Ids and Phylogenetic Comparisons of Posttranscriptional Modifications in 16 S Ribosomal RNA from *Haloferax volcanii*

Received for publication, March 13, 2000, and in revised form, May 18, 2000
Published, JBC Papers in Press, May 18, 2000, DOI 10.1074/jbc.M002153200

Jeffrey A. Kowalak‡§, Eveline Bruenger*, Pamela F. Crain†, and James A. McCloskey‡¶

From the Departments of ‡Biochemistry and ¶Medicinal Chemistry, University of Utah, Salt Lake City, Utah 84112

Small subunit (16 S) rRNA from the archaeon *Haloferax volcanii*, for which sites of modification were previously reported, was examined using mass spectrometry. A census of all modified residues was taken by liquid chromatography/electrospray ionization-mass spectrometry analysis of a total nucleoside digest of the rRNA. Following rRNA hydrolysis by RNase T1, accurate molecular mass values of oligonucleotide products were measured using liquid chromatography/electrospray ionization-mass spectrometry and compared with values predicted from the corresponding gene sequence. Three modified nucleosides, distributed over four conserved sites in the decoding region of the molecule, were characterized: 3-(3-amino-3-carboxypropyl)uridine-966, N6-methyladenosine-1501, and N6,N6-dimethyladenosine-1518 and -1519 (all *Escherichia coli* numbering). Nucleoside 3-(3-amino-3-carboxypropyl)uridine, previously unknown in rRNA, occurs at a highly conserved site of modification in all three evolutionary domains but for which no structural assignment in archaea has been previously reported. Nucleoside N6-methyladenosine, not previously placed in archaean rRNAs, frequently occurs at the analogous location in eukaryotic small subunit rRNA but not in bacteria. *H. volcanii* small subunit rRNA appears to reflect the phenotypically low modification level in the Crenarchaeota kingdom and is the only cytoplasmic small subunit rRNA shown to lack pseudouridin

* This work was supported by NIGMS Grant GM29812 from the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Present address: Laboratory of Neurotoxicology, National Institute of Mental Health, Bldg. 10, Rm. 3D40, Bethesda, MD 20892-1262.

§ To whom correspondence should be addressed. Tel.: 801-581-5581; Fax: 801-581-7457; E-mail: james.mccloskey@nm.cc.utah.edu.
Posttranscriptional Modification in *H. volcanii* 16S rRNA

**Isolation and Enzymatic Hydrolysis of rRNA—*H. volcanii* (ATCC 29805)** was grown as reported earlier (28). 30 S ribosomal subunits were prepared (29), and the 16 S rRNA was isolated by extraction with phenol and chloroform (29). Purity of the RNA was assessed by gel electrophoresis (1% agarose) using an Applied Biosystems 320A micro-preparative electrophoresis system. 16 S rRNA was hydrolyzed to nucleosides using nuclease P1 (Sigma), venom phosphodiesterase I (E.C. 3.1.15.1, Sigma), and bacterial alkaline phosphatase (E.C. 3.1.3.1, Calbiochem) (30). Digestion of rRNA (10 pmol/µl in 20 mM Tris hydrochloride, pH 7.5, 1 mM EDTA, pH 7) by RNase T1 (E.C. 3.1.27.3) (Ambion, Austin, TX) was carried out for 30 min at 37 °C at a concentration of 1000 units/50 pmol of RNA.

**LC/ESI-MS of Nucleotides from Total Digestion of *H. volcanii* 16 S rRNA**—A Quattro II mass spectrometer with MassLynx version 3.1 data system (Micromass, Beverly, MA) interfaced to an HP 1090 liquid chromatograph with diode array detector (Hewlett-Packard, Palo Alto, CA) was used for all LC/MS studies. Two hundred picomoles (~100 µg) of rRNA hydrolysate was injected directly onto a 2.1 × 100 mm LC-18S column fitted with a matching 2.1-mm precolumn (Supelco, Bellefonte, PA). The column was eluted at a flow rate of 300 µl/min using an ammonium acetate/acetonitrile gradient as described previously (31), except the concentration of ammonium acetate was decreased to 0.005M for compatibility with electrospray ionization. Diode array UV absorbance data were acquired from 240–320 nm.

