

## Loss of a Universal tRNA Feature<sup>∇</sup>

Chunxia Wang, Bruno W. Sobral, and Kelly P. Williams\*

Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

Received 2 August 2006/Accepted 5 December 2006

**tRNA<sup>His</sup> has thus far always been found with one of the most distinctive of tRNA features, an extra 5' nucleotide that is usually a guanylate. tRNA<sup>His</sup> genes in a disjoint alphaproteobacterial group comprising the Rhizobiales, Rhodobacterales, Caulobacterales, Parvularculales, and Pelagibacter generally fail to encode this extra guanylate, unlike those of other alphaproteobacteria and bacteria in general. Rather than adding an extra 5' guanylate posttranscriptionally as eukaryotes do, evidence is presented here that two of these species, *Sinorhizobium meliloti* and *Caulobacter crescentus*, simply lack any extra nucleotide on tRNA<sup>His</sup>. This loss correlates with changes at the 3' end sequence of tRNA<sup>His</sup> and at many sites in histidyl-tRNA synthetase that might be expected to affect tRNA<sup>His</sup> recognition, in the flipping loop, the insertion domain, the anticodon-binding domain, and the motif 2 loop. The altered tRNA charging system may have affected other tRNA charging systems in these bacteria; for example, a site in tRNA<sup>Gln</sup> sequences was found to covary with tRNA<sup>His</sup> among alphaproteobacteria.**

Through the unique properties of its side chain, the histidine residue makes essential contributions to protein structure and in catalytic mechanisms and must be reliably incorporated as encoded during translation. These same unique properties could make histidine especially disruptive to proteins if accidentally incorporated in place of other amino acids. Translational accuracy depends on tRNA selection in the ribosome, yet prior tRNA recognition and proper charging by the cognate aminoacyl-tRNA synthetase (aaRS) is equally important. This may explain why tRNA<sup>His</sup> has a uniquely distinguishing feature, the presence of an extra nucleotide at the 5' end (the –1 position). Bacterial, archaeal, and eukaryotic cytoplasmic and organellar tRNAs have always been found with an extra 5' guanylate residue, except for reports that the extra nucleotide is an adenylate in snail mitochondria and a uridylate in the tRNA<sup>His</sup> of bacteriophage T5 (37, 38, 42). The extra nucleotide, particularly its 5'-monophosphate, is accommodated in the histidyl-tRNA synthetase (HisRS), which validates tRNA<sup>His</sup> by charging it with histidine (10, 17, 31). Discrimination of tRNA<sup>His</sup> at its acceptor end is especially important because HisRS is less discriminating of anticodon sequence than many aaRSs (24, 29, 41).

An indication of the value of this universal feature of tRNA<sup>His</sup> comes from the very different ways that bacteria and eukaryotes produce it. In eukaryotes the extra guanylate is not encoded in the tRNA<sup>His</sup> gene but is added, subsequently to standard RNase P cleavage of the pre-tRNA, by the specialized tRNA<sup>His</sup> guanylyltransferase Thg1 (11, 18). In bacteria the extra guanylate is encoded in the tRNA<sup>His</sup> gene and is included in the mature tRNA due to an unusual cleavage reaction by RNase P, which can be considered an adaptation of the bacterial enzyme since yeast RNase P cannot perform it (9, 30).

The importance of the extra nucleotide for charging was demonstrated by analysis of the tRNA<sup>His</sup> guanylyltransferase function in *Saccharomyces cerevisiae* (18, 19). The THG1 gene is essential, and death upon its repression in a conditional strain is accompanied by replacement of His-tRNA<sup>His</sup> with a form that is both shorter by one 5' nucleotide and uncharged. This agrees with in vitro studies of the yeast HisRS that show ~500-fold reduction in charging activity upon removal of G-1 (29, 32). Reports on the importance of the extra nucleotide for charging in *Escherichia coli* differ. In vitro studies show reduction in HisRS activity by a factor of 250 or more upon removal of G-1, mainly due to loss of the unusually positioned 5' phosphate (17, 22). In contrast, an in vivo study concluded that the –1 nucleotide was not a tRNA<sup>His</sup> identity element (40).

In examining the tRNA gene content of bacterial genomes we observed that a group of alphaproteobacteria uniquely fails to encode the extra guanylate in its tRNA<sup>His</sup> gene. While this manuscript was in preparation, this same observation appeared in another publication, whose authors suggested that tRNA<sup>His</sup> in these bacteria might have the extra guanylate added by an enzyme analogous to eukaryotic Thg1 (2). Under the same hypothesis, we examined the 5' end of this tRNA<sup>His</sup> and were surprised to find instead that the otherwise universally conserved extra nucleotide was absent. This result implies that tRNA recognition by the HisRS of this group differs substantially from that by any other HisRS, and we describe striking changes in several key regions of HisRS sequences that are correlated with the unusual tRNA gene. Moreover, the tRNA<sup>His</sup> would appear to have become less distinctive among the set of tRNAs in the cell, suggesting that these bacteria may compensate for the lost identity element or simply suffer from a decreased ability to discriminate tRNA<sup>His</sup>.

### MATERIALS AND METHODS

**Bioinformatic analysis.** The script tFind.pl was written to manage the tRNA-finding programs tRNAscan-SE and Aragorn (26, 27) and improve their output and was applied to the complete or nearly complete RefSeq genome sequences from NCBI for 41 RRCPP bacteria, 25 other alphaproteobacteria, and two outgroup strains (*Escherichia coli* K-12 and *Geobacter sulfurreducens*). Of the

\* Corresponding author. Mailing address: Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg VA 24061. Phone: (540) 231-7121. Fax: (540) 231-2606. E-mail: kellwill@vt.edu.

<sup>∇</sup> Published ahead of print on 15 December 2006.

3,181 output tRNA sequences, those with undetermined identity, with introns, flagged as possible pseudogenes (except those called as tRNA<sup>Tyr</sup>, since in all of these cases, they were the only tRNA<sup>Tyr</sup> species in the genome and defects were not severe) and the small set of selenocysteinyl tRNAs, 92 in all, were rejected. Removing redundant identical copies from each species left 2,818 sequences, which were aligned in isoacceptor groups, and maximum-likelihood trees were generated by using Phylml. Inspection of sequences for outliers on the trees led to the rejection of eight very unlikely tRNA sequences.

HisRS genes and additional tRNA<sup>His</sup> genes were found by querying the National Center for Biotechnology Information for "tRNA-His," "HisRS," and "histidyl" (rejecting those not annotated as HisRS); additional eukaryotic sequences were collected by using the BLAST server for eukaryotic genome projects. The results included HisRS for all 303 complete prokaryotic genomes available at the National Center for Biotechnology Information on 12 January 2006, except for those of *Brucella melitensis* biovar Abortus 2308, *Magnetospirillum magneticum* AMB-1, and *Nanoarchaeum equitans* Kin4-M, which were identified by using BLAST. Combining identical sequences, rejecting HisRS paralogs and those not extending through motif 2, and counting multiple instances from the same gene once, there were 521 HisRS sequences (41 from the RRCPP alphaproteobacteria) and 460 tRNA sequences. Motif 2 loop sequences for 1,682 class IIa aaRS (SerRS, ThrRS, ProRS, HisRS, and some GlyRS) were taken from Pfam version 19.0 entry PF00587 after exclusion of incomplete sequences and HisRS paralogs, and the anticodon-binding domain sequences of 1,320 of the same enzymes (less SerRS) were from PF03129 (15). tRNA<sup>His</sup> genes were aligned manually, and HisRS sequences were aligned by using MUSCLE (13), adding other aaRS motif 2 sequences as described previously (12).

**RNA purification.** *Sinorhizobium meliloti* 1021 was cultured at 30°C in TY medium with shaking to an  $A_{600}$  of 0.6, and cells from 1.8 ml of culture were rapidly harvested for total RNA extraction as described previously (3); the yield was 36 µg. Total RNA from *Caulobacter crescentus* C15 was a gift from Yves Brun and Ellen Quardokus (Indiana University). Templates for in vitro transcription of *S. meliloti* tRNA<sup>His</sup> with or without G-1 and with the native U59 or the G59 variation were prepared by mutagenic PCR templated with genomic DNA. tRNAs were transcribed with T7 RNA polymerase, including GMP at a 15-fold excess over GTP to produce 5'-monophosphate ends, and purified after single-nucleotide resolution in a 12% polyacrylamide-8 M urea sequencing gel.

**RNA circularization.** Total bacterial RNA at 80 ng/µl, or a mixture of two in vitro tRNA transcripts (G<sub>-1</sub>/G<sub>59</sub> and X<sub>-1</sub>/U<sub>59</sub>) at 1.5 pM each, was incubated for 60 min at 37°C in 15 µl of RNA ligase buffer (50 mM Tris-HCl [pH 7.8] at 25°C, 10 mM MgCl<sub>2</sub>, 1 mM ATP, 10 mM dithiothreitol) with 1.3 U of T4 RNA ligase (New England Biolabs)/µl and then at 70°C for 5 min. In some experiments, RNA ligation was preceded by dephosphorylation with or without rephosphorylation. Total *S. meliloti* RNA was incubated at 200 ng/µl for 60 min at 37°C in 50 mM Tris-HCl (pH 7.9) at 25°C, 100 mM NaCl, 10 mM MgCl<sub>2</sub>, 1 mM dithiothreitol, with 0.4 U of calf intestinal phosphatase (New England Biolabs)/µl, purifying the product RNA by double phenol-CHCl<sub>3</sub> extraction and ethanol precipitation. For rephosphorylation, dephosphorylated RNA was incubated at 125 ng/µl for 50 min at 37°C in RNA ligase buffer with 0.5 U of T4 polynucleotide kinase (New England Biolabs)/µl, purifying the product RNA as described above.

