Expanded use of sense codons is regulated by modified cytidines in tRNA

William A. Cantara[a,b,1], Frank V. Murphy IV[c], Hasan Demirci[d,2], and Paul F. Agris[a,3]

[a]The RNA Institute, Department of Biological Sciences, University at Albany–State University of New York, Albany, NY 12222; [b]Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695-7622; [c]Northeastern Collaborative Access Team, Argonne National Laboratory, Argonne, IL 60439; and [d]Department of Molecular Biology, Cell Biology, and Biochemistry, Brown University, Providence, RI 02912

Edited by Dieter Söll, Yale University, New Haven, CT, and approved May 23, 2013 (received for review December 26, 2012)

Codon use among the three domains of life is not confined to the universal genetic code. With only 22 tRNA genes in mammalian mitochondria, exceptions from the universal code are necessary for proper translation. A particularly interesting deviation is the decoding of the isoleucine AUA codon as methionine by the one mitochondrial-encoded $tRNA^{Met}$. This tRNA decodes AUA and AUG in both the A- and P-sites of the metazoan mitochondrial ribosome. Enrichment of posttranscriptional modifications is a commonly appropriated mechanism for modulating decoding rules, enabling some tRNA functions while restraining others. In this case, a modification of cytidine, 5-formylcytidine (f^5C), at the wobble position-34 of human mitochondrial $tRNA^{Met}_{f5CAU}$ ($hmtRNA^{Met}_{f5CAU}$) enables expanded decoding of AUA, resulting in a deviation in the genetic code. Visualization of the codon•anticodon interaction by X-ray crystallography revealed that recognition of both A and G at the third position of the codon occurs in the canonical Watson–Crick geometry. A modification-dependent shift in the tautomeric equilibrium toward the rare imino-oxo tautomer of cytidine stabilizes the f^5C_{34}•A base pair geometry with two hydrogen bonds.

modified nucleosides | ribosome crystallography | tautomerism

The genetic code was initially deemed to be universal and frozen in time (1). However, deviations in sense and nonsense codon use are found in bacteria, archaea, and both nuclear and organellar eukaryotic genomes (2, 3). Use of genetic codes that deviate from the universal code provides insight into its evolution (4) and possibilities for investigator-initiated manipulation (synthetic biology) (5). In many cases, however, the translation of the deviant sense codons is facilitated by posttranscriptional modification chemistries that are enzymatically added to nucleosides at the first anticodon position. The modification chemistries and their impact on anticodon conformation alter the decoding capacity of the modified tRNA (6). When first proposed, modification-dependent wobble decoding was limited to inosine as the first modified anticodon residue. Inosine, a deaminated adenosine residue, expands the ability of a single isoacceptor tRNA to read three codons by base pairing with either U, C, or A at the third position of the codon (7). Many other wobble position–modified residues, mostly pyrimidines, are now known to modulate use of specific codons (8). Although modified uridines constitute a great majority of these modifications, modified cytidines are prevalent in controlling a switch between the universal genetic code and a deviant code in the shared isoleucine/methionine codon box (9, 10). The modification, 4-acetylcytidine (ac^4C), prevents $tRNA^{Met}_{CAU}$ from reading AUA through wobble geometry (11), and lysidine (k^2C) (12) and agmatidine (agm^2C) (13, 14) prevent $tRNA^{Ile}_{CAU}$ from reading AUG in bacteria and archaea. Interestingly, a fourth cytidine modification, 5-formylcytidine (f^5C), facilitates the reading of AUA and AUG as methionine by a single $tRNA^{Met}_{f5CAU}$ responding to both initiator and elongator codons in yeast and many metazoan mitochondrial genomes (6). Therefore, ac^4C, k^2C and agm^2C are restrictive modifications that alter the physicochemical properties of the Watson–Crick edge, whereas f^5C expands decoding using a modification on the C-H edge at the C5 position. It is evident that the many posttranscriptional modifications of the Watson–Crick edge alter base pairing abilities, but a clear mechanism of decoding expansion by C5 modifications remains poorly understood, especially modifications of cytidine.

