). Indeed, a tendency to disrupt polycistronic operons and also to split single genes is evident in A. aeolicus. Many genes, functionally grouped within operons in other organisms, e.g. genes for amino acid biosynthesis, are dispersed throughout the A. aeolicus genome, and subunits of the same enzyme (e.g. glutamate synthase, gltB and
Figure 4. (A) Proposed secondary structures of (putative) 6S RNAs from *E. coli, A. aeolicus* and *X. fastidiosa*. CR, conserved regions numbered I–IV according to (32). Broken lines, additional potential base pairs predicted by mfold (36); boxed nucleotides, conserved regions according to (32), with nucleotide matching the consensus shaded gray [see (B)]. (B) Alignment of (putative) 6S RNA sequences based on (32). Four prominent conserved regions (CR I–IV) are boxed, and elements matching the consensus are highlighted at the top; nucleotides of *X. fastidiosa* and *A. aeolicus* matching the consensus of all other γ-proteobacterial sequences are shaded gray. Note that there is a single U replacement at the first conserved A residue in CRI of 6S RNA from *S. onedensis*.
gltD: phenylalanyl-tRNA synthetase, pheS and pheT) are separated on the A.aeolicus chromosome (9,10). Methionyl- and leucyl-tRNA synthetases had previously been found as single polypeptide enzymes only; however, the A.aeolicus homologs are fragmented into two subunits encoded far apart on the chromosome, and in the case of leucyl-tRNA synthetase, both subunits were shown to be essential for function (9,27,28). Along the same line, tRNA CCA-addition, in most organisms catalyzed by a single enzyme, is divided between two separate enzymes in A.aeolicus, one adding the CC and the other the terminal A residue (29).

The A.aeolicus genome sequence predicts di- and tetracis-tronic tRNA transcripts. Indeed, many of the cDNA clones representing mature tRNA transcripts that were identified in our library possess 5' extensions or are fragments derived from tRNA tandem precursor transcripts (Figure 1; Table S1). On the other hand, primer extension has revealed the presence of canonical mature tRNA 5' ends in A.aeolicus total RNA extracts, as exemplified for the terminal tRNAStart of four consecutive tRNAs encoded in the tufA2 operon (11). This clearly demonstrates that transcription of tRNAs as precursors followed by canonical tRNA 5' end maturation does occur in A.aeolicus, although the identity of the 5'-processing enzyme activity remains as yet elusive.

tRNA antisense transcripts

Transcription of antisense tRNAs (Figure 2A; Tables 1 and S1) is apparently not unique to A.aeolicus, since it has also been detected in a transcriptome analysis of E.coli (30) as well as in a cDNA library of the crenarchaeon Sulfolobus solfataricus (21) similar to the one constructed for A.aeolicus. In S.solfataricus, three different antisense tRNAs were detected, two complementary to specific intron-containing precursor tRNAs. As for A.aeolicus, complementarity was complete with respect to the main target tRNA isoacceptor and spanning almost the entire (pre-)tRNA, indicating that antisense tRNAs were cis-transcribed from the same locus but opposite strand as the target RNAs. The prominent northern blot signals for two of the S.solfataricus antisense tRNAs (21) and the A.aeolicus antisense tRNAStart(UGA) (Aae-27, Figure 2C) argue against the possibility that these antisense tRNAs result from rare spurious transcription events. Potential functions include negative regulation (e.g. antisense inhibition, induction of degradation) of target tRNA isoacceptors, or may be independent of tRNA complementarity if target/antisense tRNA hybrids simply do not form in the cellular milieu.

6S RNA—function and structure

6S RNA, encoded by the ssrS gene, was demonstrated to suppress transcription from a subset of σ70-dependent promoters during stationary phase in E.coli (13). Recent data indicate that σ70-dependent promoters inhibited by 6S RNA contain an extended −10 promoter element (31). 6S RNA-deficient E.coli cells are at a disadvantage for survival in stationary phase (31), where this RNA reaches its highest
cellular concentration (~10,000 copies per cell versus ~1000 in logarithmic growth). During stationary phase, >75% of the σ70-containing RNA polymerase holoenzyme is complexed with 6S RNA (13). These findings for E.coli 6S RNA may explain the high abundance of the putative 6S RNA homolog in our A.aeolicus cDNA library derived from cells harvested in the early stationary growth phase.

