Influence of Phylogeny on Posttranscriptional Modification of rRNA in Thermophilic Prokaryotes: The Complete Modification Map of 16S rRNA of *Thermus thermophilus*†

Rebecca Guymon,‡ Steven C. Pomerantz,‡ Pamela F. Crain,‡ and James A. McCloskey*,‡,§

Departments of Medicinal Chemistry and Biochemistry, University of Utah, Salt Lake City, Utah 84112

Received December 19, 2005; Revised Manuscript Received February 13, 2006

ABSTRACT: Posttranscriptional modification in RNA generally serves to fine-tune and regulate RNA structure and, in many cases, is relatively conserved and phylogenetically distinct. We report the complete modification map for SSU rRNA from *Thermus thermophilus*, determined primarily by HPLC/electrospray ionization MS-based methods. *Thermus* modification levels are significantly lower, and structures at the nucleoside level are very different from those of the archaeal thermophile *Sulfolobus solfataricus* growing in the same temperature range [Noon, K. R., et al. (1998) *J. Bacteriol. 180*, 2883−2888]. The *Thermus* modification map is unexpectedly similar to that of *Escherichia coli* (11 modified sites), with which it shares identity in 8 of the 14 modifications. Unlike the heavily methylated *Sulfolobus* SSU RNA, *Thermus* contains a single ribose-methylated residue, N^4,2′-*O*-dimethylcytidine-1402, suggesting that O-2′-ribose methylation in this bacterial thermophile plays a reduced role in thermostabilization compared with the thermophilic archaea. Adjacent pseudouridine residues were found in the single-stranded 3′ tail of *Thermus* 16S rRNA at residues 1540 and 1541 (*E. coli* numbering) in the anti-Shine−Dalgarno mRNA binding sequence. The present results provide an example of the potential of LC/MS for extensive modification mapping in large RNAs.

Posttranscriptional modification is a nearly universal feature of the principal RNAs involved in translation of the genetic code (*1*, *2*). Individual nucleosides formed by this process are structurally diverse, generally tend to occur at conserved sequence sites, and many are phylogenetically distinct (*3*). Most detailed studies of structure−function relationships of RNA modifications have dealt with tRNA,[1] which generally possesses the greatest concentrations of modified residues, typically 8−12%, and for which the greatest amount of detailed sequence data are available (*4*). As a consequence, although many aspects of the complex details of modification−function relationships in tRNA remain obscure, significant progress in understanding has been made in recent years, for example, with regard to influence of modification on codon selection, control, and stabilization of tRNA structure and on maintenance of translational reading frames during protein synthesis (*2*).

By contrast, in rRNA, in which modification levels tend to be lower than those in tRNA (particularly in prokaryotes), far less is known concerning occurrence of modified nucleosides with regard to either structure or their functions. Essentially complete modification maps of small subunit (16/18S) or large subunit rRNAs (23/28S), in which both the chemical identities and sequence locations of modified residues are known, are limited to perhaps five eukarya (*5*), one bacterium [*Escherichia coli* 16S (leading citations in ref *6*) and 23S (tabulation summary and references in ref *3*) rRNAs], and one archaeon [*Haloferax volcanii* 16S rRNA (*7*, *8*)]. To this may be added numerous partial oligonucleotide sequences (*9*) from early work on RNase T1 cataloging of 16S rRNA in which the position but not identity of modification was often reported (e.g., ref *10*) for the goal of establishing phylogenetic status of the source organism (*11*). Also, a significant body of work has involved studies of the occurrence and distribution of pseudouridine, which is perhaps the most common of RNA modifications (*12*, *13*) and often found at conserved positions in rRNA and tRNA.

Circumstantial evidence was reported that known structurally stabilizing modifications, particularly ribose methylation at O-2′, may be used in rRNA of archaeal thermophiles for stabilization in the thermal environment. The archaeon *Sulfolobus solfataricus*, grown at 75 °C, was found to have

† Supported by NIGMS Grant GM29812 from the National Institutes of Health.

* Address correspondence to this author at the Department of Medicinal Chemistry, University of Utah. Telephone: (801) 581-5581. Fax: (801) 581-7457. E-mail: mccloskey@biochem.utah.edu.

‡ Department of Medicinal Chemistry, University of Utah.

§ Department of Biochemistry, University of Utah.

[1] Abbreviations: tRNA, transfer RNA; rRNA, ribosomal RNA; BAP, bacterial alkaline phosphatase; LC/ESI-MS, directly combined high-performance liquid chromatography/electrospray ionization mass spectrometry; TOF, time of flight; MS/MS, tandem mass spectrometry; MALDI, matrix-assisted laser desorption ionization; HFIP, 1,1,1,3,3,3-hexafluoro-2-propanol; TEA, triethylamine; E_{lab}, collision energy in the laboratory frame of reference; Ψ, pseudouridine (5-ribosyluracil); CID, collision-induced dissociation; MS-1 and MS-2, first and second mass analyzers, respectively, of a tandem mass spectrometer; MRM, multiple reaction monitoring; SSU, small ribosomal subunit; LSU, large ribosomal subunit; A>p, adenosine 2′,3′-cyclic phosphate; m^3U, 3-methyluridine; m^4Cm, N^4,2′-*O*-dimethylcytidine; m^5C, 5-methylcytidine; m^2G, N^2-methylguanosine; m_2^2G, N^2,N^2-dimethylguanosine; m^7G, 7-methylguanosine; m_2^6A, N^6,N^6-dimethyladenosine; Cm, 2′-*O*-methylcytidine; Am, 2′-*O*-methyladenosine.

10.1021/bi052579p CCC: $33.50
Published on Web 03/28/2006

approximately 38 modified sites in 16S rRNA and 50 sites in 23S rRNA, the greatest numbers reported for any prokaryote (*14*). The majority of modified residues were ribose methylated, three of which (2′-*O*-methyladenosine, 2′-*O*-methylcytidine, 2′-*O*-methylguanosine) showed increased molar levels at progressively higher culture temperatures.

Here we report the first posttranscriptional modification map for the 16S rRNA of a bacterial thermophile, *Thermus thermophilus* [growth optimum 75 °C (*15*)], including identities and sequence locations of modified residues. These data permit detailed comparison with modifications in *E. coli* 16S rRNA and allow assessment of the extent to which the extensively studied *E. coli* SSU rRNA may be representative of bacteria in general. By comparison with the highly modified 16S rRNA of the archaeal thermophile *S. solfataricus*, for which the extent and identities, but not sequence locations, of modifications have been reported (*14*), an effort was made to understand the extent to which modifications in a bacterial thermophile may be used for structural stabilization of rRNA compared with rRNA from the archaeal thermophiles.

The paucity of modification distribution data in rRNA beyond a relatively few organisms can be attributed in part to experimental difficulties in carrying out such work with large ribosomal RNAs [M_r ($\sim$1.5−3.9) × 10^6] and until recent years also to the limited availability of rRNA gene sequences, which are crucial in providing a framework for placement of oligonucleotide sequences. Mass spectrometry-based techniques have recently been shown to be effective in the identification and placement of modifications in large RNAs or subsections thereof, using either electrospray ionization (*16*−*18*) or MALDI techniques (*19*−*22*). The LC/ESI-MS-based approaches employed in the present study constitute the most extensive use of mass spectrometry for modification mapping in large RNAs and demonstrate a degree of relief from experimental difficulties associated with the mapping problem.

MATERIALS AND METHODS

Preparation and Enzymatic Digestion of rRNA. 30S ribosomal subunits from *T. thermophilus* HB8 cells, grown at 70 °C and harvested at mid-log phase, were a gift from V. Ramakrishnan, University of Utah (present address: MRC Laboratory of Molecular Biology, Cambridge, U.K.). 16S RNA was extracted with TRI Reagent (*23*) following the manufacturer's protocol. 16S rRNA (typically 100 μg, 200 pmol) was digested totally to nucleosides with nuclease P1, phosphodiesterase I, and BAP (*24*). Aliquots of RNA (5−100 pmol) were digested with 1000 units of RNase T1 (Ambion) in 10 mM Tris plus 1 mM EDTA, pH 7, for 30 min at 37 °C or 45 min at 55 °C. Alternatively, the RNA was digested with RNase U2 (Industrial Research Ltd., Lower Hutt, New Zealand) in 20 mM diammonium citrate, pH 5, and 1 mM EDTA or in 20 mM diammonium citrate, pH 5, 1 mM EDTA, and 8 M urea, with 5−10 units of RNase U2 for 15 min at 60 °C, then an additional 5−10 units of enzyme was added, and digestion was continued for 15 min.

LC/ESI-MS Analysis of Nucleosides from Total Digests of 16S rRNA. Nucleoside digests were analyzed using an HP 1090 liquid chromatograph with diode array detector, interfaced (without flow splitting) to a Micromass Quattro II mass spectrometer; MassLynx version 3.4 software was used for instrument control and data processing. Typical tuning parameters were capillary voltage in the range of −3.0 to −3.2 kV and cone voltage 40−50 V. Nucleosides were separated on a Luna C18 (Phenomenex) 250 × 2 mm column, using a multilinear gradient as described (*25*), except that the flow rate used was 300 μL/min and the composition of buffer A was reduced to 5 mM. Nucleosides were also separated on a Develosil RP-aqueous (C30) (Phenomenex) reversed-phase column (250 × 2 mm) using the same buffer system as above with a variation on the above multilinear gradient (*26*) to get better separation of early eluting nucleosides from impurities in some samples. The digest was injected into the chromatograph without prior desalting or enzyme removal. Identification of modified nucleosides was made primarily from their mass spectra and relative retention times as earlier described (*25*).