The mitochondrion's decoding of AUA as methionine is important for proper translation. In humans, this codon constitutes 20% of mRNA initiator methionines and 80% of internal methionines (15, 16). Using chemical synthesis and incorporation of the f^5C_{34} modification into the heptadecamer anticodon stem and loop domain (ASL) of human mitochondrial $tRNA^{Met}_{f5CAU}$($hmASL^{Met}_{f5CAU}$) (17), we determined that f^5C_{34} destabilized the $hmASL^{Met}_{f5CAU}$ by increasing the motional dynamics of the loop residues (17). A further study detailing the codon-binding characteristics and solution structure of $hmASL^{Met}_{f5CAU}$ agreed with the f^5C-dependent increase in residue dynamics (18). Proposed mechanisms for the decoding of AUA by $tRNA^{Met}_{f5CAU}$ depend on the f^5C_{34}•A_3 base pair forming a specific geometry (18–20). Based on molecular dynamics simulations, we suggested that this base pair could be in a sheared geometry that is supported by a bridging water molecule (18). To test this hypothesis, the geometry of the f^5C_{34}•A_3 base pair in the decoding center of the ribosomal A-site was observed in crystal structures of natively modified $hmASL^{Met}_{f5CAU}$ bound to AUA. Here, we show for the first time the modification-dependent f^5C_{34}•A_3 base pair within the codon•anticodon interaction during ribosomal A-site decoding. Surprisingly, f^5C_{34} forms a canonical Watson–Crick base pair with both the G of AUG and the A of AUA, refuting the conformation predicted from molecular dynamics simulations (18). This geometry of the f^5C_{34}•A_3 base pair requires a novel amino-imino tautomerism in f^5C_{34} similar to the keto-enol tautomerism seen for the two wobble position uridines cmo^5U_{34} and $mcm^5s^2U_{34}$ in *Escherichia coli* $tRNA^{Val3}_{UAC}$ (21) and human $tRNA^{Lys3}_{UUU}$ (22), respectively.

BIOCHEMISTRY

Author contributions: P.F.A. designed research; W.A.C., F.V.M., and H.D. performed research; W.A.C., F.V.M., H.D., and P.F.A. analyzed data; and W.A.C., F.V.M., H.D., and P.F.A. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The atomic coordinates for $hmASL^{Met}{}_{f5CAU}$-AUG and $hmASL^{Met}{}_{f5CAU}$-AUA have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 4GKJ and 4GKK, respectively).

[1]Present address: Department of Chemistry and Biochemistry, Center for Retroviral Research, and Center for RNA Biology, Ohio State University, Columbus, OH 43210.

[2]Present address: Stanford PULSE Institute, SLAC National Accelerator Laboratory, Menlo Park, CA 94025.

[3]To whom correspondence should be addressed. E-mail: pagris@albany.edu.

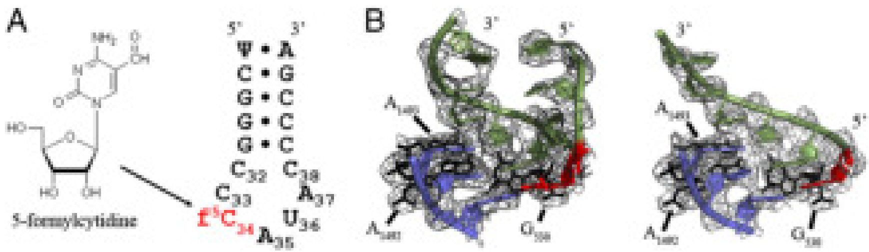

Fig. 1. hmASL$^{Met}_{f5CAU}$ secondary structure and A-site codon•anticodon interaction. (*A*) The ASL was synthesized as a heptadecamer containing two modifications: pseudouridine at position 27 and 5-formylcytidine at position 34. (*B*) The structure of hmASL$^{Met}_{f5CAU}$ bound to AUG (*Left*) showed strong electron density for ASL residues 31–39, whereas when bound to AUA (*Right*), electron density was strong for only residues 34–39. In both structures, the ASL is in green, the mRNA codon is in blue, and the A-site interacting residues (G_{530}, A_{1492}, and A_{1493}) are in black ($2mF_O$-dF_C contoured at 1.5 σ). The f^5C_{34} modification is colored red.