**Amplification and sequencing of products.** Reverse transcription-PCR (RT-PCR) primers (-R and -F) are indicated in Fig. 2A to C. RT used Superscript II (Invitrogen) according to the manufacturer's directions. PCR used GoTaq green mix (Promega) with 0.8 µM each of the RT and second PCR primers, starting from ice immediately to 3 min at 95°C, and followed by 30 cycles of 45 s at 95°C, 30 s at 50°C, and 30 s at 72°C. PCR products were analyzed in 2.5% agarose-Tris-borate-EDTA gels and in some cases gel-purified or cloned into the pGEM-T Easy vector (Promega) for sequencing by using the BigDye terminator system (Applied Biosystems) with capillary gel electrophoresis.

**Primer extension.** In vitro transcript standards (100 fmol) of *S. meliloti* tRNA<sup>His</sup>, either with or without an extra 5' guanylate, or 20 µg of total *S. meliloti* RNA were hybridized to 5'-<sup>32</sup>P-labeled primer His-AC (Fig. 2B) and subjected to RT with Superscript II according to the manufacturer's directions. Products were extracted with phenol, precipitated with ethanol, and resuspended in formamide dye, and portions were subjected to electrophoresis in a 12% polyacrylamide-8 M urea sequencing gel, which was dried for autoradiography.

## RESULTS

**Unusual tRNA<sup>His</sup> gene in an alphaproteobacterial group.** The tRNA-finding computer programs tRNAscan-SE and Aragorn are excellent tools for locating tRNA and tmRNA

genes in large sequences (26, 27), but they (i) occasionally err on gene endpoints; (ii) do not sort tRNA genes with CAT anticodons into initiator, isoleucine, and elongator methionine classes; and (iii) ignore the -1 position of tRNA<sup>His</sup> genes. We wrote a PERL script tRNAfind.pl that solves these problems for combined tRNAscan-SE and Aragorn output, using kingdom-specific rules for distinguishing genes with the CAT anticodon (28).

Examination of the corrected tRNA<sup>His</sup> gene sequences showed that a group of alphaproteobacteria fails to encode G at -1 as all other bacteria do; in addition, they have C72T and C73A changes relative to other bacterial genes (Fig. 1). The group, here referred to as RRCPP, includes members with known tRNA<sup>His</sup> genes from the orders *Rhizobiales*, *Rhodobacterales*, *Caulobacterales*, *Parvularculales*, and the species *Pelagibacter ubique* that has been placed among the *Rickettsiales*. Although the only tRNA<sup>His</sup> gene known for the order *Parvularculales* has G-1, the presence of T72 and A73 and analysis of its HisRS sequence (below) indicate that it is part of this group. The other alphaproteobacterial orders investigated, *Rhodospirillales* and *Sphingomonadales*, and the remaining *Rickettsiales*, have the canonical bacterial tRNA<sup>His</sup> gene with G-1, C72, and C73.

There is one other bacterial genome project, for *Rhodospirella baltica* of the *Planctomycetes*, with a tRNA<sup>His</sup> gene that does not encode G-1; it has T-1 and A73 but not T72. The function of this gene is questionable, since the same genome also contains a conventional bacterial tRNA<sup>His</sup> gene.

**tRNA<sup>His</sup> that lacks the extra 5' nucleotide.** Three main possibilities considered for the 5' end of tRNA<sup>His</sup> in the RRCPP group were that (i) an uncoded guanylate is added enzymatically after standard RNase P cleavage as in eukaryotes and some mitochondria, (ii) an unusual RNase P cleavage as for other bacterial tRNA<sup>His</sup> leaves whichever extra 5' (nonguanlylate) nucleotide is encoded in the gene, or (iii) standard RNase P cleavage leaves a mature tRNA<sup>His</sup> with no extra nucleotide. Concerning the first possibility, no homolog of the eukaryotic tRNA guanylyltransferase was detected in any of the alphaproteobacteria, although PSI-BLAST did identify a possible homolog in the gram-positive bacteria *Bacillus thuringiensis* and *Nocardia farcinica*.

To investigate these possibilities further, we used T4 RNA ligase to join the ends of molecules in a total RNA preparation and then amplified the ligation junction and flanking sequences of circularized tRNAs by using RT-PCR (8, 42). This method has been used to determine terminal sequences of tRNAs and a tmRNA homolog in mitochondria, including a tRNA<sup>His</sup> with an extra 5' nucleotide (23, 42). The approach has several merits. (i) Individual tRNA species need not be purified. (ii) Contaminating genomic DNA should not interfere, since PCR primers designed to converge across the junction of a circularized RNA would have a divergent orientation on genomic DNA. (iii) If products do not contain flanking sequences found in precursor RNAs, they reflect circularization of mature RNAs. (iv) Even with mispriming, primer-internal sequence from either side of the circle junction can unambiguously identify both of the ligated RNA ends.

We first examined tRNAs from *Sinorhizobium meliloti* of the order *Rhizobiales*. A test amplification for the circularized tRNA<sup>Trp</sup> yielded a ladder of PCR product bands of the sizes