LC/ESI-MS of Oligonucleotides from RNase T1 and U2 Digests of 16S rRNA. Oligonucleotide mixtures were analyzed on either the HP1090/Quattro II system described above or a Waters CapLC chromatograph interfaced to a Micromass Q-Tof2 (Q-TOF) high-resolution instrument with a MassLynx 3.4 data system. An HFIP−TEA/methanol solvent system (*27*) was used for separations on C18 reversed-phase columns, either a 150 × 1 mm Luna C18 (Phenomenex) column with a 15 × 1 mm OptiGuard C18 precolumn (Optimize Technologies) eluted at 60 μL/min (Quattro II system) or a 150 × 0.5 mm Luna C18 (Phenomenex) column with a 300 μm × 0.5 mm μ-Precolumn Cartridge C18 (LC Packings) eluted at 15 μL/min (Q-TOF system).

Detection of Modified Nucleotides in RNase Digestion Products Using Characteristic Low-Mass Fragment Ions. Oligonucleotides that contained modified residues were detected using an additional LC/MS scan function that utilized elevated declustering ("cone") voltage to produce modification-diagnostic fragments (such as B^- and $N{>}p^-$ ions), whose time vs abundance profiles were generated for comparison with normal-scan total ion current profiles to identify the oligonucleotide that generated them (*8*, *17*). These low-mass fragment ions are tabulated in Table S1 (see Supporting Information).

Sequence Determination of RNase T1 and U2 Digestion Products by Tandem Mass Spectrometry (LC/ESI-MS/MS). Modified oligonucleotides were directly sequenced by MS/MS in digestion mixtures using the Q-TOF instrument system, with preselection of sequencing targets made essentially from the molecular mass and monomer ion data shown in Table 1. Sequence data were acquired during the approximately 20 s period of passage of the component of interest through the mass spectrometer ion source. Typical parameters were as follows: TOF mass analyzer data accumulated from *m*/*z* 100 to *m*/*z* 1990 in 0.45 s, with 0.05 s delay; collision energy range 15−80 eV (E_{lab}) depending on oligonucleotide charge state and length. Sequences were derived using previously described procedures and algorithms (*28*, *29*) recently extended to include provision for the presence of Ψ (*30*) (see the following section). Multiply charged ions in product ion mass spectra have been converted to their singly charged equivalent in the relevant figures using the MassLynx MaxEnt3 algorithm.

Table 1: Modified Nucleosides in *T. thermophilus* 16S rRNA[a]

	RNase T1			RNase U2		
nucleoside	sequence	M_r meas (error)[b]	monomer ions, *m/z*[c]	sequence	M_r meas (error)[b]	monomer ions, *m/z*[c]
Ψ-516	516-**Ψ**Gp-517	669.13 (0.05)		511-CUCCG**Ψ**GCCA>p-520[d] (Figure 3)	3156.07 (0.33)	164, 189, 207, 225
m^7G-527	525-CC**m^7G**CGp-529[e] (Figure S2)	1637.24 (0.01)	164	525-CC**m^7G**CGGUA>p-532	2599.09 (0.27)	164[f]
m_2^2G-966	964-AA**m_2^2G**>p-966[e] (Figures S3 and S4)	1031.28 (0.10)	178, 372	965-A**m_2^2Gm^5C**AA>p-969[d] (Figures 4 and S4)	1679.12 (0.17)	124, 178, 372
m^5C-967	967-**m^5C**AACGp-971	1645.39 (0.14)	124, 336			
m^2G-1207	1207-**m^2G**>p-1207	359.06 (0.02)	164, 358	1205-UG**m^2G**CCCUUA>p-1213[d] (Figures 5 and S5)	2866.05 (0.31)	164, 358, 376
m^5C-1400	1393-UACACAC**m^5C**Gp-1401[d] (Figures 6 and 7)	2890.63 (0.21)	124, 336	1399-C**m^5C**G**m^4C**m**C**m**^{5}C**GU**m^5C**A>p-1408[d] (Figures 8 and 9)	3225.14 (0.36)	124, 332, 336
m^4Cm-1402	1402-**m^4C**m**C**m^5C**Gp-1405[e] (Figure S6)	1320.34 (0.13)	124, 336, 350			
m^5C-1404						
m^5C-1407	1406-U**m^5C**ACGp-1410[e] (Figure S7)	1622.36 (0.13)	124, 336			
m^3U-1498	1498-**m^3U**AACAAGp-1504[e] (Figure S8)	2304.5 (0.17)	125, 319, 337	1494-GUCG**m^3U**A>p-1499	1950.06 (0.19)	*g*
m_2^6A-1518, 1519	1518-**m_2^6A**m**$_2^6$A**Gp-1520	1077.32 (0.10)	162, 356, 374	1514-CCGG**m_2^6A**m**$_2^6$A**GG>p-1521[e] (Figure S9)	2704.14 (0.29)	162, 356
Ψ-1540, 1541	1531-AUCACCUCC**ΨΨ**UCUOH-1544	4262.87 (0.33)	164, 189, 207, 225	1535-CCUCC**ΨΨ**UCU-OH-1544[d] (Figure 10)	2993.07 (0.30)	164, 189, 207, 225

[a] *E. coli* numbering is used throughout. [b] Experimentally measured; error is found minus calculated mass for the modified oligonucleotide shown. [c] Tabulation of monomer mass ion assignments is given in Table S1 in Supporting Information. [d] Directly sequenced by MS; see Figures 3−5, 7, and 10. [e] Directly sequenced by MS; see Figures S2, S3, and S6−S9 in Supporting Information. [f] The M_r 2599.1 oligonucleotide coelutes with 1207-m^{2}GCCCUUA>p-1213, M_r 2215.3, a truncated form of the M_r 2866.1 oligonucleotide; therefore, the *m/z* 164 fragment ion assignment does not distinguish the two mGua signals. [g] Fragment ion mass signals for m^3U are not clearly resolved from background.

Detection of Pseudouridine Residues by MS/MS. Recognition of Ψ in oligonucleotides (*30*) was indicated by the occurrences of *m/z* 225 (Ψ nucleoside − $H_2O)^-$, 207 (Ψ − $2H_2O)^-$, 189 (Ψ − $3H_2O)^-$, 165 (225 − $C_2H_4O_2)^-$, or 164 (207 − HNCO)$^-$, either in primary mass spectra, at ion sampling cone ∼75 V, or in product ion mass spectra from CID of the oligonucleotide. Multiple reaction monitoring (MRM) functions were utilized for detection of Ψ in prespecified locations (*30*): *m/z* 225 (MS-1) → 165 (MS-2) for recognition of Ψ at 5′ termini; *m/z* 207 → 164 for Ψ in internal oligonucleotide positions, both using 14 eV (E_{LAB}) collision energy, dwell time 30−50 ms. Tests for ΨGp, M_r 669 (resulting from RNase T1 hydrolysis of ...GΨG... sequences), were made by MRM for *m/z* 668 → 207 at normal cone voltage and collision energy 45 eV. In the latter case the LC/MS/MS system was programed for ΨGp detection in the early part of the HPLC run (∼10 min) and for the presence of Ψ by MRM measurements (former two methods) at later elution times. In all cases careful product ion chromatographic alignments were made with oligonucleotide molecular mass chromatographic profiles derived from alternate MS scan functions to establish the identity of the Ψ-containing oligonucleotide.

Sequence Location of Pseudouridine. RNase product oligonucleotides concluded to contain Ψ were directly sequenced in the hydrolytic mixture by LC/MS/MS. In addition to basic sequence information, Ψ was placed using abundance changes in conventional sequencing ions (*28*) as described (*30*): absence of a − B ions at the site of the Ψ residue; presence of a prominent a-type ion at the Ψ position; and presence of an anomalously abundant w-type (*31*) ion immediately 3′ to the site of pseudouridylation.

Confirmation of the presence of adjacent Ψ residues in the *T. thermophilus* fragment AUCACCUCCΨΨUCU was made by preparation of a cyanoethylated derivative as described below, after the method of Kirpekar (*32*). A 165 pmol aliquot of an RNase T1 digest was HPLC-separated in the HFIP buffer system, and the fraction containing the 3′-terminal 14 residues was collected, dried in a SpeedVac, and then dissolved in 40 μL of 1.1 M triethylammonium acetate (pH 8.6)−ethanol (59:41). Eight microliters of acrylonitrile was added, and the reaction mixture was incubated at 70 °C for 3 h, dried under reduced pressure, and reconstituted in 12 μL of H_2O. An aliquot of the sample was analyzed by LC/MS on the Q-TOF instrument to determine the extent of cyanoethylation in order to establish the number of Ψ residues in the oligonucleotide.

Software for Mass Spectral Data Manipulation and Analysis. The program Mongo Oligo (http://medlib.med.utah.edu/masspec/mongo.htm) was used for mass prediction and generation of mass-ordered RNase "cut lists", derived from the corresponding *T. thermophilus* 16S rRNA gene sequence (GenBank accession M26923).

An oligonucleotide sequence extraction program (ONx) parses the gene sequence into potentially allowable corresponding RNA sequence segments and their locations using the experimental molecular mass values of RNase-derived (or other) fragments. The user provides constraints (if any) on identities of nucleotides immediately 5′ and 3′ to the cleavage site, the phosphorylation state, residue masses of potential modifications, and the number of allowable undercut residues.

SOS (*s*imple *o*ligonucleotide *s*equencer) is an interactive user-controlled graphical program for ab initio sequencing of oligonucleotides, modified or not, from their product ion mass spectra following CID (*29*).

RESULTS

Results presented in the following sections, in which posttranscriptionally modified nucleotides were identified and

placed at specific sequence locations in the rRNA molecule, were obtained using LC/ESI-MS-based methods, using combinations of the following seven protocols.

(1) A complete census of modifications present in the *Thermus* SSU RNA was obtained by positive ion LC/MS analysis of total nucleoside digests (*25*). These data also provided approximate net molar ratios of modified species (see discussion in ref *14*), with which the total rRNA modification map must ultimately be reconciled.

(2) rRNA was digested by RNase T1 to produce a mixture of several hundred oligonucleotides, each terminating with Gp-3′. Molecular masses of all oligonucleotides in the hydrolysis mixture were measured by negative ion LC/MS. Each molecular mass value was converted to exact base composition and net modification mass, if any (*33*).