The chromatographic effluent was conducted without splitting into the mass spectrometer, using the standard megaflo inlet. The ion source was 180 °C. Capillary and lens voltages were 3.1 and 0.24 kV, respectively, for measurement of positive ions. Data were acquired in “centroid” mode over the mass range 105–450 in 0.9 s (with a 0.1-s interscan delay, for a cycle time of 1 s). The “Cluster” algorithm from the MassLynx software was used to interrogate the data set for the presence of unknown nucleotides; cluster values were 132 and 146 units for normal and 2′-O-methylated nucleotides, respectively.

**LC/ESI-MS of Oligonucleotides from RNase T1 Digests of *H. volcanii* 16 S rRNA**—The mass spectrometer and liquid chromatograph are described in the preceding section; a Z-spray interface was available for these studies. Fifty pmol (~25 µg) of rRNA hydrolysate was injected directly onto a 300 × 1-mm Supelco LC-18S column (Supelco) with a 15 × 1-mm OptiGuard C-18 precolumn cartridge (Optimize Technologies, Oregon City, OR). The solvent system consisted of 0.8 M Tris hydrochloride, pH 7.0, 1 mM EDTA, pH 7.0, and 0.005M ammonium acetate, with a 0.2-s interscan delay, for a cycle time of 1 s.

The chromatographic effluent was conducted without splitting into the mass spectrometer. The ion source and desolvation temperatures were 140 °C and 300 °C, respectively. Capillary and lens voltages were ~2.75 and 0.50 kV, respectively, for measurement of negative ions. Two alternating scan functions were used for data acquisition. The first one, used to determine oligonucleotide molecular masses, utilized a 44-V cone setting, a typical value for generating mass spectra with minimal fragmentation. Data were acquired in continuous mode over the mass range 480–1380 in 3 s (with a 0.2-s interscan delay, for a cycle time of 3.2 s). The second scan function utilized a cone setting of 130 V to fragment the oligonucleotides. Data were acquired in continuous mode over the mass range 100–350 in 0.4 s (with a 0.1-s interscan delay, for a cycle time of 0.5 s).

**RESULTS**

The posttranscriptional modification status of *H. volcanii* rRNA was examined using a combination of LC/MS-based methods (33) involving analysis of mixtures of nucleotides produced by total enzymatic hydrolysis (31) and of oligonucleotides from RNase T1 digestion. The latter analysis, carried out directly on the total rRNA digest, provides accurate molecular mass values for oligonucleotide products, which can in turn be converted to base compositions (34) and correlated with specific oligonucleotide sequences in the rRNA through comparison with the corresponding gene sequence (35).

A chromatogram based on UV detection from LC/MS analysis of a total nucleoside digest of 16 S rRNA is shown in Fig. 1 and indicates the presence of three modified nucleotides, acpU, m6A, and m3A. The assignments shown are based on HPLC retention times compared with tabulated values for RNA nucleoside standards (31) and on mass values for the protonated molecule and the protonated base produced as a fragment ion (31). The molar ratio of m3A:m6A is approximately 2, based on chromatographic peak areas by UV detection using mixture standards (data not shown). The presence of acpU was unexpected, but of particular interest because it previously was known to occur only in tRNA (36). Both the masses of the protonated nucleoside and base (346 and 214, respectively) and the relatively early retention time of 6.4 min are highly distinctive compared with values for other RNA nucleosides (31, 36). Examination of the mass spectra recorded during HPLC elution of A, U, G, and C provided no evidence for additional modified nucleotides that might have co-eluted with the major nucleotides and not have been evident in the UV
Posttranscriptional Modification in H. volcanii 16 S rRNA

To assign each modified nucleoside to the rRNA sequence, molecular masses of oligonucleotides from an RNase T1 digest of the 16 S rRNA were determined by LC/ESI-MS for comparison with calculated masses of (unmodified) T1 oligonucleotides predicted from the gene sequence (35). Shown in Fig. 2 are the masses for absorbance at 250 nm (extracted from photodiode array data; panel A) and for base fragment ions of the modified nucleosides (acp3U, m6A, m2A) expected from results of the initial modification screen shown in Fig. 1 (extracted from the high cone voltage scan function; panels B–D). The chromatographic profiles generated by these three ions (m/z 212, 148, 162) thus mark the elution times of oligonucleotides that contain them. These mass channels were time aligned with total ion current profiles from a normal cone voltage scan function, recorded in the same analysis, from which full mass spectra of the modified oligonucleotides were derived.