		tDNA		HisRS		SerRS		HisRS Motif 2 Loop	
		5' End		3' End <sup>73</sup>		Flipping Loop		ProRS ThrRS GlyRS	
E	Hsa	CGCCGTGA	TCACGGCA	Hsa	ETLMGKYGEDS---KLI	Hsa	RSEAGSYGRDTRGLIRMH	Hsa	RRDNPAM----TRGRYR
	Dme	AGCCGTGA	TCACGGCA	Dme	EVLTGKYGEDS---KLI	Dme	RDEVRPRF---GVMRSR	Dme	RRDNPAM----TKGRYR
	Cel	GGCCTGCG	CGCAGGCA	Cel	DVLMGKYGEEGG---KLV	Cel	RNEPSGSLH---GLMRVR	Cel	RRDQPVM----SRGRYR
	Scd	TGCCATCT	AGATGGCA	Scd	EILAGKYGEDS---KLI	Scd	RNEISPRS---GLIRVR	Scd	RRDQPAM----TKGRMR
	Ehi	AGCTCACA	TGTGAGCA	Ehi	DILTISKYGEEGG---KLI	Ehi		Ehi	RRDQPYM----TRGRQR
Ath	AGTGGCTG	CAGCCACA	Ath	ETLMGKYGEDS---KLV	Ath		Ath	RRDNP----SKGRYR	
A	Neq	GGCCCGCG	CGGGGGCC	Neq	DTLKGKYGEAAE---KKEI	Neq		Neq	RYDEP----QKGRYR
	Hma	GTCCGGGT	ACCCGGAC	Hma	EMYVDKSGEEIV---EEL	Hma		Hma	RYEEP----QQRFR
	Mth	TGCCGGGG	CCCCGGCC	Mth	KLFTVKSGEVIV---NQI	Mth		Mth	RYERP----QAGRFR
	Sso	GGCTCGGG	CCCCGAGC	Sso	KLFLKGGEEELR---QTM	Sso		Sso	RYDEP----QYGRYR
	Ape	GGCCCGGG	CCCCGGCC	Ape	ETLSGKYGEAAE---SRLI	Ape		Ape	RHEEP----QKGRYR
BII	Pub	TGCTGCGG	CGCGGGTA	Pub	ENIGSFLADDDSNPMSDV	Pub	RNEKA----GNARYR	Pub	RNEKA----GNARYR
	Pbe	GGCCCGCT	AGCGGGTA	Pbe	DAL-GKFLPDRPNQGV	Pbe	RNEKP----GPGFRF	Pbe	RNEKP----GPGFRF
	Ccr	AGCCCGCT	AGCTGGTA	Ccr	DAL-GKFLPDRPNQGV	Ccr	RNEKP----GPGFRF	Ccr	RNEKP----GPGFRF
	Rpa	TGCAGTTG	CAACTGTA	Rpa	DAL-GKFLPDRPNQGV	Rpa	RNEKP----GPGFRF	Rpa	RNEKP----GPGFRF
	Bab	CGCGGTCG	CGACCGTA	Bab	DAL-GKFLPDRPNQGV	Bab	RNEKP----GPGFRF	Bab	RNEKP----GPGFRF
	Sme	TGCGGTTG	CAACCGTA	Sme	DAL-GKFLPDRPNQGV	Sme	RNEKP----GPGFRF	Sme	RNEKP----GPGFRF
	Rsp	TGCCGGGT	AGTCGGTA	Rsp	EAL-GKFLPDRPNQGV	Rsp	RNEKP----GPGFRF	Rsp	RNEKP----GPGFRF
	Cau	GGTGGCTG	CAGCCACC	Cau	ETLEGLKGDDEE---KLI	Cau	RGERP----AARGYR	Cau	RGERP----AARGYR
	Tpa	GGTGGGTT	CACCTACC	Tpa	EVLLRKSSEGDTE---KQM	Tpa	RGEKP----QMGRYR	Tpa	RGEKP----QMGRYR
	Rba	GGTGGTTG	CAGCCACC	Rba	EILTKGSDDET---RQL	Rba	RGEKP----QEGRYR	Rba	RGEKP----QEGRYR
BI	Xax	GGTGGGCT	CAGCCACC	Xax	DVLLTKSGGETE---RQV	Xax	RGERA----QRGRFR	Xax	RGERA----QRGRFR
	Nos	GGCGGGCG	CGTTCGCC	Nos	EVLQAK-GNQGD---NII	Nos	RGERA----KDGRFR	Nos	RGERA----KDGRFR
	BanII	GGCGGTTG	CAGTCGCC	BanII	ELMSYKYGGEDEL---KEI	BanII	RDGPI----KQGRFR	BanII	RDGPI----KQGRFR
	Blo	GGTGGCCA	TGGCCACC	Blo	ASLLKK-GETS---KEI	Blo	RGERP----QEGRFR	Blo	RGERP----QEGRFR
	Pgi	GGTGGGCG	CGTTCACC	Pgi	STLMGKYGEEGD---RLL	Pgi	RADRP----QKGRYR	Pgi	RADRP----QKGRYR
	Cac	GGTGGGTA	TATTCACC	Cac	EVLLAKSAGETE---KQV	Cac	RGERN----QKGRYR	Cac	RGERN----QKGRYR
	Nar	GGCGACCG	CGGTTCGCC	Nar	AVFSRSLGETTDDVVSKEK	Nar	RYERP----QKGRYR	Nar	RYERP----QKGRYR
	Mma	GGCGGGTG	CACTCGCC	Mma	DVFARTLGETSDVVTKEK	Mma	RHERP----QKGRFR	Mma	RHERP----QKGRFR
	Ama	GGCGGGCG	CGTTCGCC	Ama	DVFLTKLGDSSDIIGKEM	Ama	RYERP----QKCRQR	Ama	RYERP----QKCRQR
	Pma	GGCGGGCG	CGTTCGCC	Pma	ELFMRIIGEGTDVVSKEK	Pma	RYERP----QAGRQR	Pma	RYERP----QAGRQR
V	Cab	GGCGAACG	CGTTCGCC	Cab	EVFL-HVGEQSDIVKEM	Cab	RYERQ----QSGRYR	Cab	RYERQ----QSGRYR
	BanI	above	above	BanI	ELFLRGVGDTTDIVQEM	BanI	RYERP----QAGRYR	BanI	RYERP----QAGRYR
	Bsu	GGCGGTTG	CAATCGCC	Bsu	ELFARGVGESTDIVQEM	Bsu	RYERP----QTGRYR	Bsu	RYERP----QTGRYR
	Aae	GGAGGGTG	CACCTACC	Aae	ELFQRSVGETTDDIVQEM	Aae	RYERP----QAGRFR	Aae	RYERP----QAGRFR
	Th	GGTGAAGC	CGTTCACC	Th	QVFEEKVGAATDIVRKEK	Th	RAERP----QKGRYR	Th	RAERP----QKGRYR
	Eco	GGTGGCTA	TAGCCACC	Eco	PLFKRAIGEVTDVVEKEM	Eco	RHERP----QKGRYR	Eco	RHERP----QKGRYR
	T5	TGTGGCTA	TAGCCACC						
	Mimi	AGATCCGT	ACGGGTCA						

HisRS Insertion Domain		HisRS Catalytic Domain		HisRS Anticodon Domain	
E	KLDK--VSWEEVK-----NEM	Hsa	EAVL-----LQTPAQAGEEPLGVGVS	Hsa	YKYRRNG-----EKQLI
	KLDK--SPWADVR-----KEM	Dme	EGVLKGESATV-ASPAKTSQQNGEQANEPATVGSV	Dme	VRTRRGK-----DLGSM
	KLDK--TPWEDVE-----QEM	Cel	EAVA-----PK-ALEGTAVENSEDTAGQPVGVGVS	Cel	LRDRDSM-----RORTA
	KLDK--SPWEAVK-----KEM	Scd	EVVTSASAPPENASELKKKAKSAEDASEFVGVGSI	Scd	LRSVTSR-----EEVDV
	KLDK--VSWEDVE-----KEM	Ehi	EAVM-----VVGNTSVGSI	Ehi	LRVVTR-----EEINV
KLDK--QSFEQVK-----KEM	Ath	EAVC-----IGAEVGS	Ath	LRNVVTR-----DEQTI	
A	KWKD--IGEEGVL-----EKL	Neq	EVMI-----EGFNRL	Neq	VKRLGQFEFA-----DDGELV
	KRAK--VDHDEYL-----DAL	Hma	ECFD-----STGEVSRV	Hma	LKNLITQ-----KQLTI
	K----GDVEELE-----ALL	Mth	EYIV-----PELGAQKQI	Mth	LKLEKQSE-----EKEDOTC
	K----NKVDEAL-----VIL	Sso	EVLH-----PSVQFSI	Sso	FQDRLTR-----ERKKI
	KLDK--IGVEGVR-----RLL	Ape	EVVL-----DEPRIGSV	Ape	LKEMDDG-----EQVSV
BII	KLDKPGFGLRGVEDLLKEERVAS---GAV	Pub	ETNL-----NFKAKNSKGEVDIGSI	Pub	VKDMERN-----SQETL
	KYDR--LGAEGVALLGPGRKDAS---GDF	Pbe	EAEI-----TFPVNKDKGEVQVFGSV	Pbe	VRNLITR-----KQISI
	KLDR--IGVDGVRLLGEGRLDES---GDF	Ccr	EAEI-----LLSTDERGNKVSFSGSI	Ccr	LYSAAKG-----ERVAV
	KLDK--FPITEIEKLLGNRWDGEGEGKDF	Rpa	EVEL-----LLDTKDKGRPIRFQSV	Rpa	LKNL-LGVKGN-----NQLTF
	KLDK--FGPEGVRLLLGKRLDES---GDF	Bab	EAEI-----LFDVNTEDGQKVVFQSV	Bab	VKDLILGAKLSEEEIESNA-EWRAGQPAQVTV
	KLDK--FGPEGVRLLLGEGRKDES---GDF	Sme	EAEI-----QFAVTNEKGEKVVFQSV	Sme	IKDLDAGRMAEGLTDNA-AWKAERPQQTII
	KIDR--LGETGVRALLGSGRKDES---GDF	Rsp	EAEI-----TFEIPDEKGRKRQVFGSV	Rsp	IKDLILGAEIASLEKREDEYLLKQADAPQSC
	KFDK--IGADGVR-----AEL	Cau	EAIT-----PHWPEGSL	Sme	IKDLVEGKRLSAEIEDNV-TWLESRPAQITV
	KRAK--MAPHVLT-----AQL	Tpa	ETFL-----TQLPHIGSV	Sme	IKDLIEGKRLSAGEIEDNV-AWREARVAQVSV
	KLK--IGREKTA-----NEM	Rba	ETTL-----DELPFIGSV	Sme	LKDLVLGAQIAAGASLE--EW-KSRPAQVEV
BI	KIDK--RGADYVR-----DTL	Xax	ETTL-----TDHPQIGSI	Cau	IKDLQRG-----EQRSL
	NLEK--IGEAKVR-----LEL	Nos	ETTL-----IGHEALGSI	Tpa	LRDLSTR-----KDTRL
	KIEK--IGIDGVR-----KDV	BanII	EIFL-----KDGSISSSI	Rba	VKTLATKES-----QDVAY
	KLDK--IGADEVA-----RLL	Blo	ETFL-----DGAASLGS	Xax	VKDLVRE-----QQFDV
	KLDK--IGLDKVN-----DEL	Pgi	EVKA-----LDAEMGSI	Nos	LKDLQSG-----EQVEV
	KYDK--ITEEQFL-----SEL	Cac	ETTL-----TGNEYSYGI	BanII	LRNMKEG-----SEVKV
	YFRA--HK-----AEL	Nar	EFVT-----DRLGAQGTV	Blo	VKNIVTG-----EQVAA
	YLSE--FR-----EAL	Mma	EFTT-----TELGAQGTV	Pgi	LKMLTG-----EQQML
	YFEK--HE-----HAL	Ama	EFKT-----TSLGSQDAV	Cac	IKDMITG-----EOKKL
	WLEV--NK-----NSL	Pma	EITT-----GTLGSQATV		
YFLD--HL-----SSL	Cab	EAIT-----TCRDHSYAL			
HFEP--RI-----GEF	BanI	EIMS-----EAEGFGAITTL			
HFEP--RI-----EEF	Bsu	EIMS-----NAEGFGAITTL			
YLTG--VK-----EEL	Aae	EAVS-----DELGL--TL			
VLSP--HR-----EAL	Th	EVHH-----EEIGAQSAL			
FLEQ--HK-----EKL	Eco	EWVT-----NSLGSQGTV			