(3) The experimentally derived mass values were next compared with the list of RNase T1 fragment masses calculated from the 16S rRNA gene sequence. With the exception of pseudouridine, the presence of modifications was then recognized by deviations in measured mass value from those in the calculated RNase T1-derived mass list (*34*).

(4) Alternatively, the rRNA was digested with RNase U2, which has a different cleavage specificity than RNase T1 and is therefore complementary to it. RNase U2 generally cleaves on the 3′ side of A, resulting in an A>p 3′ terminus, and in principle reduces the difficulty in placing modifications in G-rich regions (an increased problem in the case of G-rich thermophiles) when RNase T1 is used. Unfortunately, RNase U2 has a tendency to overcut by cleavage at some Gs and to undercut at A, resulting in formation of oligonucleotides having multiple As, and thus no rigorously predictable purine termini. As a consequence, RNase U2 is most effectively used in conjunction with RNase T1 experiments by calculation of potential U2 oligonucleotide masses based on mass analysis of T1 digestion products, which are then compared with the experimentally observed U2 values. We note that, in general, in our hands RNase U2 was found to be less satisfactory than T1, not only from the standpoint of fidelity of cleavage but also the reproducibility. The resulting mass and base composition values of U2-generated fragments were used to adjudicate or confirm uncertainties of modification placement, either in oligonucleotides from RNase T1 hydrolysis or in the 16S rRNA molecule.

(5) During analysis of RNase T1 or U2 digests mass spectrometer scan functions were also employed in which elevated declustering voltage (nozzle-skimmer potential difference, also termed cone or "orifice" voltage) was used to dissociate oligonucleotides after HPLC separation but prior to mass analysis, to produce a range of "fingerprint" mass signals in the low *m*/*z* range that were characteristic of modified nucleotide residues (*8*, *17*). Examples include fragments representing base ions (e.g., *m*/*z* 164 for methyl-Gua$^-$), mononucleotide ions (e.g., *m*/*z* 376 for mGp$^-$ or Gmp$^-$) or their dehydration products (*m*/*z* 358 for mG>p$^-$ or Gm>p$^-$), and *m*/*z* 207 for Ψ. Correlations between the diagnostic monomer mass signals and the oligonucleotides in the mixture from which they were derived were established by careful time alignment of chromatographic ion profiles of monomer fragment ions with molecular ion signals from the oligonucleotide [e.g., $(M - 4H)^{4-}$], recorded using normally low nozzle-skimmer potential conditions in every other scan.

(6) The detection and sequence location of pseudouridine residues, which are isomeric with uridine and thus mass-silent, were somewhat more difficult and were made by recently developed tandem mass spectrometry-based methods (*30*). The presence of Ψ in the oligonucleotide was recognized using characteristic low mass fragment ions [see (5) above] and whether the Ψ residue was in an internal or 5′-terminus position, based on monitoring of selected reaction channels using MS/MS. Placement of Ψ was made in favorable cases by abundance changes in CID spectra resulting from the presence of Ψ rather than by changes in mass as for other modified residues.

(7) Selected oligonucleotides that were believed to be modified from the foregoing experiments were directly sequenced in the RNase digest mixture by using LC/MS/MS (*28*, *29*). In general, the results of direct sequencing were somewhat redundant with, but complementary to, evidence derived from experimental mass values of RNase digestion products plus analysis of monomer fragment ion data. In many instances, mass measurement and sequencing of RNase T1 fragments, plus mass analysis of targeted RNase U2 fragments, were sufficient for unambiguous placement of the modified residue in the rRNA molecule.

Identities and Net Stoichiometry of Modified Residues in T. thermophilus 16S rRNA. Identities and approximate numbers of modified residues, established by LC/MS analysis of total nucleoside digests of rRNA (see Figure S1 in Supporting Information), are Ψ (2.9), m^3U (1.0), m^4Cm (1.2), m^5C (3.9), m^2G (1.4), m_2^2G (1.0−1.4), m^7G (1.1), and m_2^6A (1.8).

Structure Assignments and Sequence Locations of Modified Residues from RNase Digestion Products. Chromatographic separation of products from digestion using RNase T1 is shown in Figure 1A and RNase U2 in Figure 1B, each annotated to show those oligonucleotides found to be modified. A summary of modified oligonucleotides detected and placed from the RNase hydrolysates using LC/MS methods is given in Table 1 along with a listing of monomer fragment ions used for identification of the modified residue in each oligonucleotide (see also Table S1 in Supporting Information). These data show eight structurally different modified nucleosides at 14 sequence sites in the rRNA molecule.

(A) Ψ*-516.* The occurrence of the MRM reaction channel indicative of the dinucleotide ΨGp in Figure 2A implies the presence of the subsequence GΨG in the rRNA molecule. An expanded sequence context is gained from the M_r 3156 oligonucleotide eluting at 43 min in the RNase U2 digest (Table 1, Figure 1B), which contains Ψ at an internal position as demonstrated by the MRM reaction channel response at 43 min in Figure 2B. Use of the ONx program established two possible GΨG-containing subsequences from molecular mass 3156: 511-CUCCGUGCCA>p and 743-UCCAC-CCGUG>p. These two possibilities were clearly distinguished from the CID mass spectrum of the M_r 3156 oligonucleotide in Figure 3. Enhanced abundances of the w_4 and y_4 ion species and the abundant a_6 ion mark the presence of Ψ in the sixth position, excluding the second sequence above and placing Ψ at position 516 rather than 512. The unusual adenine-loss ions annotated in Figure 3 are characteristic of the A>p 3′ terminus (R. Guymon et al., manuscript

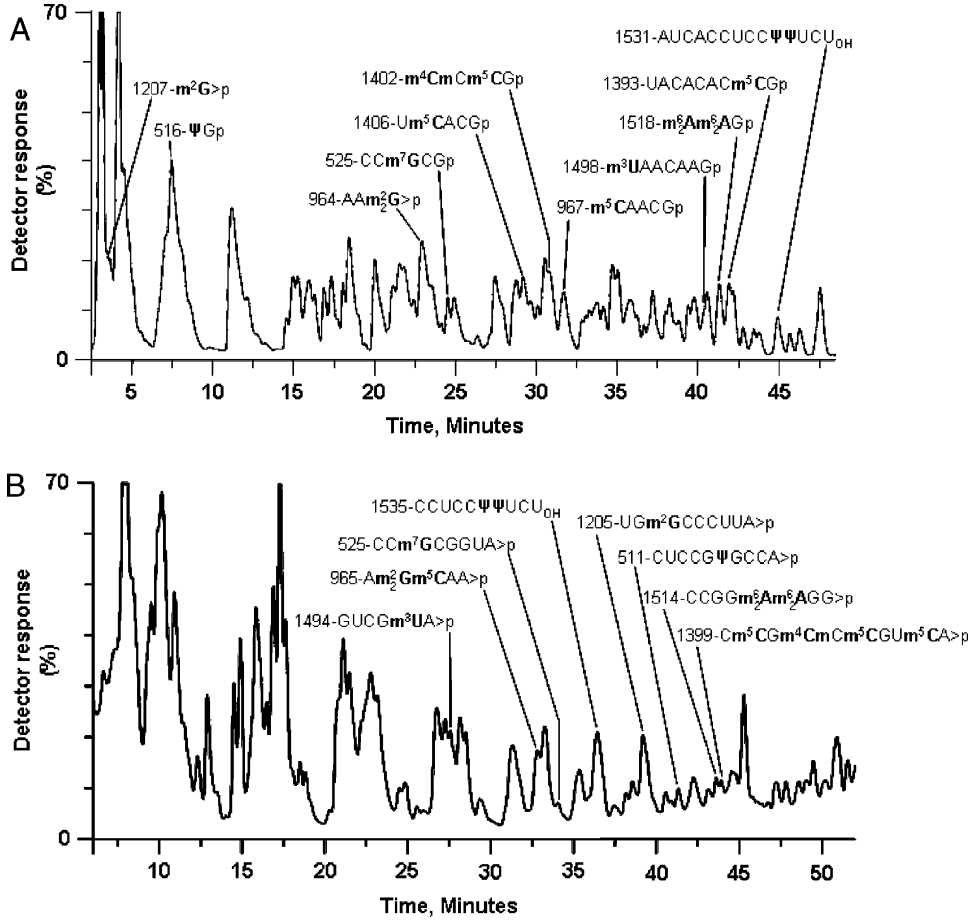

FIGURE 1: RNase hydrolysis products from *T. thermophilus* 16S rRNA showing elution positions of modified oligonucleotides determined by mass spectrometry. Detection at 260 nm. See Table 1 for corresponding mass data. Panels: A, RNase T1; B, RNase U2.