Shown in Fig. 3 are five summed mass spectra across the apex of the peak eluting at 25.8 min (Fig. 2, B and C), which is expected to include m6A- and acp3U-containing T1 oligonucleotides. Although the coincident elution times of the two base fragment ions suggests that they may belong to the same oligonucleotide component, the corresponding molecular mass values dictate their presence in different oligonucleotides, as follows. Comparison of the measured masses of the three oligonucleotides A, B, and C with all masses predicted from the gene sequence (19) allowed component B (M2, 2574.5) to be readily assigned as unmodified 1093-UACAUUAGp-1100 (calculated M2, 2574.5). The relative masses of the remaining oligonucleotides (A, M1, 2305.5 and C, M2, 2673.8) are not present in the calculated T1 catalog, so each one contains one of the two modified nucleotides. To derive the T1 oligonucleotide to which each belongs, the residue mass of each of the modified residues (14 Da for methyl in m6A; 101 Da for aminocarboxypropyl in acp3U) was subtracted in turn from the molecular masses of oligonucleotides A and C. Allowable T1 compositions were obtained only for M1, 2305.5 – 14 and for M2, 2673.8 – 101. The m6A is therefore confined to the oligonucleotide UAACAGp, whereas the acp3U can be accommodated within either of two (A3,C3,U)Gp sequences. The base composition of the acp3U-containing RNase T1 fragment inferred from the measured molecular mass was independently confirmed by isolation of the 8-mer oligonucleotide by anion exchange and reversed-phase chromatographies (35), digestion to nucleosides, and LC/MS analysis. These results (data not shown) confirm the composition acp3U plus (A3,C3)Gp (derived from chromatographic peak heights) with no unmodified uridine. The published sequence (19) indicated unspecified modified A and U in the sequences UAACAGp and ACUCAACp, respectively; the mass data in the present study define the corresponding modified oligonucleotides as 1498-UAm3CAAGp-1505 and 964-Acap3UCAACGp-971 (E. coli numbering).

The remaining modified nucleotide, m5A, is present in an oligonucleotide component eluting at 30.7 min (Fig. 2D); five summed mass spectra spanning the apex of this peak are shown in Fig. 4. Subtraction of the modification element CH3 = 2 (28 Da for one m5A) from the indicated molecular mass of the single oligonucleotide in this peak does not yield a mass value allowed from the gene sequence-based T1 catalog. Subtraction of CH4 = 4 (56 Da), however, yields an allowed composition of (A3,C3,U)Gp, represented in five oligonucleotides in the RNA. Reference to the original RNase T1 catalogs (19) allows assignment of this modified oligonucleotide to one of three AAUCUGp oligonucleotides: 1518-m5A3UUAUAGp-1523 (E. coli numbering).

In summary, the experiments described revealed the presence of three posttranscriptionally modified species, each of which was localized to a specific sequence location within four sites in the RNA. Structure assignments and sequence locations (E. coli numbering) are summarized in Table 1. No evidence was found for an additional modified C assigned from...
RNase catalog data, 1401-GCCCGp-1405 (19) (see “Discussion”), as a result of two key experiments: failure to observe additional modified nucleoside(s) in the total nucleoside digest (Fig. 1) and failure to find additional base ions released during analysis of RNase T₁ digestion products as in Fig. 2, B–D. Specific attention was paid to mass values corresponding to the known modified C bases in RNA (36) and to any other low mass fragment ions that could be candidates for new bases of unknown or unexpected structure, which would have been revealed in the analysis shown in Fig. 2. In addition, no oligonucleotide molecular mass values were found within the ≥3-mer products (represented by Fig. 2A) that were unassignable based on expected RNA masses calculated from the gene sequence. However, this latter experiment alone is not considered conclusive because of the difficulty in making M₁ measurements in the (mass spectrally) complex 3- and 4-mer elution region in the chromatogram. Finally, no modified nucleosides or sites were found that might have evaded detection in small oligonucleotides (e.g. NGp) from partial RNase sequencing and cataloging (19).