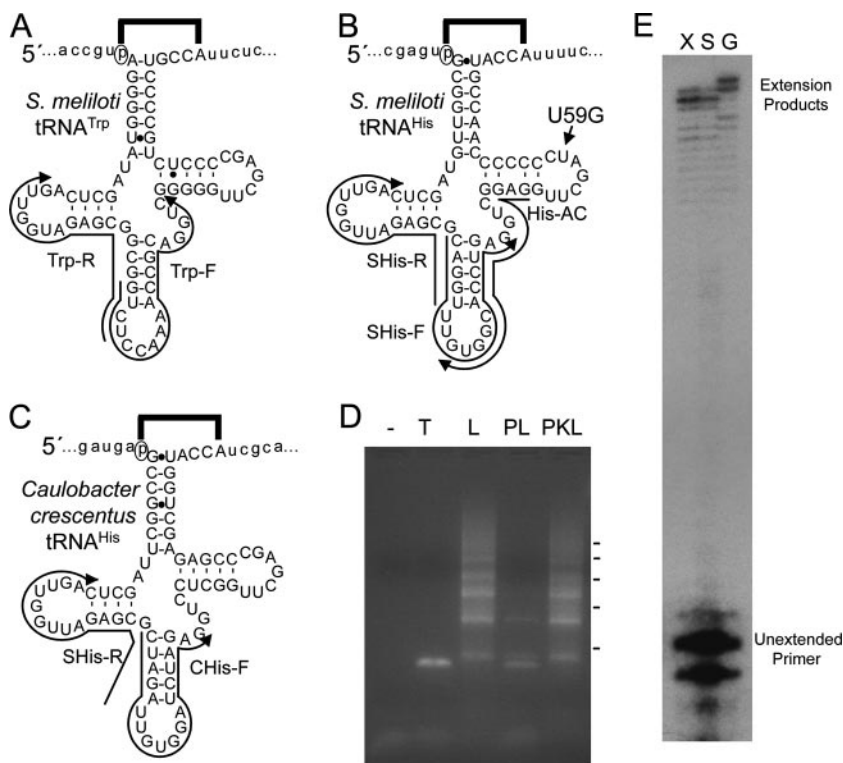


FIG. 2. tRNA end sequence analysis. A-C. RNA ligation products. Joints formed by T4 RNA ligase (brackets), involving the circled 5'-monophosphate ends, as revealed by RT-PCR with the indicated primers. Sequences of precursor RNA sequences are shown with portions removed by tRNA processing in lowercase. (D) Amplification of tRNA<sup>His</sup> circles. *S. meliloti* total RNA was treated with T4 RNA ligase (L), calf intestinal phosphatase and then RNA ligase (PL), or phosphatase and then T4 polynucleotide kinase then RNA ligase (PKL). These samples, total RNA with none of these treatments (T), and no RNA (-) were subjected to RT with primer SHis-R and then PCR with SHis-R and SHis-F. Dashes on the right mark the positions of the lowest five bands of a 100-bp ladder after electrophoresis in an agarose gel. E. Primer extension. The 5'-<sup>32</sup>P-labeled primer His-AC was extended on 50 fmol in vitro transcript *S. meliloti* tRNA<sup>His</sup> with (G) or without (X) the extra 5' guanylate or on 10 µg of total *S. meliloti* RNA (S). Portions of the reaction products (7% for G and X and 40% for S) were analyzed by electrophoresis in a denaturing gel, followed by autoradiography.

expected for a long rolling-circle product of RT. Ensemble sequencing of all products or of the gel-purified smallest band gave the expected end ligation sequence of tRNA<sup>Trp</sup> (Fig. 2A).

RT-PCR on untreated *S. meliloti* RNA using primers for tRNA<sup>His</sup> circles yielded a small artifactual product (Fig. 2D, lane T). (The artifact of lane T of Fig. 2D was independent of RNA ligation but did appear to depend on RT, even in PCR with annealing temperatures as low as 42°C, which ruled out simple mispriming on any contaminating genomic DNA. It had the terminal sequence 5'-. . .CCCAACCGTACcagtggttagag

cgcaggt [where the sequence from primer SHis-R is in lowercase and the sequence from the 3' terminus of tRNA<sup>His</sup> is underlined], and thus showed effective priming at the penultimate nucleotide of tRNA<sup>His</sup>. A single band of the same size was also observed in no-ligation controls with purified in vitro transcript tRNA.) After treatment with RNA ligase to circularize tRNAs, the smallest RT-PCR product was larger than the artifact and was associated with a ladder of bands as for the tRNA<sup>Trp</sup> circle (Fig. 2D, lane L). The gel-purified smallest and next-smallest amplification products both had the sequence

FIG. 1. Unique tRNA<sup>His</sup> gene and HisRS sequences in an alphaproteobacterial group. Selected segments of a sequence alignment are shown from eukaryotic cytoplasmic (E), archaeal (A), and bacterial type I (BI) and II (BII) HisRS (5), along with portions of tRNA<sup>His</sup> genes from the same strains and from viruses (V). For eukaryotes with multiple tRNA<sup>His</sup> genes, only the most abundant is shown. Coordinates of protein sequences are for *E. coli* HisRS. Abbreviations: Hsa, *Homo sapiens*; Dme, *Drosophila melanogaster*; Cel, *Caenorhabditis elegans*; Sce, *Saccharomyces cerevisiae*; Ehi, *Entamoeba histolytica*; Ath, *Arabidopsis thaliana*; Neq, *Nanoarchaeum equitans*; Hma, *Haloarcula marismortui*; Mth, *Methanobacterium thermoautotrophicum*; Sso, *Sulfolobus solfataricus*; Ape, *Aeropyrum pernix*; Pub, *Pelagibacter ubique*; Pbe, *Parvularcula bermudensis*; Ccr, *Caulobacter crescentus*; Rpa, *Rhodospseudomonas palustris*; Bab, *Brucella abortus*; Sme, *Sinorhizobium meliloti*; Rsp, *Rhodospseudomonas sphaeroides*; Cau, *Chloroflexus aurantiacus*; Tpa, *Treponema pallidum*; Rba, *Rhodopirellula baltica*; Xax, *Xanthomonas axonopodis*; Nos, *Nostoc* sp. strain PCC 7120; BanII and BanI, *Bacillus anthracis*; Blo, *Bifidobacterium longum*; Pgi, *Porphyromonas gingivalis*; Cac, *Clostridium acetobutylicum*; Nar, *Novosphingobium aromaticivorans*; Mma, *Magnetospirillum magneticum*; Ama, *Anaplasma marginale*; Pma, *Prochlorococcus marinus*; Cab, *Chlamydomonas abortus*; Bsu, *Bacillus subtilis*; Aae, *Aquifex aeolicus*; Tth, *Thermus thermophilus*; Eco, *Escherichia coli*; Rba2, a second tRNA gene from Rba; Mimi, Mimivirus. Species of the RRCPP bacteria are grouped in a thick box, surrounded by their two closest HisRS homologs (2), and other alphaproteobacterial species are grouped in a thin box. Alignable portions of SerRS, ProRS, ThrRS (Eco), and GlyRS (Hsa) are shown.

from end ligation of tRNA<sup>His</sup> and showed that no extra nucleotide was present at the 5' end of the circularized tRNA<sup>His</sup> (Fig. 2B). This result was confirmed in both tested clones of the PCR products.

To confirm the correlation of this unusual tRNA structure with the unusual gene, we determined the terminal sequence of tRNA<sup>His</sup> from a different order in the RRCPP group, *Caulobacterales*. As for *S. meliloti*, no extra nucleotide was found on the tRNA<sup>His</sup> of *Caulobacter crescentus* (Fig. 2C) in six of seven clones of the amplified circularization junction. (The remaining clone appeared to originate from a circularized tRNA<sup>His</sup> precursor; its junction sequence contained two additional 3' nucleotides and ten additional 5' nucleotides beyond the mature tRNA sequence).

**Possible terminus other than 5'-monophosphate.** These results were surprising given that all other known tRNA<sup>His</sup> have an extra 5' nucleotide and were subject to an objection. Circularization by T4 RNA ligase requires 5'-monophosphate and 3'-hydroxyl ends. It was possible that the circles we detected were from a minor variant form of tRNA<sup>His</sup> and that the major species has a 5' terminus other than a monophosphate; a precedent comes from a study of chicken mitochondrial tRNA<sup>His</sup>, at least a fraction of which was shown to have a 5'-di- or triphosphate (25). To address this issue, we treated *S. meliloti* RNA first with calf intestinal phosphatase to remove terminal mono-, di-, or triphosphates and produce uniform 5'-hydroxyl ends and then with T4 polynucleotide kinase to produce uniform 5'-monophosphate ends. This was followed by RNA ligation and amplification for tRNA<sup>His</sup> circles as before. Dephosphorylation reduced circularization substantially; the main product after RNA ligation and amplification was the same artifact seen when the RNA ligation step was omitted (Fig. 2D, lane PL; compare to lanes T and L). Rephosphorylation was also effective; the artifact was not apparent among the RT-PCR products and only the circularization ladder seen for the original RNA was detected (Fig. 2D, lane PKL [compare to lane L]). Any tRNA<sup>His</sup> 5' ends that were originally hydroxyl, diphosphate, or triphosphate would have been substantially converted to monophosphate by rephosphorylation, and yet these tRNA circles still lacked any -1 nucleotide, as revealed in all of six clones of the PCR products. Thus, the terminal nucleotide sequence of tRNA<sup>His</sup> in *S. meliloti* is uniform and lacks the extra 5' nucleotide found in all other tRNA<sup>His</sup>. Canonical RNase P processing as for non-tRNA<sup>His</sup> is indicated by the observations that the mature 5' sequence ends at the +1 position, and at least a fraction of the tRNA<sup>His</sup> had the 5' monophosphate expected from RNase P cleavage.