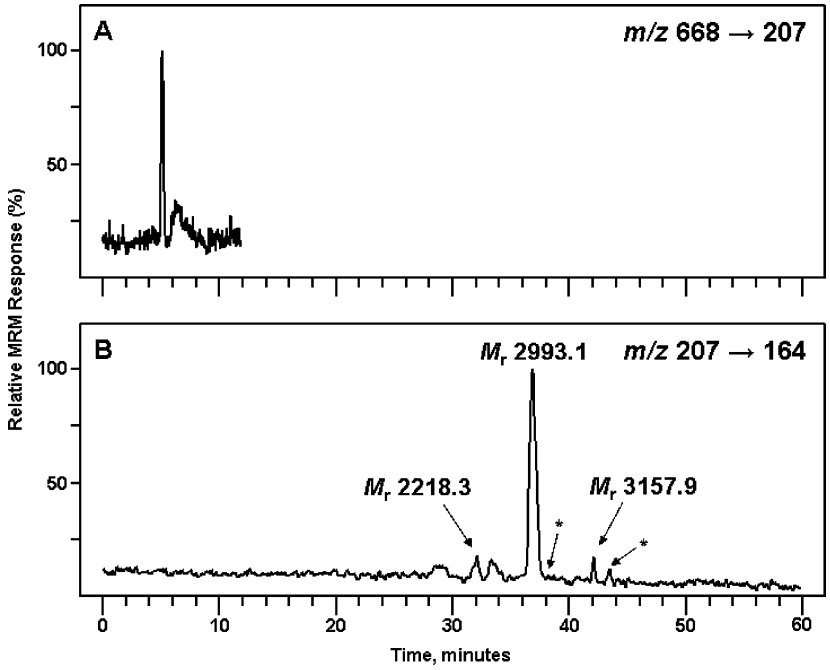

FIGURE 2: Pseudouridine detection in RNase fragments from *T. thermophilus* 16S rRNA. (A) MRM channel *m/z* 668 → 207 showing the presence of Ψ in the M_r 669 dinucleotide eluting at 7.5 min (Figure 1A, RNase T1 digest). (B) MRM channel *m/z* 207 → 164 showing the presence of internal Ψ in the RNase U2 oligonucleotides M_r 2993 (37 min, Figure 1B), M_r 3156 (43 min, also Figure 1B), and M_r 2218 (32 min, not annotated in Figure 1B) deduced to be a truncated form of M_r 3156. Mass values shown in panel B are estimates derived from an alternate scan function not using MRM (see Materials and Methods). Asterisks denote other minor enzyme miscut products in the vicinity of Ψ-516 (see text).

in preparation), resulting from RNase U2 hydrolysis that are not found in RNase T1 products (Gp 3′ termini).

The placement of Ψ-516 also permits assignment of the minor 32.3 min eluant in Figure 2B (not separately annotated

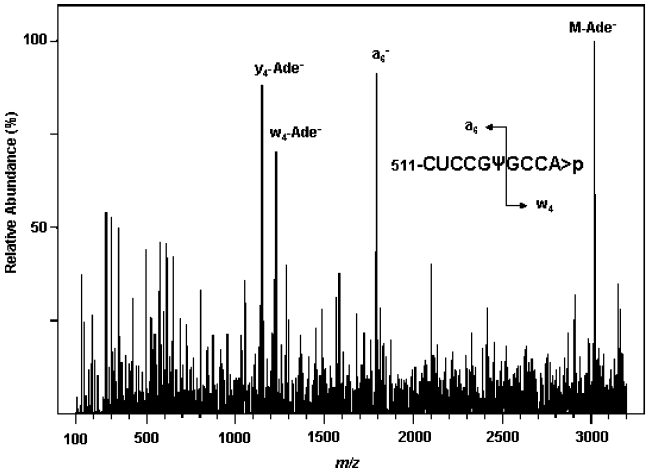

FIGURE 3: Product ion mass spectrum of RNase U2 oligonucleotide CUCCGΨGCCA>p (M_r 3156) used for placement of Ψ-516 and showing adenine-loss ions characteristic of A>p 3′-terminal oligonucleotides.

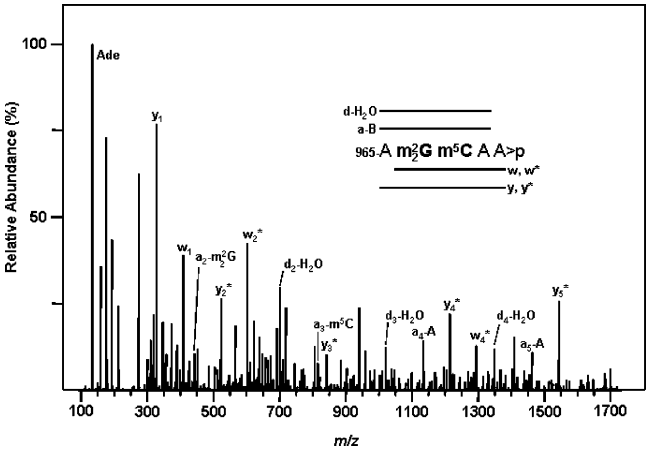

FIGURE 4: Product ion mass spectrum from CID of 965-$Am^2_2Gm^5CAA$>p-970 (Figure 1B) used for placement of modifications containing methyl groups in the loop of helix 31.

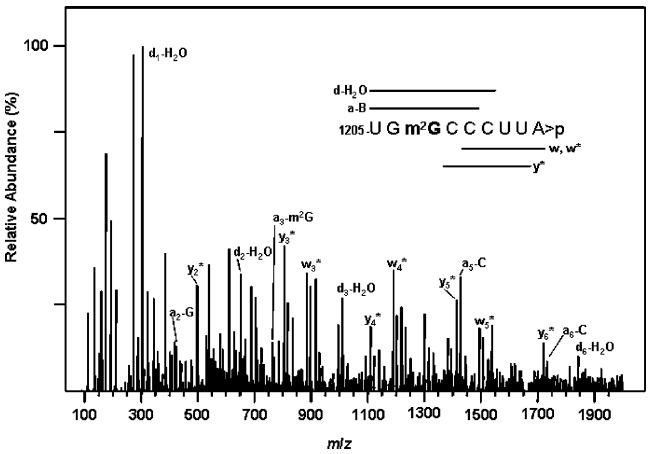

FIGURE 5: Product ion mass spectrum from CID of 1205-UGm²-GCCCUUA>p-1213 (Figure 1B) used for placement of m^2G-1207.

in Figure 1B), which contains an internal Ψ. The molecular mass associated with this response is 2217 (spectrum not shown), which corresponds to 511-CUCCGΨG>p-517, a truncated version of the M_r 3156 oligonucleotide. With the aid of the ONx program, the only rational alternative placement of the CUCCGUG sequence would be at position 91−98. However, no confirmatory sequences (M_r values cotracking with an appropriate MRM response) were found that would sustain this positioning. Ancillary evidence in support of the placement of Ψ-516, as opposed to Ψ-96, was obtained from careful analysis of the oligonucleotide mass values corresponding to the minor MRM responses labeled by asterisks in Figure 2B. Both of the M_r values associated with these responses were consistent with RNase U2 miscuts of sequences surrounding Ψ-516: M_r 2547.5, 510-ACUCCGΨG>p (37.4 min); M_r 3503.1, 511-CUCCGΨGCCAG>p (43.7 min). (Assignment of the 37 min oligonucleotide M_r 2993 is discussed in the later section on Ψ-1540,1541.)

(B) m^7G-527. The pentanucleotide sequence resulting from T1 hydrolysis (Table 1) was confirmed by direct sequencing of the fragment (Figure S2). The RNase U2 product of M_r 2599.1 supports the sequence context of the methylated G and places it uniquely at position 527 in the rRNA. Identity of the base as 7-methyl rather than N^2-methylguanine is made by the characteristic and exceptionally facile loss of the base from the molecular ion (*m*/*z* 1636 → 1471, Figure S2) and by the enhanced $a_3 - m^7Gua$ base-loss ion, *m*/*z* 723. The enhanced base loss is a characteristic effect of guanine 7-methylation. Although the sequence CCGCG occurs in three places in the rRNA, it is assigned as shown, in analogy to *E. coli* and *Proteus vulgaris* rRNAs (*9*) containing the same unusual modification sequence at this site in the loop of stem 18.

(C) m^2_2G-966, m^5C-967. The tandem modifications at 966−967, which represent modification sites that are nearly universal in bacteria (*9*), are not distinguished with regard to sequence location in the RNase T1 digest because they occur in different fragments (Table 1) but occur together in the U2 fragment (M_r 1679) which was directly sequenced (Figure 4), as was the expected T1 fragment, AAm^2_2G >p (Figure S3). The presence of the monomer base ion *m*/*z* 178 in the T1 and U2 oligonucleotides (Figure S4) confirms the location of both methyl groups to be in the guanine base and accounts for the m^2_2G residue observed in the total nucleoside digest (Figure S1).

(D) m^2G-1207. The sequence context for this modification is provided by the U2 fragment, which was sequenced (Figure 5), and as a result has a single possible placement in the rRNA molecule. The presence of monomethylguanine is indicated by ion *m*/*z* 164 in the RNase U2 (Figure S5) and T1 fragments. Position 1207 was not previously reported as modified in any SSU rRNA (*9*) (all three phylogenetic domains) with the sole exception of *E. coli*, which also has m^2G-1207.

(E) m^5C-1400, m^4Cm-1402, m^5C-1404, m^5C-1407. The composition and sequence location of the remarkable pentamethylated sequence occurring in the U2 fragment assigned as residues 1399−1408, M_r 3225.14, is uniquely defined by four lines of data: (1) the molecular masses of three T1 fragments and one U2 fragment as diagramed in Figure 6, from which a "sequence ladder" has been constructed using the four RNase fragments; (2) constraints in base sequence mandated by the 16S rRNA gene sequence in the segment 1493−1410 through use of the ONx program; (3) direct mass spectrometric sequencing of the three T1 pieces (Figures 7, S6, and S7); (4) identification of appropriate modified residues from monomer ions in all four RNase fragments

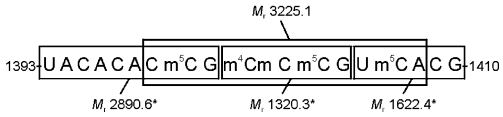

FIGURE 6: Sequence relationships of the RNase fragments from the highly methylated section of helix 44 in *T. thermophilus* 16S rRNA as defined primarily by molecular mass and sequence data. Directly sequenced oligonucleotides are denoted by asterisks.