DISCUSSION

In the present study three different modified nucleotide species were structurally identified and placed at four sites in the 16 S rRNA sequence. These modification sites (Table I) correlate with those reported from RNase T₁ catalogs in conjunction with the corresponding gene sequence (19), but a modified cytosine (shown in the reported sequence as 1401-GCCCG-1405 (19)) was not found in the total nucleoside digest, as base fragment ion in LC/MS analysis of the RNase T₁ digest, or in oligonucleotide molecular masses reflecting incremental mass additions to M₁ 1278.8 (unmodified CCCGp). Were such a modified C present (and sufficiently stable to survive isolation and digestion protocols), it would have to elute underneath one of the seven nucleosides apparent in Fig. 1, would have to have a base fragment coincident in mass with a limited number of ubiquitous sugar-phosphate backbone fragment ions (present in significant excess in the analysis shown in Fig. 2), and in any instance would have a previously unknown structure.

The finding of acp³U in H. volcanii 16 S rRNA (as 964-Acapʰ³UCAACGp-971) was unexpected. This hypermodified uridine derivative (37, 38) was previously unknown in rRNA. It occurs in bacterial tRNAs, where it is restricted to position 47 in the extra loop (21 of 45 reported sequences in which acp³U occurs) or positions 20, 20a, or 20b in the dihydrouridine loop (24 sequences) (10). Although there are 23 different modifications reported in rRNA from all sources (36), the structures of only two uridine derivatives, pseudouridine and 2′-O-methyluridine, were previously established in archaeal SSU RNAs (39). The SSU rRNA position analogous to the location of acp³U in H. volcanii (E. coli 966) appears to be nearly universally modified in other phylogenetic domains (e.g. bacteria (24) and human and yeast (11)), although the extent of conservation between domains of the modified structure itself is limited but

![TABLE I](image)

| Molecular mass (error, Da) | Modified residue | Sequence                  |
|----------------------------|-----------------|---------------------------|
| 2673.8 (0.1)               | acp³U           | 964-Acapʰ³UCAACGp-971     |
| 2305.5 (0.1)               | m⁶A             | 1498-UAm⁶ACAApG-1504      |
| 1995.3 (0)                 | m⁰A             | 1518-m⁰Am⁰AUCUApG-1523    |

* See Figs. 3 and 4.
* See Figs. 1 and 2.
* E. coli numbering; derived (35) in conjunction with the 16 S rRNA gene sequence (19).
intriguing. As shown in Fig. 5 the 966 position in *E. coli* is occupied by N7-methylguanosine, adjacent to a second modification, 5-methylcytidine-967. Oligonucleotide catalog data from other bacterial SSU rRNAs (Halobacteroides halobus (40) and *Brevisacculus brevis* (41)) support the same tandem modification pattern of 966-GGC-967. However, in eukaryotes (11) N-966 is usually a conserved U that appears to be commonly modified to the structurally related compound 1-methyl-3-(3-amino-3-carboxypropyl) pseudouridine (23) (structure in Fig. 5).

RNase T1 catalogs of SSU rRNAs have been published by Woese and co-workers for *S. solfataricus*, *S. acidocaldarius*, and *T. tenax* (25), for *M. jannaschii* (27), and for 15 additional methanogens (17) in addition to *H. volcanii* (19). In all of these archaea the position analogous to 966 is designated as an unknown modified residue N in the conserved sequence 966-NCACG. The extent to which the assignment of N-966 as acp3U is common to archaea must await study of a broader range of organisms.

In tRNA, the side chain of nucleoside acp3U was found to be biosynthesized from adenosynmethionine, based on 14C and 3H labeling (42). Although the occurrence of acp3U in tRNA, and possibly rRNA, is conserved, its biological function is presently unknown. The effect of side chain substitution by the 3-amino-3-carboxy moiety at N-3 of uridine was studied by recently unknown. The effect of side chain substitution by the 3-amino-3-carboxy moiety at N-3 of uridine was studied by

---

**REFERENCES**

1. Agria, P. F. (1996) *Prog. Nucleic Acids Res. Mol. Biol.* 53, 79–129

2. Davis, D. R. (1998) in *Modification and Editing of RNA* (Grosjean, H., and Benne, R., eds) pp. 85–102, ASM Press, Washington, D. C.