**Circularizability of tRNA containing G<sub>-1</sub>.** Our circularization assay may have failed to detect tRNA with an extra guanylate if that nucleotide specifically prevents circularization, for example, by leaving the 5' and 3' ends in such a configuration that they are unable to approach each other. To examine whether tRNA<sup>His</sup> with an extra 5' guanylate can circularize efficiently, we prepared synthetic (unmodified) versions of *S. meliloti* tRNA<sup>His</sup> either with (G<sub>-1</sub>) or without (X<sub>-1</sub>) the extra 5' guanylate by in vitro transcription. Titration in a circularization assay showed that both tRNAs produced RT-PCR products at concentrations as low as 3 pM (data not shown).

Dilution favors circularization but does not rule out bimolecular ligation. To test whether bimolecular ligation might

TABLE 1. Ligation products from a dilute mixture of G<sub>-1</sub>G<sub>59</sub> and X<sub>-1</sub>U<sub>59</sub> versions (with and without the extra 5' guanylate) of *S. meliloti* tRNA<sup>His</sup> transcripts

Ligation partners (5' partner/3' partner)	No. of clones	
	Perfect tRNA	Imperfect 3' end <sup>a</sup>
G <sub>-1</sub> /G <sub>59</sub> (self)	6	4
X <sub>-1</sub> /U <sub>59</sub> (self)	12	5
G <sub>-1</sub> /U <sub>59</sub> (cross)	0	0
X <sub>-1</sub> /G <sub>59</sub> (cross)	0	0

<sup>a</sup> The 3' partner had a base-substitution or single-nucleotide deletion or insertion involving the CCA tail.

explain the result with G<sub>-1</sub> tRNA<sup>His</sup> and, if not, to better compare circularization efficiencies, the assay was performed on a dilute mixture of G<sub>-1</sub> and X<sub>-1</sub> tRNAs, with one tRNA marked near its 3' end by an innocuous sequence variation. In this way, circularization and self-ligation products could be distinguished from cross-molecule ligation products. The innocuous variation was placed at position 59, which is the most variable position in the T-loop of tRNAs in general and is unoccupied by secondary or tertiary contacts in tRNA crystal structures. *S. meliloti* has U<sub>59</sub>, and among tRNA<sup>His</sup> sequences from 41 RRCPP strains, all four bases are found at this position, with G<sub>59</sub> in 26 of them. Accordingly a G<sub>-1</sub>G<sub>59</sub> version of *S. meliloti* tRNA<sup>His</sup> was prepared and mixed with X<sub>-1</sub>U<sub>59</sub> tRNA<sup>His</sup> at 1.5 pM each, and the mixture was treated with T4 RNA ligase. RT-PCR products were cloned and sequenced to identify ligation partners for individual product molecules (Table 1). Eighteen of the tested clones had ligation junctions with perfect tRNA sequences, and these were exclusively from the two types of self-ligation (G<sub>-1</sub>/G<sub>59</sub> and X<sub>-1</sub>/U<sub>59</sub>); none had the G<sub>-1</sub>/U<sub>59</sub> or X<sub>-1</sub>/G<sub>59</sub> ligation junction that would have indicated cross-ligation. If no circularization occurred in the reaction, and all four bimolecular ligations were equiprobable, this result would be expected with a probability of  $0.5^{18} = 4 \times 10^{-6}$ . Even if the input ratio of molecules was as skewed as the output ratio (2:1) might suggest, the probability that the result was due to bimolecular ligation would still be negligible. Four of these clones were tandem repeats of the same self-ligation junction, further supporting self-ligation (and implying that RT had proceeded twice around a circle). The exclusive detection of self-ligation products indicates that circularization predominated under these dilute conditions. Thus, G<sub>-1</sub> tRNA<sup>His</sup> can circularize, although possibly with a mild (~2-fold) disadvantage in circularization efficiency relative to X<sub>-1</sub> tRNA<sup>His</sup>. Nine additional clones came from tRNAs with imperfect 3' ends; despite tRNA purification in a sequencing gel, these clones showed an extra terminal nucleotide, single-nucleotide deletion, or base substitution in the CCA tail. All nine were self-ligation products, which does not provide information on the relative circularization efficiency of the two (perfect) tRNA types but does support the conclusion that cross-molecule ligation was insubstantial under these conditions. We conclude that the failure to detect G<sub>-1</sub> tRNA<sup>His</sup> in total *S. meliloti* RNA was not because an extra guanylate prevents circularization.

**Primer extension.** Modifications at the 5' end of mature tRNA<sup>His</sup>, possibly containing an extra nucleotide, may have prevented circularization of the mature tRNA, allowing detec-



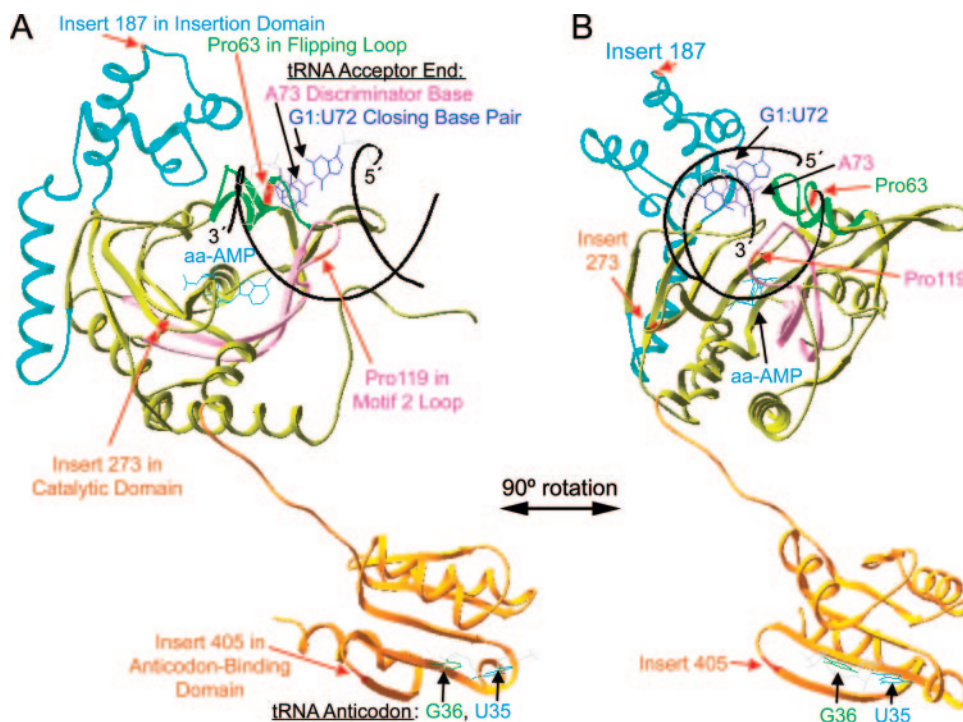


FIG. 3. Map of changes to histidyl-tRNA synthetase. A complex between HisRS and tRNA<sup>His</sup> was modeled from available structures, with five sites (Fig. 1) where RRCPP HisRS differs from canonical bacterial HisRS marked in red. (A) The ribbon is monomer A from a crystal structure (1ADJ) of the *Thermus thermophilus* HisRS homodimer (1). Its catalytic domain backbone was superposed by using the iterative fit function of the Swiss-Pdb viewer 3.7 (20) (552 atoms, RMSD 1.43Å) with that from the *E. coli* AspRS-tRNA<sup>Asp</sup> crystal structure (1COA) to mark the aspartyl-adenylate (aa-AMP) and the acceptor arm (black) of tRNA<sup>Asp</sup> (14). By analogy with AspRS-tRNA<sup>Asp</sup>, the insertion domain here in an open conformation would be expected to close toward the tRNA acceptor end, clamping it to the catalytic domain. The anticodon-binding domain backbone was superposed (320 atoms, RMSD 1.39Å) with that from the *E. coli* ThrRS-tRNA<sup>Thr</sup> crystal structure (1QF6) to mark two anticodon bases of tRNA<sup>Thr</sup> (34). The elbow of tRNA<sup>His</sup> would contact the back of the other HisRS monomer (16). Positions (*E. coli* numbering) of residues and inserts unique to the RRCPP HisRS and bases of RRCPP tRNA<sup>His</sup> are shown. (B) View from the left of panel A.

tion only of immature processing intermediates. To investigate this possibility, the 5' end of tRNA<sup>His</sup> was examined by primer extension, using as standards the *in vitro* transcripts containing or lacking the G-1 nucleotide (Fig. 3). The RT primer used in the previous experiments (Fig. 2B) was designed and demonstrated to perform well in RT-PCR of circularized tRNA but may not have been sufficiently specific in a single round of RT; a new primer was designed to maximize specificity for tRNA<sup>His</sup> by positioning the 3' end in the anticodon sequence. The primer extension products from *S. meliloti* total RNA (lane S) matched those of the standard lacking the extra nucleotide (lane X) and not those of the standard containing an extra 5' guanylate (lane G). This result, which did not depend on the use of T4 RNA ligase, is compatible with that from circularization, again suggesting that mature *S. meliloti* tRNA<sup>His</sup> lacks any extra nucleotide.