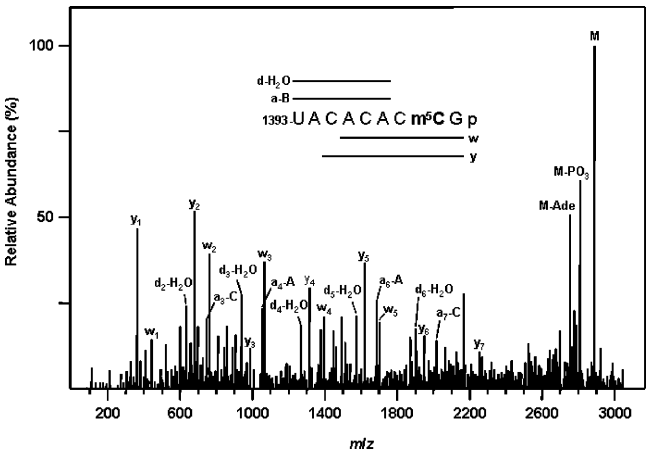

FIGURE 7: Product ion mass spectrum from CID of 1393-UACACACm^{5}CGp-1401 (Figure 1A), M_r 2891, a subcomponent of the sequence shown in Figure 6.

(Table 1), including the unusual posttranscriptionally modified nucleoside m^{4}Cm-1402 (Figure 8) (*35*).

(F) m^{3}U-1498. The sequence of the RNase T1 product M_r 2304.5 was determined by direct sequencing (Figure S8), showing a base-methylated U at the 5′ terminus, and is supported by presence of the methyluracil anion *m*/*z* 125 and nucleotide ions *m*/*z* 319 and 337 (data not shown); see Table S1. Overlap with the M_r 1950.1 U2 fragment extends the implied sequence segment to residues 1494−1504 in agreement with the corresponding gene sequence. These data uniquely define the rRNA position and identity of the 1498 modification, as also concluded from analogous experiments with *E. coli* 16S rRNA (*34*), which contains the same modified sequence.

(G) m$_2^6$A-1518, 1519. Molecular masses 1077 from RNase T1 and 2704 from RNase U2 hydrolysis are consistent with the tandem m$_2^6$A-containing sequences shown in Table 1 and represent a modification motif at the apex of helix 45 that is conserved in nearly all SSU rRNAs (*36*, *37*) in all three phylogenetic domains (*9*). The assignment is supported by the presence of the characteristic dimethyladenine base ion *m*/*z* 162 and by the sequence mass spectrum of the U2 digestion product (Figure S9), indicating the presence of adjacent dimethyladenines in the expected sequence context.

(H) Ψ-1540, 1541. The strongest signals for detection of internal-Ψ residues by the MRM method occur at 36.2 min in the RNase U2 digest (Figure 2B) and 46.2 min in the T1 digest (Figure 9). Minor responses (data not shown) were also observed at the same retention times in the 5′-Ψ MRM detection channel, each corresponding to ∼0.05 of the internal-Ψ signals for the same oligonucleotides. Secondary responses of this magnitude are typical for cross-channel interferences using the MRM detection method (*30*). The oligonucleotide M_r values, 2993 and 4263, corresponding to the strong internal-Ψ signals are assigned as interrelated

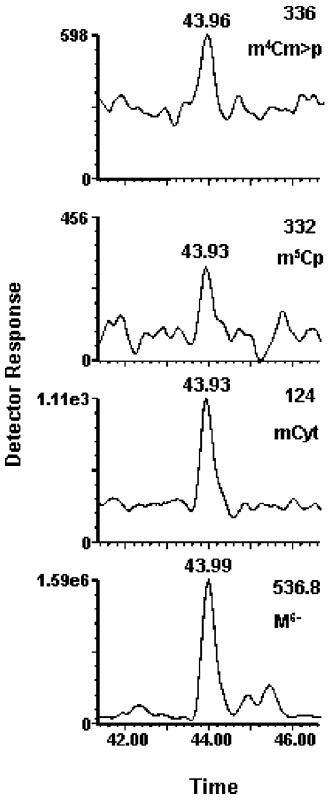

FIGURE 8: Monomer mass ion alignment profiles for the RNase U2 digestion product Cm^{5}CGm^{4}CmCm^{5}CGUm^{5}CA>p (Figure 1B), showing the presence of cytidine methylated only in the base and of dimethylcytidine with one methyl each in the base and ribose.

oligonucleotides from the 3′ terminus of the rRNA molecule, both representing the same Ψ content and placement in the rRNA molecule. On the basis of computer-generated RNase product masses calculated from the gene sequence, both M_r values were found to represent unique base compositions and placements within the rRNA molecule: nucleotides 1535−1544 and 1531−1544, respectively (see Table 1).

Initial placement of Ψ-1540 was accomplished by LC/MS/MS sequencing of both the RNase U2 10mer (Figure 10) and RNase T1 14mer (data not shown). In both cases, the unusually abundant w_4 ion, present at greater than twice the abundance of this ion from the unmodified oligonucleotide (data not shown), was used to place the Ψ at position 1540. The placement was confirmed by the significantly intense complementary a_6 ion in the U2 10mer, although the corresponding a_{10} ion, not surprisingly, was not observed above background in the T1 14mer spectrum. The w_3 ion in Figure 10 (and to a lesser extent the w_2 ion) is also present at twice the normal abundance. When relative abundances of the w series ions from Figure 10 are represented as enhancements (*30*) relative to those in the control mass spectrum of the unmodified 10mer, clear enhancements of w_4 and w_3 ion abundances are observed (see Figure S10). These data are consistent with the presence of Ψ at positions 1540 and 1541, supported by four observations: (1) a census of nucleoside modification, which indicates a net of approximately three Ψ residues in the RNA molecule; (2) the unusually intense MRM responses; (3) the lack of any other MRM response that would indicate placement of a Ψ in an oligonucleotide other than the already characterized Ψ-516;

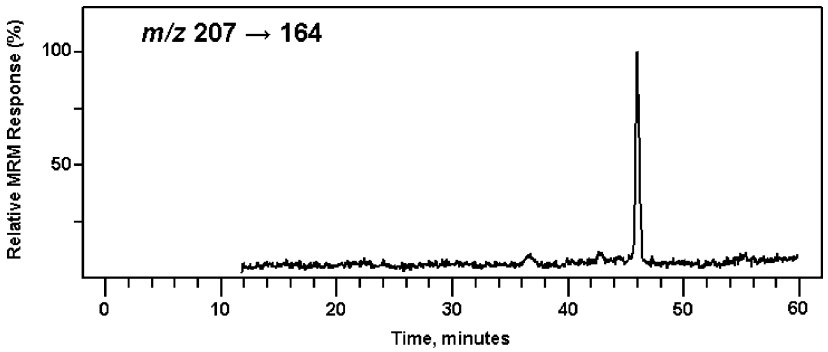

FIGURE 9: MRM response for the *m*/*z* 207 → 164 channel showing the presence of internal Ψ in the RNase T1 oligonucleotide M_r 4263 (44.9 min eluant in Figure 1A).

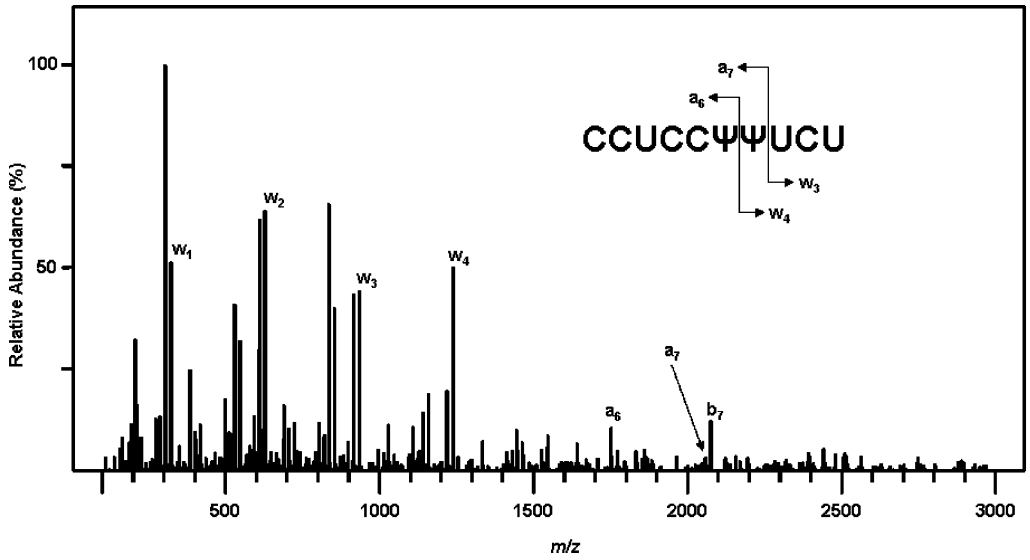

FIGURE 10: Product ion mass spectrum from CID of RNase U2 oligonucleotide M_r 2993 exhibiting anomalously high abundances of ions w_3 (*m*/*z* 934) and w_4 (*m*/*z* 1240) and showing sites of pseudouridylation corresponding to positions 1540 and 1541.

and (4) the above-mentioned enhanced w_4 and w_3 responses in the mass spectrum of the U2 10mer digestion product.

The hypothesis that two Ψs are present in the single-stranded 3′ terminus of the rRNA molecule was tested by preparing the cyanoethyl derivative of the RNase T1 14mer, which eluted in a clearer part of the chromatogram and was thus easier than the 10mer to collect as a clean fraction. The mass spectrum (not shown) of the reaction product indicated the presence of zero (M_r 4262), one (M_r 4315), and two (M_r 4368) cyanoethyl groups in the reaction mixture, in the ratio 0.29:1.0:0.35, of the derivatized 14mer. These results clearly show the presence of two cyanoethyl additions to the molecule, but because of difficulty in driving the reaction to completion, the occurrence of lesser amounts of mono-Ψ components (at either position 1540 or position 1541) cannot be excluded. [Our reaction yield is qualitatively similar to earlier results using this method of Ψ detection, in which 50% reaction yield was reported for one Ψ (*32*).]