3. Watase, K., Shimo, M., and Oshima, T. (1976) Biochem. Biophys. Res. Commun. 72, 1137–1144

4. Edmonds, C. G., Crain, P. F., Gupta, R., Hashizume, T., Hocart, C. H., Kowalak, J. A., Pomerantz, S. C., Stetter, K. O., and McCloskey, J. A. (1991) *J. Bacteriol.* 173, 3138–3148

5. Yokoyama, S., and Nishimura, S. (1995) in *RNA: Structure, Biosynthesis, and Function* (Soll, D., and RajBhandary, U. L., eds) pp. 297–233, ASM Press, Washington, D. C.

6. Ofengand, J., and Fournier, M. J. (1998) in *Modification and Editing of RNA* (Grosjean, H., and Benne, R., eds) pp. 229–253, ASM Press, Washington, D. C.

7. Tollervey, D. (1996) *Science* 273, 1056–1057

8. Barcheller, J.-P., and Cavaille, J. (1999) in *Modification and Editing of RNA* (Grosjean, H., and Benne, R., eds) pp. 255–272, ASM Press, Washington, D. C.

9. Woese, C. R., Gutell, R., Gupta, R., and Noller, H. F. (1983) *Microbiol. Rev.* 47, 621–669

10. Sprinzal, M., Horn, C., Brown, M., Ioudovitch, A., and Steinberg, S. (1998) *Nucleic Acids Res.* 26, 148–153

11. Maden, B. E. H. (1990) *Prog. Nucleic Acids Res. Mol. Biol.* 39, 241–293

12. Limbach, P. A., Crain, P. F., and McCloskey, J. A. (1994) *Nucleic Acids Res.* 22, 2183–2196

13. Bakin, A., Kowalak, J. A., McCloskey, J. A., and Ofengand, J. (1994) *Nucleic Acids Res.* 22, 3681–3690

14. Kowalak, J. A., Bruenger, E., Hashizume, T., Pelletier, J. M., Ofengand, J., and McCloskey, J. A. (1996) *Nucleic Acids Res.* 24, 688–693