**Sequence covariations between a protein and its RNA substrate.** The class II aaRS share sequence motifs and structure in the catalytic domain, and two additional domains are usually present, an insertion domain and an anticodon-binding domain (Fig. 3). A stepwise pathway for tRNA binding has emerged from multiple crystallographic studies of the class II AspRS (7, 35). The tRNA is first recognized by the anticodon-binding domain, which rotates, bending the tRNA elbow and driving the acceptor end toward the catalytic domain. The insertion

domain also rotates, beginning to clamp the acceptor end onto the catalytic domain. When the aminoacyl-adenylate substrate is in the active site, the acceptor end can also be properly accommodated in the active site, with increased clamping by the insertion domain. The acceptor end is further anchored by contacts to two loops within the catalytic domain, the flipping loop and the motif 2 loop, after conformational changes in the loops. Analogs or homologs of all of these critical parts of AspRS can be identified in crystal structures of HisRS (16), and even though there is no structure available for a HisRS complex with tRNA, there are indications that lessons from AspRS apply (6).

HisRS sequences from RRCPP bacteria form a discrete cluster, with major differences from other HisRS in all of the regions that might be expected to be involved in tRNA binding. Attention has previously focused on the motif 2 loop of HisRS, where functionally relevant covariation between HisRS and tRNA<sup>His</sup> sequences has been noted (21). The tRNA<sup>His</sup> discriminator position 73 is a C in bacteria and an A in eukaryotes, while HisRS position 118 (*E. coli* numbering) is a Gln in bacteria and differs (typically found as an Ala-Met-Thr tripeptide) in eukaryotes. Likely contact between Gln118 and C73, as the corresponding position of AspRS makes with the discriminator base of tRNA<sup>Asp</sup>, was demonstrated by directed mutational analysis (21). The RRCPP group shows a different

covariation involving the same tRNA and HisRS positions: the tRNA change from C73 to A correlates with a HisRS change from Gln118 to Gly (Fig. 1). The Gly may not directly recognize A73, but there are additional changes in the motif 2 loop not found in other HisRS that could affect tRNA recognition, Asn114 and strikingly Pro119 (not present in the *Pelagibacter ubique* HisRS, the most divergent of the RRCPP sequences). These changes are unique among HisRS proteins except for a few occurrences of Gly118 in other bacteria but can also be assessed by comparison to the motif 2 loop of other class IIa aaRS. In this broader context, Gly118 is not unusual and is highly conserved among SerRS and ThrRS. Asn114 is likewise not unusual outside of HisRS. Pro119, however, is unique among all class IIb enzymes.

The flipping loop region of the RRCPP also stands out among HisRS; in particular, Pro63 is found uniformly and uniquely in the RRCPP. Unique insertions occur at other sites in the HisRS of RRCPP, in the insertion domain (Fig. 1 and 3, insert 187), in the catalytic domain (insert 273), and in the anticodon-binding domain (insert 405). This latter insert can also be assessed in a broader context because its region is well aligned with ThrRS, ProRS, and GlyRS; it is unique among all class IIa aaRS.

The other alphaproteobacteria, in the *Rhodospirillales*, *Sphingomonadales*, and *Rickettsiales* other than *Pelagibacter*, do not have the unusual HisRS, except that genes for both types of bacterial HisRS are found in data from the incomplete *Magnetospirillum magnetotacticum* genome project. However, the source strain for this project was isolated based on its magnetic properties (4) and may be impure, as is also suggested by the unprecedented finding of two dissimilar tmRNA genes among the same genomic sequence; the complete genome of its congener *M. magneticum* encodes only one of these two HisRS, similar to those of other *Rhodospirillales*.

**Effects on charging systems for other tRNAs.** When initially established in an RRCPP ancestor, the unusual tRNA<sup>His</sup> and HisRS might have affected global tRNA charging accuracy in the cell, with increases in the histidylolation of non-tRNA<sup>His</sup> and in the mischarging of tRNA<sup>His</sup>. This situation might have selected for compensatory changes in other aaRS or other tRNAs or for further compensation in the HisRS or tRNA<sup>His</sup> molecules themselves. As an initial exploration of possible effects on other charging systems, we examined whether any other tRNAs covary with tRNA<sup>His</sup>. All unique tRNA sequences were collected from each of 41 RRCPP strains, 25 non-RRCPP alphaproteobacteria, and two outgroup proteobacteria (*Escherichia coli* K-12 and *Geobacter sulfurreducens*), for each standard isoacceptor family. The sequences were aligned for each family, and each position in each alignment was tested for whether a single base was present in >90% of the RRCPP sequences and a different base was present in >90% of the other bacteria. Only the three previously noted positions in tRNA<sup>His</sup> (-1, 72, and 73) were identified by this test. Reducing the cutoff to 80% added a fourth tRNA<sup>His</sup> site (position 70) and two pairing positions (51 and 65) in the tRNA<sup>Glu</sup> T-stem. Retaining the 90% cutoff, but allowing any set of two or more of the four bases to occur in one bacterial group, with a mutually exclusive set of bases in the other bacterial group, identified one new position, again in tRNA<sup>Glu</sup>, where position 72 is 100% C or G in RRCPP bacteria and

100% A or U in the other bacteria. The pairing partner of position 72 (position 2) in tRNA<sup>Glu</sup> was not detected this way because the non-RRCPP alphaproteobacteria frequently have G<sub>2</sub> paired with U<sub>72</sub>. Thus, a site in tRNA<sup>Glu</sup> covaries perfectly with tRNA<sup>His</sup> among the tested alphaproteobacteria, producing a three-H-bond acceptor stem pair (C<sub>2</sub>:G<sub>72</sub> or G<sub>2</sub>:C<sub>72</sub>) in RRCPP bacteria and a two-H-bond pair (A<sub>2</sub>:U<sub>73</sub>, U<sub>2</sub>:A<sub>73</sub>, or G<sub>2</sub>:U<sub>73</sub>) in the other alphaproteobacteria.

## DISCUSSION

**Loss of a universal tRNA<sup>His</sup> feature.** tRNA<sup>His</sup> in all previously studied organisms has a regular nucleotide at the extra -1 position, usually a guanylate with a 5'-monophosphate. If an extra guanylate or other regular nucleotide were present on the RRCPP tRNA<sup>His</sup>, with either a hydroxyl or mono-, di-, or triphosphate 5' end, our experiments could have detected it. The primary conclusion from our exclusive detection of tRNA<sup>His</sup> species lacking the extra nucleotide is that tRNA<sup>His</sup> does not contain a regular nucleotide at the -1 position in RRCPP bacteria, as is otherwise universal.

Remaining possibilities are that (i) the tRNA species that we detected is the mature tRNA which simply has no extra nucleotide or (ii) an immature tRNA species was detected while the mature tRNA has an irregular -1 nucleotide that prevents the action of either RNA ligase or reverse transcriptase. If the primer extension experiment had demonstrated an extra 5' nucleotide, this would have argued against the first possibility and for the second, but it did not reveal an extra 5' nucleotide and therefore did not resolve the issue. Definitive resolution will probably await direct analysis of the purified tRNA<sup>His</sup> molecule. For now we note that a mature RRCPP tRNA<sup>His</sup> lacking an extra nucleotide can be readily explained by standard RNase P cleavage of the precursor and that there is no precedent for 5' addition of an irregular nucleotide. Thus, the most parsimonious interpretation of the data in hand is that the mature tRNA simply lacks an extra nucleotide.

It was demonstrated here that versions of *S. meliloti* tRNA<sup>His</sup> that contain or lack G<sub>-1</sub> can both be circularized by using T4 RNA ligase. This contrasts with a report that neither of the corresponding versions of an *E. coli* tRNA<sup>His</sup> microhelix could be circularized (36). Key differences are positions 72 and 73, which are U<sub>72</sub>A<sub>73</sub> in *S. meliloti* and C<sub>72</sub>C<sub>73</sub> in *E. coli*.