DISCUSSION

The posttranscriptional modification map of *T. thermophilus* SSU rRNA allows insight into two areas of principal interest: first, how rRNA modification patterns of a bacterial thermophile differ from those of archaeal thermophiles, particularly with regard to the extent of modification, and the chemical structures of individual modified nucleosides, some of which may contribute to thermal stabilization. This comparison relates to phylogenetic differences (bacteria vs archaea) as well as the extent to which posttranscriptional modification has evolved as a significant contributor to RNA thermostabilization in each phylogenetic domain. Second, by comparison with the *E. coli* SSU rRNA modifications (tabulated at http://medlib.med.utah.edu/RNAmods/), insights may be gained into how typical (i.e., conserved) the modifications in *E. coli* are as a model of bacterial rRNA modifications in general. Until the present work *E. coli* was the only bacterial SSU rRNA to have been completely modification-mapped and has generally been regarded as typical of bacteria, a concept supported on a limited basis by early reports of modifications in RNase T1 oligonucleotide fingerprint catalogs (*11*, *38*). Also, we view the overall modification characteristics of this bacterial thermophile SSU RNA as a likely predictor of the corresponding LSU RNA, not previously reported.

The availability of two complete bacterial SSU modification maps, taken together with various incomplete sequence-modification files (*9*), should now permit the modeling of the most important (conserved) modified residues into existing high-resolution SSU structures of *T. thermophilus* (*39*) and *E. coli* (*40*). Finally, the present study constitutes the most extensive effort to date for use of mass spectrometry for modification mapping of large RNAs, thus providing an

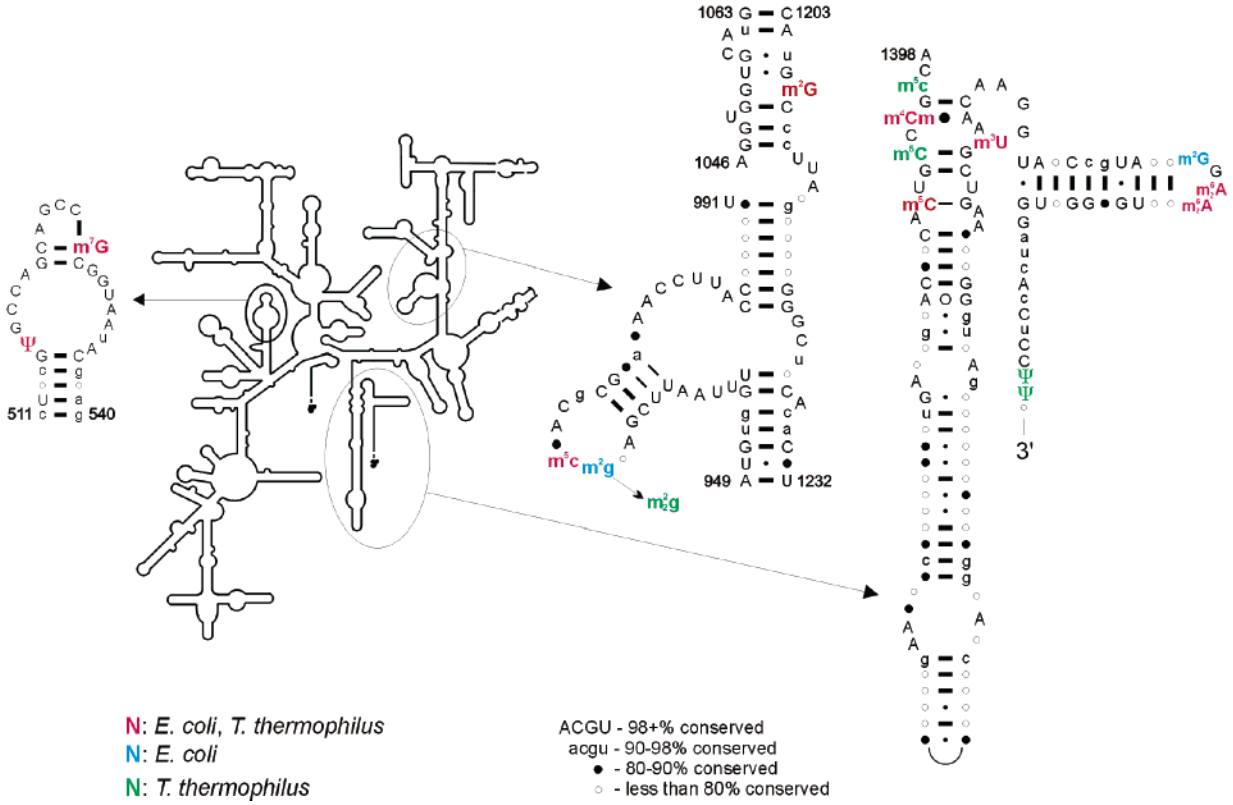

FIGURE 11: Secondary structure representation of 16S rRNA showing sites and identities of posttranscriptional modifications for *T. thermophilus* and *E. coli*. Referenced to the *E. coli* sequence and numbering from http://www.rna.icmb.utexas.edu, courtesy of R. Gutell.

opportunity for critique of the experimental problems and prospects of this methodology.

Results from the present study are summarized in the two-dimensional secondary structure in Figure 11, which includes *E. coli* modification sites for comparison. Although *Thermus* and *Escherichia* are not closely related (*41*) on the basis of comparative SSU rRNA sequences, the most widely used method of phylogenetic placement (*11*), their 16S rRNA modification maps are more similar than we had expected, showing a degree of conservation not only in the modification sites but also in structures of the modified nucleotides. Nine of the 14 modified residues in *T. thermophilus* are identical to those in *E. coli*. [One more residue, m_2^2G-966, differs from *E. coli* only by addition of a methyl group in the base. N^2,N^2-Dimethylguanosine had earlier been reported as a modification at this site in some bacteria but without discussion or further molecular characterization (*42*).]

Compared to 14 modified sites in *Thermus*, *E. coli* 16S rRNA utilizes seven different modifications at 11 sites, which we judge to be typical or slightly high for bacteria in general based on numerous RNase T1 catalogs (*9*), in which oligonucleotide fragments smaller that 6mers are generally not reported. Therefore, to the extent that modification in general supports RNA tertiary structure and thus plays an overall stabilizing role (*43*−*45*), modification in bacterial thermophiles probably plays a modestly increased role in stabilization compared to the mesophile *E. coli* but does not appear to be a major vehicle for structural stabilization. Interestingly, the modification levels of *Thermus* tRNAs similarly do not appear to be particularly high and are, in fact, slightly lower than in *E. coli* tRNA (*4*), when comparing the same isoacceptors.

By sharp contrast, SSU RNA of the archaeal thermophile *S. solfataricus*, which grows in the same temperature range as *T. thermophilus* (70−75 °C), was reported to contain a net of approximately 38 modified sites (*14*), the largest number reported for a prokaryotic SSU RNA. Even so, the *Sulfolobus* level is significantly below that of the archaeon *Pyrococcus furiosus* (culture temperature 100 °C), which has ~80 modified residues (J. A. McCloskey, P. F. Crain, and K. O. Stetter, unpublished results). *Sulfolobus* 16S RNA (75 °C culture) contains 24 2′-O-methylated nucleotides, providing an effective form of RNA thermostabilization resulting from thermodynamic preference for the C3′-*endo* sugar conformation in order to avoid base−O-2′ steric interactions in the alternate C2′-*endo* conformation (*46*−*48*). As a notable difference *T. thermophilus* contains only a single ribose-methylated residue, m^{4}Cm-1402, which is common to *E. coli* and *P. vulgaris* (*9*, *49*) and probably other bacteria.

Modification at position 1401 (m^{5}C) is presently unique to all reported bacterial sequences, and two others (Ψ 1540, 1541; discussed below) are unique to all SSU sequences (*9*). Position 1498 is usually modified in bacterial sequences but was structurally characterized (as m^{3}U) only in the case of *E. coli* (*9*). U-1498 is tabulated as modified, but with unknown structure, in the single reported modified RNase T1 oligonucleotide from *Thermus* (*50*). m_2^2G-1207 and Ψ-516 are the same as in *E. coli*, but interestingly, they are not reported modification sites in any other SSU RNA of all three evolutionary domains (*9*).

The *Thermus* modification map reveals two unusual areas of modification not previously observed. The remarkable sequence 1399-Cm^{5}CGm^{4}CmCm^{5}CGUm^{5}C-1407 (see Figures 6 and 11) occurs in the decoding center of the ribosome

at the head of SSU RNA helix 44 (*39*) and contains an unusual concentration of five methyl groups including three m^5C residues, thus far unique to all SSU RNAs (*9*). Although the general role of m^5C is understudied in proportion to its relatively common occurrence in RNA, it is known to provide a stabilizing influence on structure (*48*), which would be a reasonable occurrence in the functional center of the SSU RNA (*51*) of a thermophile. While evidence from studies of tRNA indicates that stabilization afforded by m^5C can result from Mg^{2+} binding (*52*), earlier work concluded that the principal effect results from enhanced base stacking interactions derived from 5-methyl substitution (*48*, *53*, *54*). The occurrence of ribose methylation would also offer some stabilization from residue m^4Cm-1402, but the influence of the unusual exocyclic N^4-methyl moiety is not known.

The second particularly notable site of modification in the *T. thermophilus* 16S RNA molecule is the unexpected occurrence of tandem pseudouridine residues in the single-stranded 3′ tail of the molecule, at positions 1540 and 1541 within the (approximate) expected site of the anti-Shine−Dalgarno mRNA-binding sequence. The crystal structure of the *T. thermophilus* 30S ribosomal subunit refined to 3 Å resolution has been reported and terminates at residue 1534 in the 3′ tail (*39*). The occurrence of adjacent Ψ residues has been reported in numerous tRNAs, mainly in anticodon stems and in the T loop (*4*). In rRNA, however, the occurrence of ΨΨ is rare; a report of tandem Ψs in 23S rRNA of the archaeon *Haloarcula marismortui* (*55*) was not supported by later studies (*22*, *56*), and concerns were raised of false positives in the earlier report because of a methodological error (*56*). Otherwise, modification in the 3′ tail of SSU RNA is unusual and is limited to several archaea in which modified Cs of unknown structure further upstream were reported (*57*).