15. Ofengand, J., and Bakin, A. (1997) *J. Mol. Biol.* 266, 246–268

16. Choi, Y. C., and Busch, H. (1978) *Biochemistry* 17, 2551–2560

17. Bales, W. E., Fox, G. E., Magrum, L. J., Woese, C. R., and Wolfe, R. S. (1979) *Microbiol. Rev.* 43, 260–296

18. Stackebrandt, E., Fowler, V. J., and Woese, C. R. (1983) *Syst. Appl. Microbiol.* 4, 326–337

19. Gupta, R., Lanter, J. M., and Woese, C. R. (1983) *Science* 221, 656–659

20. Brimacombe, R., Mitchell, P., Oswalt, M., Stade, K., and Bohrhar, D. (1985) *FASEB J.* 7, 161–167

21. Maden, B. E. H. (1986) *J. Mol. Biol.* 190, 681–699

22. Raue, H. A., Kloutwijk, J., and Musters, W. (1988) *Prog. Biophys. Mol. Biol.* 51, 77–129

23. Saponara, A. G., and Enger, M. D. (1974) *Biochim. Biophys. Acta* 349, 61–77

24. Bristos, J., Palmer, M. L., Kennedy, P. J., and Noller, H. F. (1978) *Proc. Natl. Acad. Sci. U. S. A.* 75, 4801–4805

25. Woese, C. R., Gupta, R., Hahn, C. M., Zilling, W., and Tu, J. (1984) *Syst. Appl. Microbiol.* 5, 97–105

26. Olsen, G. J., Pace, N. R., Neille, M., Kaine, B. P., Gupta, R., and Woese, C. R. (1987) *J. Mol. Biol.* 202, 301–307

27. Jones, W. A., Leigh, J. A., Mayer, F., Woese, C. R., and Wolfe, R. S. (1983) *FASEB J.* 7, 161–167

28. Gupta, R. (1984) *J. Biol. Chem.* 259, 9461–9471

29. Moazed, D., Van Stolk, B. J., Douthwaite, S., and Noller, H. F. (1986) *J. Mol. Biol.* 191, 483–493

30. Crain, P. F. (1990) *Methods Enzymol.* 193, 857–865

31. Pomerantz, S. C., and McCloskey, J. A. (1990) *Methods Enzymol.* 193, 794–824

32. Apfel, A., Chakel, J. A., Fischer, S., Lichtenwalter, K., and Hancock, W. S. (1997) *Anal. Chem.* 69, 1320–1325
33. Crain, P. F., Ruffner, D. E., Ho, Y., Qiu, F., Rozenski, J., and McCloskey, J. A. (1999) in *Mass Spectrometry in Biology and Medicine* (Burlingame, A. L., Carr, S. A., and Baldwin, M. A., eds) pp. 531–551, Humana Press, Totowa, NJ
34. Pomerantz, S. C., Kowalak, J. A., and McCloskey, J. A. (1993) *J. Am. Soc. Mass Spectrom.* 4, 204–209
35. Kowalak, J. A., Pomerantz, S. C., Crain, P. F., and McCloskey, J. A. (1993) *Nucleic Acids Res.* 21, 4577–4585
36. Rozenski, J., Crain, P. F., and McCloskey, J. A. (1999) *Nucleic Acids Res.* 27, 186–197
37. Ohashi, Z., Maeda, M., McCloskey, J. A., and Nishimura, S. (1974) *Biochemistry* 13, 2620–2625
38. Friedman, S. L., Li, H. J., Nakanishi, K., and Van Lear, G. (1974) *Biochemistry* 13, 2932–2937
39. Noon, K. R., Bruenger, E., and McCloskey, J. A. (1988) *J. Bacteriol.* 180, 2883–2898
40. Oren, A., Weiburg, W. G., Kessel, M., and Woese, C. R. (1984) *Syst. Appl. Microbiol.* 5, 58–70
41. Kop, J. A., Kopylov, A. M., Magrum, L., Siegel, R., Gupta, R., Woese, C. R., and Noller, H. F. (1984) *J. Biol. Chem.* 259, 15287–15293
42. Nishimura, S., Taya, Y., Kuchino, Y., and Oashi, Z. (1974) *Biochem. Biophys. Res. Commun.* 57, 702–708
43. Smith, W. S., Nawrot, B., Malkiewicz, A., and Agris, P. F. (1992) *Nucleosides Nucleotides* 11, 1683–1694
44. Stuart, J. W., Basti, M. M., Smith, W. S., Forrest, B., Guenther, R., Sierzputowska-Gracz, H., Nawrot, B., Malkiewicz, A., and Agris, P. F. (1986) *Nucleosides Nucleotides* 15, 1099–1028
45. Chan, Y. L., Gutell, R., Noller, H. F., and Wool, I. G. (1984) *J. Biol. Chem.* 259, 224–230
46. Woese, C. R., Blanz, P., and Hahn, C. M. (1984) *Syst. Appl. Microbiol.* 5, 179–175
47. Ofengand, J., Bakin, A., Wrzesinski, J., Nurse, K., and Lane, B. (1995) *Biochem. Cell Biol.* 73, 915–924
48. McCarroll, R., Olsen, G. J., Stahl, Y. D., Woese, C. R., and Sogin, M. L. (1983) *Biochemistry* 22, 5858–5868
49. Mitchell, P., Oswald, M., and Brimacombe, R. (1992) *Biochemistry* 31, 3004–3011
50. van Knippenberg, P. H., Van Kimmenade, J. M., and Heus, H. A. (1984) *Nucleic Acids Res.* 12, 2595–2604
51. van Knippenberg, P. H. (1985) in *Structure, Function, and Genetics of Ribosomes* (Hardesty, B., and Kramer, G., eds) pp. 412–424, Springer-Verlag New York Inc., NY
52. Woese, C. R., Kandler, O., and Wheelis, M. L. (1990) *Proc. Natl. Acad. Sci. U. S. A.* 87, 4576–4579
53. Fellner, P. (1969) *Eur. J. Biochem.* 11, 12–27
54. Youvan, D. C., and Hearst, J. E. (1981) *Nucleic Acids Res.* 9, 1723–1741
55. Tautz, D., Hancock, J. M., Webb, D. A., Tautz, C., and Dover, G. A. (1988) *Mol. Biol. Evol.* 5, 366–376

*Posttranscriptional Modification in H. volcanii 16 S rRNA* 24489