The anomalous tRNA<sup>His</sup> was found in a group of alphaproteobacteria. Another case where tRNA<sup>His</sup> may prove to lack an extra 5' nucleotide is among bacteriophages. Several phages (with hosts as diverse as gammaproteobacteria, gram-positive bacteria, and cyanobacteria) have a tRNA<sup>His</sup> gene, but all current examples have the unusual termini T-1/G1/C72/C73. A similar mutant of the *E. coli* tRNA<sup>His</sup> gene generates a tRNA lacking an extra 5' nucleotide (40), which calls into question the report in an early tRNA sequence compilation (itself re-reported from a meeting abstract and since removed from the compilation) that the tRNA<sup>His</sup> of *E. coli* phage T5 retains U-1 (38). In other cases there is no compelling reason to question the presence of the extra nucleotide. Whereas phage tRNA<sup>His</sup> genes differ from those of their hosts, the tRNA<sup>His</sup> gene of mimivirus, which infects amoebae, matches those of eukaryotes and so therefore should its product (Fig. 1). Other bacteria, plastids, many mitochondria, and most archaea en-

code G-1 in their tRNA<sup>His</sup> genes. A small number of archaea fail to encode G-1, but these species are all equipped with a Thg1 homolog and so can be expected to add G-1 posttranscriptionally by the eukaryotic mechanism.

The novel tRNA<sup>His</sup> might function more efficiently during translation. By occupying its discriminator nucleotide in a base-pair with G-1, typical bacterial tRNA<sup>His</sup> has a shorter 3' tail than does standard tRNA. The potential G-1:A73 non-canonical pair in eukaryotic tRNA<sup>His</sup> may likewise limit the flexibility of the tail. These unconformities could impair accommodation in the active site of the ribosome during translation of His codons, an effect that would be relieved with the standard 3' tail of the novel tRNA<sup>His</sup> in the RRCPP group.

**Additional novelty of RRCPP tRNA<sup>His</sup>.** The tRNA<sup>His</sup> genes in the RRCPP group resemble those of eukaryotes in encoding A73 at the discriminator position and failing to encode G-1. Recent phylogenetic analyses of tRNA<sup>His</sup> gene sequences and HisRS suggest that (i) the novel RRCPP tRNA gene originated by divergence of an ancestral alphaproteobacterial tRNA<sup>His</sup> gene, not by lateral transfer from a eukaryotic donor; (ii) that the RRCPP HisRS is of the type found in eukaryotes which has replaced the typical bacterial HisRS in prokaryotes in several cases; and (iii) that a prokaryote is a more likely proximal donor of this HisRS to the RRCPP bacteria (2). The links to the eukaryotic system led to the presumptions by these authors that tRNA<sup>His</sup> identity rules would be like those of eukaryotes and that the G-1 would still be present in the mature RRCPP tRNA<sup>His</sup>. Instead, the absence of the -1 nucleotide shows that tRNA<sup>His</sup> identity rules in the RRCPP group are unlike those in any other system.

An additional novelty in the RRCPP group is that the base pair between positions +1 and 72, which is a G:C pair in nearly all other tRNA<sup>His</sup>, has become a G:U pair. This serves as the closing base pair for the acceptor stem and is a good candidate for an identity determinant for the novel tRNA (39).

**Disjoint ancestry of RRCPP bacteria.** If the species of the RRCPP group comprise a single clade, the distribution of the novel HisRS/tRNA<sup>His</sup> system could be explained by vertical inheritance after an instance of coevolution of the molecules. If the species are split into multiple clades within the alphaproteobacteria, this might point to lateral transfer of one or both genes. Published genome-based species trees of the alphaproteobacteria have placed *Rhizobiales*, *Rhodobacterales*, and *Caulobacterales* together in a clade, apart from the *Sphingomonadales*, *Rhodospirillales*, and *Rickettsiales*, but do not address the affiliations of the more recently sequenced *Parvularcula* and *Pelagibacter* (33). A recent multiprotein study of 72 alphaproteobacterial strains (K. P. Williams, B. W. Sobral, and A. Dickerman, unpublished data) agrees with the above findings and additionally places *Parvularcula* close to *Caulobacter* but places *Pelagibacter* at the base of the *Rickettsiales* branch with high support, apart from the other species with the novel tRNA<sup>His</sup> system. This suggests lateral transfer of the novel HisRS gene within the alphaproteobacteria, most likely directed from a member of the *Rhizobiales/Rhodobacterales/Caulobacterales/Parvularculales* clade into an ancestor of *Pelagibacter*. The native tRNA<sup>His</sup> gene of the recipient could have undergone the two key base substitutions (C72T and C73A) after transfer of the HisRS gene alone, but the juxtaposition of the HisRS and tRNA<sup>His</sup> genes within an apparent

operon in *Pelagibacter* suggests instead that the two genes were transferred together.

**Correlated variation in HisRS.** In vivo and in vitro studies of yeast HisRS and in vitro study of *E. coli* HisRS show that the presence of G-1 is the major identity element of tRNA<sup>His</sup>, whereas an in vivo study in *E. coli* indicates that the discriminator base C73, unique among *E. coli* tRNAs, is a more important identity element (18, 22, 29, 40). The RRCPP group has lost both of these key tRNA<sup>His</sup> identity elements. This would appear to be detrimental to translational accuracy unless the losses are compensated. Without implying any particular ordered evolutionary sequence, compensation for these tRNA changes might occur at three levels: (i) HisRS must efficiently recognize the novel tRNA<sup>His</sup>; (ii) the HisRS changes, allowing this recognition may require further changes in HisRS or in other tRNAs, so that other tRNAs (which the novel tRNA<sup>His</sup> resembles more than standard tRNA<sup>His</sup> does) are not mischarged with histidine; and (iii) other aaRS-tRNA recognition rules may need to change so that the less distinctive tRNA<sup>His</sup> is not mischarged with other amino acids. Regarding the first point, any or all of the unique features of the RRCPP HisRS described in Results may prove to enable recognition. This is especially likely for the motif 2 and flipping loops, and the inserts in the insertion and catalytic domains are roughly positioned such they could also plausibly interact with the acceptor end. It should be noted that very large displacements of the motif 2 and flipping loops are not likely since these same loops also contribute to the activation of histidine.

An insert at position 405 is unique among all homologous anticodon-binding domains of class IIa aaRS. The position of insert 405 is also interesting because an amino acid substitution at this same site was obtained in a selection for *E. coli* HisRS that could compensate for a mutation at the acceptor end of suppressor tRNA<sup>His</sup> (41). In vitro, this HisRS mutant had tighter binding to each form of tRNA tested, although with reduced discrimination between cognate and noncognate tRNAs (6). Insert 405 may function similarly in RRCPP HisRS, compensating for weakened interaction with tRNA in the catalytic domain by promoting tighter binding in the anticodon-binding domain.

**Effects on other systems in RRCPP cells.** Regarding competition among other tRNAs for charging by HisRS, the RRCPP G<sub>+</sub>:U<sub>72</sub> pairing, although phylogenetically unique among tRNA<sup>His</sup>, is not unique among the tRNAs within a cell; for example, an elongator tRNA<sup>Met</sup> of *S. meliloti* has the same closing base pair and discriminator base as tRNA<sup>His</sup>.

An initial investigation into whether other tRNAs may have adapted to the unusual His system revealed that a three-H-bond base pair is found in the second position of the acceptor stem in tRNA<sup>Glu</sup> of all RRCPP bacteria, where a two-H-bond base pair is found in all non-RRCPP alphaproteobacteria. This may represent an adaptation to prevent mischarging by the unusual HisRS of the RRCPP group, although we do not have a ready structural explanation. Because one RRCPP genus (*Pelagibacter*) is not of the same ancestry as the others, the phylogenetic distribution of tRNA<sup>Glu</sup> is not a simple coincidence of two variants vertically inherited from a single ancestor. The *Pelagibacter* tRNA<sup>Glu</sup> gene may have hitchhiked on the same presumptive transfer event that brought in the tRNA<sup>His</sup>-HisRS operon, but these units are currently distant in the genome.



Full compensation for the loss of the extra guanylate may also require changes in other aaRS so that they can discriminate against the less distinctive tRNA<sup>His</sup>; it will be interesting to investigate whether non-HisRS aaRS signatures change in the RRCPP group. As an alternative to compensation at the level of global tRNA charging, it is at least possible that the unusual tRNA<sup>His</sup> and HisRS have resulted in an uncompensated reduction in tRNA charging accuracy that these bacteria simply abide.

A second arena where there might be a change correlated with the novel tRNA<sup>His</sup> is in RNase P structure and function. It will be interesting to investigate whether RNase P in the RRCPP group, like eukaryotic RNase P (9), is unable to perform the unusual cleavage required for the canonical bacterial tRNA<sup>His</sup>. If so this could help elucidate features of bacterial RNase P responsible for the unusual cleavage.

**Conclusion.** While the loss of a universal and apparently ancient tRNA feature was especially surprising, the additional changes to both tRNA<sup>His</sup> and HisRS in the alphaproteobacterial group further indicate a radical departure from previously known tRNA<sup>His</sup> identity rules. It will certainly be interesting to investigate the structure and in vitro tRNA recognition properties of this novel HisRS.

#### ACKNOWLEDGMENTS

This study was supported by the Virginia Bioinformatics Institute and by U.S. Department of Defense grant W911SR-04-0045 to B.W.S.