The gene for 6S RNA is located immediately upstream of the ygfa gene in γ-proteobacteria and A.aeolicus. Mfold analyses of this region combined with 6S RNA blastn searches enabled us to identify additional ssrS homologs within the γ-proteobacteria beyond those already noticed. These include the conjunctal 6S RNAs of Xylella fastidiosa and Haemophilus ducreyi. The known 6S RNAs, so far exclusively from γ-proteobacteria, have been proposed to adopt a rod-like, mainly double-stranded structure with a central bulged region that is reminiscent of an ‘open complex’ found during transcription initiation (13). In fact, the currently identified γ-proteobacterial 6S RNAs (with the exception of X.fastidiosa) and the A.aeolicus homolog share conserved sequence motifs in this central region (5'-GRGCCNAYA in CR I, 5'-CCUUR- AA in CR III and 5'-GGYYCANGG in CR IV; Figure 4A and B). An alignment of known and putative 6S RNAs, based on that by (32) and including the novel candidates, is shown in Figure 4B, with the conserved regions CR I–IV highlighted by boxes as in Figure 4A. CR II is not present in the A.aeolicus homolog. However, the A.aeolicus RNA shares a further structural commonality with 6S RNA from E.coli and close relatives, centering around the single U bulge in the terminal helix (Figure 4A, U15 in E.coli and U17 in A.aeolicus) and including the flanking 2 and 3 bp on the left and right, respectively. In several cases, mfold offers multiple folding solutions for the central loop, for example one in which elements CR III and IV form a stem-loop structure as suggested by (32). Alternative base pairs predicted by mfold are indicated by broken lines in Figure 4A. The conformational ambiguity in the central loop region may suggest that the biological function of 6S RNA is associated with switches of conformation in this part of the molecule. Finally, the example of the putative 6S RNA homolog of X.fastidiosa (Figure 4) predicts even more degeneracy within the observed consensus sequence motifs and underlines the substantial structure and length variation in the stem regions flanking the central bulge.

We have found the majority of A.aeolicus 6S RNA molecules to be nicked in the 5' proximal bulge of CR I (Figures 2C, 4A and 5). Although this nick is unlikely to result in significant dissociation of the two 6S RNA fragments, it may nevertheless play a functional role in the regulation of RNA polymerase activity.

**Phylogeny of ssrS and ygfa**

Since RNA polymerase phylogeny places A.aeolicus close to the δε-proteobacteria, we hoped to identify ssrS homologs in this branch, e.g. in Helicobacter or Campylobacter, which further share with Aquifex the idiosyncrey to encode a predicted HD superfamily hydrolase (COG1418) immediately downstream of ygfa (Figure S1 of the Supplementary Material). However, we could not identify a single convincing ssrS candidate outside the γ-proteobacteria, neither by mfold analyses of intergenic regions upstream of the respective ygfa genes nor by blastn searches with different 6S RNAs including the one from A.aeolicus. Blast searches with the A.aeolicus ygfa (aq_1731) gene product as the query sequence gave best matches to ygfa homologs in the Bacteroidetes branch (Bacteroides thetaiotaomicron, Porphyromonas gingivalis). However, in these two bacteria the ygfa homolog is directly preceded and actually overlaps with a gene predicted to encode a C-terminal protease (or its precursor), thus leaving no intergenic space to accommodate an ssrS homolog (Figure S1). In summary, the genetic coorganization of ssrS and ygfa as found in γ-proteobacteria and Aquifex seems to be exceptional among bacteria.

Our discovery of a 6S RNA homolog in A.aeolicus adds another piece of evidence in support of the evolutionary proximity of the transcription machineries in Aquifex and γ-proteobacteria, one of the conceivable consequences being similar modes of promoter recognition. In fact, the identification of promoter elements similar to the E.coli consensus upstream of the ssrS transcription initiation site in A.aeolicus (Figure 5B) is consistent with this assumption. Likewise, the β, β' and σ70 subunits of RNA polymerase from A.pyrophilus cross-reacted with antisera raised against the corresponding subunits from E.coli, and A.pyrophilus RNA polymerase was able to initiate to some extent transcription at the E.coli-specific T7 A1 promoter, although the enzyme from Thermotoga maritima did so as well (4). A.aeolicus encodes an alternative sigma factor σN (σN2, RpoN; aq_599). Together with one of its σN-dependent transcriptional activators (aq_218, NifA homolog), A.aeolicus σN was shown to associate with the E.coli core RNA polymerase and to be able to initiate transcription at the σN-specific E.coli glnHp2 promoter (33). As the NifA homolog had not been identified in any bacteria other than A.aeolicus and azotrophic proteobacteria, the authors discussed their findings to suggest that A.aeolicus either acquired the NifA gene from proteobacteria by horizontal gene transfer or that the evolutionary split of A.aeolicus from the proteobacteria was a more recent event than implied by 16S rRNA-based phylogeny (33).