Pseudouridine is a common component of eukaryotic rRNA (*5*) although there is only one residue in *E. coli* 16S rRNA (*6*), at the same site as found in the present study (position 516, in loop 18). Despite considerable interest in the occurrence and function of Ψ (*12*, *13*, *55*), until the present work, *E. coli* and *Bacillus subtilis* are the only bacterial SSU RNAs in which Ψ has been placed (both at position 516) (*58*). The reason for this may be in part due to methodology: position 516 in the bacterial rRNA occurs in the highly conserved sequence 515-GUG-517 (*9*) which would not have been included in the numerous 16S RNase T1-based catalogs published in the earlier literature for phylogenetic placement of bacteria (*11*).

No single function for Ψ in RNA has emerged, and it is generally thought that it is capable of playing a number of structure context-dependent roles (*12*, *59*) as a consequence of the availability of N-1H (not available in uridine) for H-bonding interactions. It is agreed that Ψ serves a role in RNA stabilization in both single and double strands (*48*), principally through enhanced base stacking or imino proton−H_2O coordinations (*60*). It is reasonable/tempting to assume in the present case that Ψ-1540, −1541 in the anti-Shine−Dalgarno sequence of *T. thermophilus* functions to stabilize rRNA−mRNA binding in the thermophilic environment. This prediction has analogy in the stabilizing influence of Ψ−A base pairs at stem−loop junctions (*48*), first established for Ψ-39 at the base of the anticodon stem in tRNA (*60*, *61*). Perhaps more relevant are studies of the stabilizing effect of Ψ in models of anticodon−codon interactions in which replacement of U by Ψ in the second anticodon position of a $tRNA^{Tyr}$ model (34-G-U/Ψ-A-36) resulted in a remarkable T_m increase of 5.3 °C (*62*). However, this postulate regarding the influence of tandem Ψs in *Thermus* is presently without evidence and must be tested experimentally.

Comments on Experimental Methodology. The correct identification and placement of posttranscriptional modifications in large rRNAs are experimentally difficult enterprises, no matter what methods are brought to bear. Mass spectrometry provides unique capabilities for such measurements, based on the fact that all 31 known rRNA modifications with the exception of pseudouridine are expressed as distinct changes in mass compared with the parent nucleotide. The most useful approach involves comparison of experimental molecular masses of RNase-derived oligonucleotides with those predicted by established gene or RNA sequences (*34*). Both ESI- and MALDI-based methods, with mass measurement errors below 0.01%, are basically applicable to such measurements, and much of the experimental difficulty and effort falls to sample preparation prior to mass spectrometry. For example, large RNAs can be enzymatically subsectioned and re-isolated before mass spectrometry in order to reduce the complexity of the nucleotide population to be measured (e.g., refs *16*, *20*, and *22*). While such efforts can reduce ambiguities in placement of modified oligonucleotides in the overall rRNA sequence (placement based on either oligonucleotide mass or sequence), sample consumption, and re-isolation time, can be important when extensive or complete rRNA sequence coverage is desired.

We have elected to approach these mapping difficulties through use of negative ion LC/ESI-MS (or LC/ESI-MS/MS), which permits measurements to be made directly on complex oligonucleotide mixtures, for example, containing several hundred oligonucleotides (*17*). The LC/MS approach permits simultaneous use of methods designed for detection of specific RNA modifications, such as ribose methylation (*63*) or pseudouridylation (*30*) by time alignment of appropriate MS signals with those associated with HPLC separation. Equally important, the ∼20 s passage time of HPLC eluants (whether or not chromatographically resolved) through the electrospray ion source is sufficient, under the conditions we describe, for acquisition of oligonucleotide MS/MS sequence data (*28*) from preselected targets of interest.

We judge that the limit of complexity allowed by LC/MS methods will be reached in the region of RNase digests of RNAs containing 1800−2200 nucleotides, due to two factors: first, ambiguities of possible placement of individual oligonucleotides in the overall RNA sequence, and second, increasing difficulty in making unambiguous time alignments between *m/z* channels. In the latter event, initial subsectioning of the RNA will be required.

ACKNOWLEDGMENT

We are grateful to V. Ramakrishnan for preparations of purified *T. thermophilus* 30S ribosomal subunits and to D. R. Davis for helpful discussions. J. Rozenski generously provided the ONx software used for sequence data analysis.

SUPPORTING INFORMATION AVAILABLE

Table S1, assignments for monomer mass fragment ions used for detection of modified residues in RNase digest oligonucleotides; Figure S1, census of modified nucleosides in *T. thermophilus* 16S rRNA by LC/MS analysis of a total nucleoside digest; Figure S2, mass spectrum from CID of 525-CCm^{7}GCGp; Figure S3, mass spectrum from CID of 964-AAm$_2^2$G>p; Figure S4, chromatographic time alignments between characteristic low-mass ions and molecular ions in RNase T1 and U2 digests; Figure S5, chromatographic time alignments between methylguanine fragment ions and the molecular ion from U2 product M_r 2866; Figure S6, mass spectrum from CID of RNase T1 oligonucleotide M_r 1320; Figure S7, mass spectrum from CID of RNase T1 oligonucleotide M_r 1622; Figure S8, mass spectrum from CID of RNase T1 oligonucleotide M_r 2304; Figure S9, mass spectrum following CID of RNase U2 oligonucleotide M_r 2704; and Figure S10, normalized sum of the w series of sequence ions from the mass spectrum of RNase U2 oligonucleotide M_r 2993, compared with unmodified control oligonucleotide. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Grosjean, H., and Benne, R., Eds. (1998) *Modification and Editing of RNA*, pp 596, ASM Press, Washington, DC.
2. Agris, P. F. (2004) Decoding the genome: a modified view, *Nucleic Acids Res. 32*, 223−238.
3. Rozenski, J., Crain, P. F., and McCloskey, J. A. (1999) The RNA modification database−1999 update, *Nucleic Acids Res. 27*, 196−197.
4. Sprinzl, M., and Vassilenko, K. S. (2005) Compilation of tRNA sequences and sequences of tRNA genes, *Nucleic Acids Res. 33*, D139−D140.
5. Maden, B. E. H. (1990) The numerous modified nucleosides in eukaryotic ribosomal RNA, *Prog. Nucleic Acid Res. Mol. Biol. 39*, 241−303.
6. Bakin, A., Kowalak, J. A., McCloskey, J. A., and Ofengand, J. (1994) A single pseudouridine residue in *E. coli* 16S RNA is located at position 516, *Nucleic Acids Res. 22*, 3681−3684.
7. Gupta, R., Lanter, J. M., and Woese, C. R. (1983) Sequence of the 16S ribosomal RNA from *Halobacterium volcanii*, an archaebacterium, *Science 221*, 656−659.
8. Kowalak, J., Bruenger, E., Crain, P. F., and McCloskey, J. A. (2000) Identities and phylogenetic comparisons of posttranscriptional modifications in 16S ribosomal RNA from *Haloferax volcanii*, *J. Biol. Chem. 275*, 24484−24489.
9. Rozenski, J., and McCloskey, J. A. (2005) The small subunit rRNA modification database, *Nucleic Acids Res. 33*, D135−D138.
10. Balch, W. E., Fox, G. E., Magrum, L. J., Woese, C. R., and Wolfe, R. S. (1979) Methanogens: reevaluation of a unique biological group, *Microbiol. Rev. 43*, 260−296.
11. Fox, G. E., Pechman, K. R., and Noese, C. R. (1977) Comparative cataloging of 16S ribosomal RNA: molecular approach to procaryotic systematics, *Int. J. Syst. Bacteriol. 27*, 44−57.
12. Ofengand, J., and Fournier, M. J. (1998) The pseudouridine residues of rRNA: number, location, biosynthesis, and function, in *Modification and Editing of RNA* (Grosjean, H., and Benne, R., Eds.) pp 229−253, ASM Press, Washington, DC.
13. Charette, M., and Gray, M. W. (2000) Pseudouridine in RNA: what, where, how, and why, *IUBMB Life 49*, 341−351.
14. Noon, K. R., Bruenger, E., and McCloskey, J. A. (1998) Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile *Sulfolobus solfataricus*, *J. Bacteriol. 180*, 2883−2888.
15. Oshima, T., and Imahori (1974) Description of *Thermus thermophilus* (Yoshida and Oshima) comb. nob., a nonsporulating thermophilic bacterium from a Japanese thermal spa, *Int. J. Syst. Bacteriol. 24*, 102−112.
16. Kowalak, J. A., Bruenger, E., and McCloskey, J. A. (1995) Posttranscriptional modification of the central loop of domain V in *E. coli* 23S ribosomal RNA, *J. Biol. Chem. 270*, 17758−17764.
17. Crain, P. F., Ruffner, D. E., Ho, Y., Qiu, F., Rozenski, J., and McCloskey, J. A. (1999) Problems and prospects in the characterization of posttranscriptional modifications in large RNAs, 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.
18. Kowalak, J. A., Bruenger, E., Crain, P. F., and McCloskey, J. A. (2000) Identities and phylogenetic comparisons of posttranscriptional modifications in 16S ribosomal RNA from *Haloferax volcanii*, *J. Biol. Chem. 275*, 24484−24489.
19. Hansen, L. H., Kirpekar, F., and Douthwaite, S. (2001) Recognition of nucleotide G745 in 23S ribosomal RNA by the rrmA methyltransferase, *J. Mol. Biol. 310*, 1001−1010.
20. Hansen, M. A., Kirpekar, F., Ritterbusch, W., and Vester, B. (2002) Posttranscriptional modifications in the A-loop of 23S rRNAs from selected archaea and eubacteria, *RNA 8*, 202−213.
21. Andersen, T. E., Porse, B. T., and Kirpekar, F. (2004) A novel partial modification at C2501 in *Escherichia coli* 23S ribosomal RNA, *RNA 10*, 907−913.
22. Kirpekar, F., Hansen, L. H., Rasmussen, A., Poehlsgaard, J., and Vester, B. (2005) The archaeon *Haloarcula marismortui* has few modifications in the central parts of its 23S ribosomal RNA, *J. Mol. Biol. 348*, 563−573.
23. Chomczynski, P., and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction, *Anal. Biochem. 162*, 156−159.
24. Crain, P. F. (1990) Preparation and enzymatic hydrolysis of RNA and DNA for mass spectrometry, *Methods Enzymol. 193*, 782−790.
25. Pomerantz, S. C., and McCloskey, J. A. (1990) Analysis of RNA hydrolyzates by LC/MS, *Methods Enzymol. 193*, 796−824.
26. Noon, K. R., Guymon, R., Crain, P. F., McCloskey, J. A., Thomm, M., Lim, M., and Cavicchioli, R. (2003) Influence of temperature on tRNA modification in Archaea: *Methanococcoides burtonii* (T_{opt} 23 °C) and *Stetteria hydrogenophila* (T_{opt} 95 °C), *J. Bacteriol. 185*, 5483−5490.
27. Apffel, A., Chakel, J. A., Fischer, S., Lichtenwalter, K., and Hancock, W. S. (1997) Analysis of oligonucleotides by HPLC-electrospray ionization mass spectrometry, *Anal. Chem. 69*, 1320−1325.
28. Ni, J., Pomerantz, S. C., Rozenski, J., Zhang, Y., and McCloskey, J. A. (1996) Interpretation of oligonucleotide mass spectra for determination of sequence using electrospray ionization and tandem mass spectrometry, *Anal. Chem. 68*, 1989−1999.
29. Rozenski, J., and McCloskey, J. A. (2002) SOS: a simple interactive program for ab initio oligonucleotide sequencing by mass spectrometry, *J. Am. Soc. Mass Spectrom. 13*, 200−203.
30. Pomerantz, S. C., and McCloskey, J. A. (2005) Detection of the common RNA nucleoside pseudouridine in mixtures of oligonucleotides by mass spectrometry, *Anal. Chem. 77*, 4687−4697.
31. McLuckey, S. A., Van Berkel, G. J., and Glish, G. L. (1992) Tandem mass spectrometry of small, multiply charged oligonucleotides, *J. Am. Soc. Mass Spectrom. 3*, 60−70.
32. Mengel-Jørgensen, J., and Kirpekar, F. (2002) Detection of pseudouridine and other modifications in tRNA by cyanoethylation and MALDI mass spectrometry, *Nucleic Acids Res. 30*, e135.
33. Pomerantz, S. C., Kowalak, J. A., and McCloskey, J. A. (1993) Determination of oligonucleotide composition from mass spectrometrically measured molecular weight, *J. Am. Soc. Mass Spectrom. 4*, 204−209.
34. Kowalak, J. A., Pomerantz, S. C., Crain, P. F., and McCloskey, J. A. (1993) A novel method for the determination of posttranscriptional modification in RNA by mass spectrometry, *Nucleic Acids Res. 21*, 4577−4585.
35. Limbach, P. A., Crain, P. F., and McCloskey, J. A. (1994) Summary: the modified nucleosides of RNA, *Nucleic Acids Res. 22*, 2183−2196.
36. van Knippenberg, P. H., Van Kimmenade, J. M., and Heus, H. A. (1984) Phylogeny of the conserved 3′ terminal structure of the RNA of small ribosomal subunits, *Nucleic Acids Res. 12*, 2595−2604.
37. van Knippenberg, P. H. (1985) Structural and functional aspects of the N^6,N^6-dimethyladenosine in 16S ribosomal RNA, in *Structure, Function, and Genetics of Ribosomes* (Hardesty, B., and Kramer, G., Eds.) pp 412−424, Springer-Verlag, New York.