#### REFERENCES

- Aberg, A., A. Yaremchuk, M. Tkalco, B. Rasmussen, and S. Cusack. 1997. Crystal structure analysis of the activation of histidine by *Thermus thermophilus* histidyl-tRNA synthetase. *Biochemistry* **36**:3084–3094.
- Ardell, D. H., and S. G. Andersson. 2006. TFAM detects co-evolution of tRNA identity rules with lateral transfer of histidyl-tRNA synthetase. *Nucleic Acids Res.* **34**:893–904.
- Barnett, M. J., C. J. Toman, R. F. Fisher, and S. R. Long. 2004. A dual-genome symbiosis chip for coordinate study of signal exchange and development in a prokaryote-host interaction. *Proc. Natl. Acad. Sci. USA* **101**:16636–16641.
- Blakemore, R. P., D. Maratea, and R. S. Wolfe. 1979. Isolation and pure culture of a freshwater magnetic spirillum in chemically defined medium. *J. Bacteriol.* **140**:720–729.
- Bond, J. P., and C. Francklyn. 2000. Proteobacterial histidine-biosynthetic pathways are paraphyletic. *J. Mol. Evol.* **50**:339–347.
- Bovee, M. L., W. Yan, B. S. Sprout, and C. S. Francklyn. 1999. tRNA discrimination at the binding step by a class II aminoacyl-tRNA synthetase. *Biochemistry* **38**:13725–13735.
- Briand, C., A. Poterszman, S. Eiler, G. Webster, J. Thierry, and D. Moras. 2000. An intermediate step in the recognition of tRNA<sup>Asp</sup> by aspartyl-tRNA synthetase. *J. Mol. Biol.* **299**:1051–1060.
- Bruce, A. G., and O. C. Uhlenbeck. 1978. Reactions at the termini of tRNA with T4 RNA ligase. *Nucleic Acids Res.* **5**:3665–3677.
- Burkard, U., I. Willis, and D. Söll. 1988. Processing of histidine transfer RNA precursors: abnormal cleavage site for RNase P. *J. Biol. Chem.* **263**:2447–2751.
- Connolly, S. A., A. E. Rosen, K. Musier-Forsyth, and C. S. Francklyn. 2004. G-1:C73 recognition by an arginine cluster in the active site of *Escherichia coli* histidyl-tRNA synthetase. *Biochemistry* **43**:962–969.
- Cooley, L., B. Appel, and D. Söll. 1982. Post-transcriptional nucleotide addition is responsible for the formation of the 5' terminus of histidine tRNA. *Proc. Natl. Acad. Sci. USA* **79**:6475–6479.
- Cusack, S., M. Hartlein, and R. Leberman. 1991. Sequence, structural and evolutionary relationships between class 2 aminoacyl-tRNA synthetases. *Nucleic Acids Res.* **19**:3489–3498.
- Edgar, R. C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**:113.
- Eiler, S., A.-C. Dock-Bregeon, L. Moulinier, J.-C. Thierry, and D. Moras. 1999. Synthesis of aspartyl-tRNA<sup>Asp</sup> in *Escherichia coli*: a snapshot of the second step. *EMBO J.* **18**:6532–6541.
- Finn, R. D., J. Mistry, B. Schuster-Bockler, S. S. Griffiths-Jones, V. Hollich, T. Lassmann, S. Moxon, M. Marshall, A. Khanna, R. Durbin, S. R. Eddy, E. L. Sonnhammer, and A. Bateman. 2006. Pfam: clans, web tools and services. *Nucleic Acids Res.* **34**:D247–D251.
- Francklyn, C., and J. Arnez. 2005. Histidyl-tRNA synthetases, p. 847–858. *In* M. Iba, C. Francklyn, and S. Cusack (ed.), *The aminoacyl-tRNA synthetases*. Landes Bioscience, Georgetown, TX.
- Fromant, M., P. Plateau, and S. Blanquet. 2000. Function of the extra 5'-phosphate carried by histidine tRNA. *Biochemistry* **39**:4062–4067.
- Gu, W., R. L. Hurto, A. K. Hopper, E. J. Grayhack, and E. M. Phizicky. 2005. Depletion of *Saccharomyces cerevisiae* tRNA<sup>His</sup> guanylyltransferase Thg1p leads to uncharged tRNA<sup>His</sup> with additional m<sup>5</sup>C. *Mol. Cell. Biol.* **25**:8191–8201.
- Gu, W., J. E. Jackman, A. J. Lohan, M. W. Gray, and E. M. Phizicky. 2003. tRNA<sup>His</sup> maturation: an essential yeast protein catalyzes addition of a guanine nucleotide to the 5' end of tRNA<sup>His</sup>. *Genes Dev.* **17**:2889–2901.
- Guex, N., and M. C. Peitsch. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**:2714–2723.
- Hawko, S. A., and C. S. Francklyn. 2001. Covariation of a specificity-determining structural motif in an aminoacyl-tRNA synthetase and a tRNA identity element. *Biochemistry* **40**:1930–1936.
- Himeno, H., T. Hasegawa, Ueda, K. Watanabe, K. Miura, and M. Shimizu. 1989. Role of the extra G-C pair at the end of the acceptor stem of tRNA<sup>His</sup> in aminoacylation. *Nucleic Acids Res.* **17**:7855–7863.
- Jacob, Y., E. Seif, P. O. Paquet, and B. F. Lang. 2004. Loss of the mRNA-like region in mitochondrial tmRNAs of jakobids. *RNA* **10**:605–614.
- Kleina, L. G., J. M. Masson, J. Normanly, J. Abelson, and J. H. Miller. 1990. Construction of *Escherichia coli* amber suppressor tRNA genes. II. Synthesis of additional tRNA genes and improvement of suppressor efficiency. *J. Mol. Biol.* **213**:705–717.
- L'Abbé, D., B. F. Lang, P. Desjardin, and R. Morais. 1990. Histidine tRNA from chicken mitochondria has an uncoded 5'-terminal guanylate residue. *J. Biol. Chem.* **265**:2988–2992.
- Laslett, D., and B. Canback. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **32**:11–16.
- Lowe, T. M., and S. R. Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**:955–964.
- Marck, C., and H. Grosjean. 2002. tRNomics: analysis of tRNA genes from 50 genomes of *Eukarya*, *Archaea*, and *Bacteria* reveals anticodon-sparing strategies and domain-specific features. *RNA* **8**:1189–1232.
- Nameki, N., H. Asahara, M. Shimizu, N. Okada, and H. Himeno. 1995. Identity elements of *Saccharomyces cerevisiae* tRNA<sup>His</sup>. *Nucleic Acids Res.* **23**:389–394.
- Orellana, O., L. Cooley, and D. Söll. 1986. The additional guanylate at the 5' terminus of *Escherichia coli* tRNA<sup>His</sup> is the result of unusual processing by RNase P. *Mol. Cell. Biol.* **6**:525–529.
- Rosen, A. E., and K. Musier-Forsyth. 2004. Recognition of G-1:C73 atomic groups by *Escherichia coli* histidyl-tRNA synthetase. *J. Am. Chem. Soc.* **126**:64–65.
- Rudinger, J., C. Florentz, and R. Giegé. 1994. Histidyl-tRNA synthetase of yeast HisRS of tRNA or tRNA-like structure relies on residues –1 and 73 but is dependent on the RNA context. *Nucleic Acids Res.* **22**:5031–5037.
- Sällström, B., and S. G. Andersson. 2005. Genome reduction in the alpha-Proteobacteria. *Curr. Opin. Microbiol.* **8**:579–585.
- Sankaranarayanan, R., A.-C. Dock-Bregeon, P. Romby, J. Caillet, M. Springer, B. Rees, C. Ehresmann, B. Ehresmann, and D. Moras. 1999. The structure of threonyl-tRNA synthetase-tRNA<sup>Thr</sup> complex enlightens its receptor activity and reveals an essential zinc ion in the active site. *Cell* **97**:371–381.
- Sauter, C., B. Lorber, J. Cavarelli, D. Moras, and R. Giege. 2000. The free yeast aspartyl-tRNA synthetase differs from the tRNA<sup>Asp</sup>-complexed enzyme by structural changes in the catalytic site, hinge region, and anticodon-binding domain. *J. Mol. Biol.* **299**:1313–1324.
- Seetharaman, M., C. Williams, C. J. Cramer, and K. Musier-Forsyth. 2004. Effect of G-1 of histidine tRNA microhelix conformation. *Nucleic Acids Res.* **31**:7311–7321.
- Sprinzl, M., and K. S. Vassilenko. 2005. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **33**:D139–D140.
- Sprinzl, M., F. Grueter, A. Spelzhaus, and D. H. Gauss. 1980. Compilation of tRNA sequences. *Nucleic Acids Res.* **8**:r1–r22.
- Varani, G., and W. H. McClain. 2000. The G-U wobble base pair: a fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep.* **1**:18–23.
- Yan, W., and C. Francklyn. 1994. Cytosine 73 is a discriminator nucleotide in vivo for histidyl-tRNA in *Escherichia coli*. *J. Biol. Chem.* **269**:10022–10027.
- Yan, W., J. Augustine, and C. Francklyn. 1996. A tRNA identity switch mediated by the binding interaction between a tRNA anticodon and the accessory domain of a class II aminoacyl-tRNA synthetase. *Biochemistry* **35**:6559–6568.
- Yokobori, S., and S. Pääbo. 1995. Transfer RNA editing in land snail mitochondria. *Proc. Natl. Acad. Sci. USA* **92**:10432–10435.