A recent analysis of the domain architecture of bacterial β, β' and σ70 subunits has further substantiated the evolutionary proximity of transcription machineries in A.aeolicus and proteobacteria (34). On the other hand, phylogeny of ribosomal proteins again supported the existence of an original Thermotoga–Aquifex clade. To explain this discrepancy, the authors suggested a scenario according to which Aquifex and Thermotoga evolved from a common progenitor, but the Aquifex lineage at some point replaced its RNA polymerase subunits by horizontal gene transfer from proteobacteria. The importance of horizontal gene transfer in the evolution of the A.aeolicus genome is moreover corroborated by evidence that at least 10% of its protein-coding genes may have been acquired by this mode from Archaea (35). Such multiple transfer events could in fact be related to the mosaic-like appearance of the A.aeolicus genome. In view of these findings, the identification of a 6S RNA homolog in A.aeolicus makes its horizontal co-transfer together with components of the RNA polymerase from a γ-proteobacterial ancestor to the progenitor of Aquifex likely, further suggesting that the importance of 6S RNA for transcription regulation may be more substantial than currently thought.
SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank K.M. Wassarman for discussion and encouragement. This work has been supported by grant HA 1672/13-1 from the Deutsche Forschungsgemeinschaft to R.K.H and by a grant of the German Human Genome Project (DHGP) through the BMBF (#01KW9966) to A.H. Funding to pay the Open Access publication charges for this article was provided by Fonds der Chemischen Industrie.

Conflict of interest statement. None declared.