38. Woese, C. R. (1987) Bacterial evolution, *Microbiol. Rev. 51*, 221−271.
39. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit, *Nature 407*, 327−339.
40. Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., and Cate, J. H. (2005) Structures of the bacterial ribosome at 3.5 Å resolution, *Science 310*, 827−834.
41. Pace, N. R. (1997) A molecular view of microbial diversity and the biosphere, *Science 276*, 734−740.
42. Woese, C. R., Gutell, R., Gupta, R., and Noller, H. F. (1983) Detailed analysis of the higher-order structure of 16S-like ribonucleic acids, *Microbiol. Rev. 47*, 621−669.
43. Sampson, J. R., and Uhlenbeck, O. C. (1988) Biochemical and physical characterization of an unmodified yeast phenylalanine transfer RNA transcribed *in vitro*, *Proc. Natl. Acad. Sci. U.S.A. 85*, 1033−1037.
44. Derrick, W. B., and Horowitz, J. (1993) Probing structural differences between native and *in vitro* transcribed *Escherichia coli* valine transfer RNA: evidence for stable base modification-dependent conformers, *Nucleic Acids Res. 21*, 4948−4953.
45. Kowalak, J. A., Dalluge, J. J., McCloskey, J. A., and Stetter, K. O. (1994) Role of posttranscriptional modification in stabilization of transfer RNA from hyperthermophiles, *Biochemistry 33*, 7869−7876.
46. Kawai, G., Ue, H., Yasuda, M., Sakamoto, K., Hashizume, T., McCloskey, J. A., Miyazawa, T., and Yokoyama, S. (1991) Relation between functions and conformational characteristics of modified nucleosides found in tRNAs, *Nucleic Acids Symp. Ser.*, 49−50.
47. Kawai, G., Yamamoto, Y., Kamimura, T., Masegi, T., Sekine, M., Hata, T., Iimori, T., Watanabe, T., Miyazawa, T., and Yokoyama, S. (1992) Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2′-hydroxyl group, *Biochemistry 31*, 1040−1046.
48. Davis, D. R. (1998) Biophysical and conformational properties of modified nucleosides in RNA (nuclear magnetic resonance studies), in *Modification and Editing of RNA* (Grosjean, H., and Benne, R., Eds.) pp 85−102, ASM Press, Washington, DC.
49. Carbon, P., Ebel, J. P., and Ehresmann, C. (1981) The sequence of the ribosomal 16S RNA from *Proteus vulgaris*. Sequence comparison with *E. coli* 16S RNA and its use in secondary structure modeling, *Nucleic Acids Res. 9*, 2325−2333.
50. Hensel, R., Demharter, W., Kandler, O., Kroppenstedt, R. M., and Stackebrandt, E. (1986) Chemotaxonomic and molecular-genetic studies of the genus *Thermus*. Evidence for a phylogenetic relationship of *Thermus aquaticus* and *Thermus ruber* to the genus *Deinococcus*, *Int. J. Syst. Bacteriol. 36*, 444−453.
51. Brimacombe, R., Mitchell, P., Osswald, M., Stade, K., and Bochkariov, D. (1993) Clustering of modified nucleotides at the functional center of the ribosome, *FASEB J. 7*, 161−167.
52. Chen, Y., Sierzputowska-Gracz, H., Guenther, R., Everett, K., and Agris, P. F. (1993) 5-Methylcytidine is required for cooperative binding of Mg^{2+} and a conformational transition at the anticodon stem-loop of yeast phenylalanine tRNA, *Biochemistry 32*, 10249−10253.
53. Sowers, L. C., Shaw, B. R., and Sedwick, W. D. (1987) Base stacking and molecular polarizability: effect of a methyl group in the 5-position of pyrimidines, *Biochem. Biophys. Res. Commun. 148*, 790−794.
54. Wang, S., and Kool, E. T. (1995) Origins of the large differences in stability of DNA and RNA helices: C-5 methyl and 2′-hydroxyl effects, *Biochemistry 34*, 4125−4132.
55. Ofengand, J. (2002) Ribosomal RNA pseudouridines and pseudouridine synthases, *FEBS Lett. 514*, 17−25.
56. Del Campo, M., Recinos, C., Yanez, G., Pomerantz, S. C., Guymon, R., Crain, P. F., McCloskey, J. A., and Ofengand, J. (2005) Number, position, and significance of the pseudouridines in the large subunit ribosomal RNA of *Haloarcula marismortui* and *Deinococcus radiodurans*, *RNA 11*, 210−219.
57. Woese, C. R., Gupta, R., Hahn, C. M., Zillig, W., and Tu, J. (1984) The phylogenetic relationships of three sulfur dependent archaebacteria, *Syst. Appl. Microbiol. 5*, 97−105.
58. Wrzesinski, J., Bakin, A., Nurse, K., Lane, B. G., and Ofengand, J. (1995) Purification, cloning, and properties of the 16S RNA pseudouridine 516 synthase from *Escherichia coli*, *Biochemistry 34*, 8904−8913.
59. Meroueh, M., Grohar, P. J., Qiu, J., SantaLucia, J., J., Scaringe, S. A., and Chow, C. S. (2000) Unique structural and stabilizing roles for the individual pseudouridine residues in the 1920 region of *Escherichia coli* 23S rRNA, *Nucleic Acids Res. 28*, 2075−2083.
60. Davis, D. R. (1995) Stabilization of RNA stacking by pseudouridine, *Nucleic Acids Res. 23*, 5020−5026.
61. Davis, D. R., and Poulter, C. D. (1991) ^{1}H-^{15}N NMR studies of *Escherichia coli* tRNA(Phe) from hisT mutants: a structural role for pseudouridine, *Biochemistry 30*, 4223−4231.
62. Davis, D. R., Veltri, C. A., and Nielsen, L. (1998) An RNA model system for investigation of pseudouridine stabilization of the codon-anticodon interaction in $tRNA^{Lys}$, $tRNA^{His}$ and $tRNA^{Tyr}$, *J. Biomol. Struct. Dyn. 15*, 1121−1132.
63. Qiu, F., and McCloskey, J. A. (1999) Selective detection of ribose-methylated nucleotides in RNA by a mass spectrometry based method, *Nucleic Acids Res. 27*, e20.

BI052579P