REFERENCES

1. Huber,R., Wilharm,T., Huber,D., Trincone,A., Burggraf,S., König,H., Rachel,R., Rockinger,I., Fricke,H. and Stetter,K.O. (1992) *Aquifex pyrophilus* gen. sp. nov., represents a novel group of marine hyperthermophilic hydrogen-oxidizing bacteria. *System. Appl. Microbiol.*, 15, 340–351.
2. Burggraf,S., Olsen,G.J., Stetter,K.O. and Woese,C.R. (1992) *Aquifex aeolicus* gen. nov., sp. nov., a hyperthermophilic hydrogen-oxidizing bacterium. *J. Bacteriol.*, 174, 3525–3541.
3. Bocchetta,M., Gribaldo,S., Sanangelantoni,A. and Cammarano,P. (2000) Phylogenetic depth of the bacterial genera *Aquifex* and *Thermotoga* inferred from analysis of ribosomal protein, elongation factor, and RNA polymerase subunit sequences. *J. Mol. Evol.*, 50, 366–380.
4. Klenk,H.P., Meier,T.D., Durovic,P., Schwass,V., Lottspeich,F., Dennis,P.P. and Zillig,W. (1999) RNA polymerase of *Aquifex pyrophilus*: implications for the evolution of the bacterial rpoBC operon and extremely thermophilic bacteria. *J. Mol. Evol.*, 48, 528–541.
5. Gruber,T.M. and Bryant,D.A. (1998) Characterization of the group 1 and group 2 sigma factors of the green sulfur bacterium *Chlorobium tepidum* and the green non-sulfur bacterium *Chloroflexus aurantiacus*. *Arch. Microbiol.*, 170, 285–296.
6. Gupta,R.S. (2000) The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. *EMBS Microbiol.*, 24, 367–402.
7. Huttenhofer,A., Cavaillé,J. and Bacherelli,J.-P. (2004) Experimental RNomics: a global approach to identifying small nuclear RNAs and their targets in different model organisms. *Methods Mol. Biol.*, 265, 409–428.
8. Vogel,J., Bartels,V., Tang,T.H., Lehrach,H., Bacherelli,J.P. and Brosius,J. (2001) RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J.*, 20, 2943–2953.
9. Wassarman,K.M. and Storz,G. (2000) 6S RNA regulates *E. coli* RNA polymerase activity. *Cell*, 101, 613–623.
10. Swanson,R.V. (2001) Genome of *Aquifex aeolicus*. *Methods Enzymol.*, 330, 158–169.
11. Willkomm,D.K., Feltens,R. and Hartmann,R.K. (2002) 6S RNA maturation in *Aquifex aeolicus*. *Biochimie*, 84, 713–722.
12. Huttenhofer,A., Kiefmann,M., Meier-Ewert,S., O’Brien,J., Lehrach,H., Bacherelli,J.P. and Brosius,J. (2001) RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J.*, 20, 2943–2953.
13. Washio,T., Sasayama,J. and Tomita,M. (1998) Analysis of complete *E. coli* genome complement. *Nucleic Acids Res.*, 30, 6057–6068.
14. Williams,K.P. (2002) Descent of a split RNA. *Nucleic Acids Res.*, 30, 2025–2030.
15. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
16. Wagner,E.G.H. and Vogel,J. (2005) Approaches to identify novel non-messenger RNAs in bacteria and to investigate their biological functions: functional analysis of identified non-mRNA. In Hartmann,R.K., Bindereif,A., Schön,A. and Westhof,E. (eds), *Handbook of RNA Biochemistry*. WILEY-VCH, Weinheim, Germany, pp. 614–642.
17. Lee,S.Y., Baily,S.C. and Apirion,D. (1978) Small stable RNA from *Escherichia coli*: evidence for the existence of new molecules and for a new ribonucleoprotein particle containing 6S RNA. *J. Bacteriol.*, 133, 1015–1023.
18. Iyer,L.M., Koonin,E.V. and Aravind,L. (2004) Evolution of bacterial riboswitches: implications for the evolution of the bacterial rpoBC operon. *Methods Mol. Biol.*, 265, 528–541.
19. Kim,K.S. and Lee,Y. (2004) Regulation of 6S RNA biogenesis by switching utilization of both sigma factors and endoribonucleases. *Nucleic Acids Res.*, 32, 6057–6068.
20. Vogel,D.W., Hartmann,R.K., Struck,J.C., Ulbrich,N. and Erdmann,V.A. (1987) The sequence of the 6S RNA gene of *Pseudomonas aeruginosa*. *Nucleic Acids Res.*, 15, 4583–4591.
21. Keiler,K., Shapiro,L. and Williams,K.P. (2000) tmRNAs that encode proteolysis-inducing tags are found in all known bacterial genomes: A two-pieces tmRNA functions in *Caulobacter*. *Proc. Natl Acad. Sci. USA*, 97, 7778–7783.
22. Williams,K.P. (2002) Descent of a split RNA. *Nucleic Acids Res.*, 30, 2025–2030.
23. Gouda,M., Yokogawa,T., Asahara,H. and Nishikawa,K. (2002) Leucyl-tRNA synthetase from the extreme thermophile *Aquifex aeolicus* has a heterodimeric quaternary structure. *FEBS Lett.*, 518, 139–143.
24. Gouda,M., Yokogawa,T. and Nishikawa,K. (2002) The β subunit of *Aquifex aeolicus* leucyl-tRNA synthetase is responsible for cognate tRNA recognition. *Biochem. Biophys. Res. Comm.*, 297, 950–955.
25. Tomita,K. and Weiner,A.M. (2001) Collaboration between CC- and AA-adding enzymes to build and repair the 3′-terminal CCA of tRNA in *Aquifex aeolicus*. *Science*, 294, 1334–1336.
26. Li,Y., Cole,K. and Altman,S. (2003) The effect of a single, temperature-sensitive mutation on global gene expression in *Escherichia coli*. *Trends Genet.*, 19, 161–169.
27. Trotochad,A.E. and Wassarman,K.M. (2004) 6S RNA function enhances long-term cell survival. *J. Bacteriol.*, 186, 4978–4985.
28. Brown,J.W. and Ellis,J.C. (2005) Comparative analysis of secondary structure: 6S RNA. In Hartmann,R.K., Bindereif,A., Schön,A. and Westhof,E. (eds), *Handbook of RNA Biochemistry*. WILEY-VCH, Weinheim, Germany, pp. 491–512.
29. Studholme,D.J., Wigneshweraraj,S.R., Gallegos,M.T. and Buck,M. (2000) Functionality of purified sigma(N) (sigma(54)) and a NifA-like protein from the hyperthermophile *Aquifex aeolicus*. *Nature New Biol.*, 310, 340–345.
30. Vogel,D.W., Hartmann,R.K., Struck,J.C., Ulbrich,N. and Erdmann,V.A. (1987) The sequence of the 6S RNA gene of *Pseudomonas aeruginosa*. *Nucleic Acids Res.*, 15, 4583–4591.