Results

Codon Bound hmASL$^{Met}_{f5CAU}$ Structure. The ASL domain of hmtRNA$^{Met}_{f5CAU}$ was chemically synthesized with the wobble modification f^5C_{34}, the native C_{33}, and a pseudouridine, Ψ_{27}, at the 5′ terminus (Fig. 1*A*). C_{33} is quite rare. The uridine at position 33 in tRNAs is considered invariant and recognized for its contribution to the "U-turn" structural motif. There are only 21 known instances of C_{33}, 13 of which are found in initiator tRNAs (23). The hmASL$^{Met}_{f5CAU}$ bound both AUG and AUA codons with significant affinity in the A-site of the bacterial ribosome (17) and within the A-site of the bovine mitochondrial 55S ribosome (18). In determining the crystallographic structure of the modified hmASL$^{Met}_{f5CAU}$ bound to the AUG or the AUA codons in the ribosomal A-site, native *Thermus thermophilus* 30S ribosomal subunit crystals were soaked with the hmASL$^{Met}_{f5CAU}$ and hexameric oligonucleotides, each containing either AUG or AUA codons. Under the conditions used, both crystals diffracted to 3.3-Å resolution (Table 1). Unbiased difference Fourier electron density maps were generated and used for building the hmASL$^{Met}_{f5CAU}$ and mRNA residues into the X-ray crystal structures. The structure of the 30S ribosomal subunit (including all RNA and protein) was nearly identical to those reported previously (21, 22). The hmASL$^{Met}_{f5CAU}$ took a conformation nearly identical to that of the ASL of a ribosome-bound tRNA-EF-Tu complex (24, 25), demonstrating the biological relevance of the present structures. More importantly, the conserved A-site residues G_{530}, A_{1492}, and A_{1493} were in the correct orientation to constrain the codon•anticodon pair residues into the proper geometry for recognition (Fig. 1*B*) (26, 27). As such, differences in the characteristics of the codon•anticodon interaction can be attributed to the specific conformation of hmASL$^{Met}_{f5CAU}$ and properties of the f^5C_{34} modification.

Table 1. Data collection and refinement statistics

Dataset	ASL$^{Met}_{f5CAU}$;Ψ_{27}-AUG	ASL$^{Met}_{f5CAU}$;Ψ_{27}-AUA
Data collection		
Space group	$P4_12_12$	$P4_12_12$
Cell dimensions		
a, b, c (Å)	400.6, 400.6, 176.2	402.4, 402.4, 175.6
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	80.1–3.3 (3.38–3.30)	97.6–3.3 (3.38–3.30)
R_{merge}	22.5 (133.7)	19.8 (113.3)
I/σI	6.35 (1.19)*	5.50 (1.25)*
$CC_{1/2}$	0.99 (0.41)	0.99 (0.49)
Completeness (%)	97.4 (95.0)	94.7 (93.5)
Redundancy	3.3 (3.2)	2.6 (2.5)
Refinement		
No. reflections	207,954	230,910
R_{work}/R_{free}	18.8/22.2	19.5/23.6
No. atoms	52,227	52,187
RNA	32,766	32,726
Protein	19,231	19,231
Ion	188	188
Paromomycin	42	42
B-factor	80.1	87.4
r.m.s. deviations		
Bond lengths (Å)	0.008	0.008
Bond angles (°)	1.36	1.35

Parentheses indicate highest-resolution shell.
*I/σI = 2 at 3.48 Å for both structures.

Conformational Characteristics of hmASL$^{Met}_{f5CAU}$. Nearly the entire hmASL$^{Met}_{f5CAU}$ exhibited strong electron density when bound to the cognate AUG codon (Fig. 1*B*). In the structure of hmASL$^{Met}_{f5CAU}$ bound to the wobble AUA codon, the codon•anticodon interaction and the 3′ side are clearly resolved. However, there was poor electron density from the 5′-end to the anticodon similar to the poor electron density observed in crystal structures of the mitochondrial ASL$^{Leu}_{\tau m5UAA}$ (28) and 8-nt loop containing ASLs (29), both in the *T. thermophilus* ribosomal A-site. Therefore, the overall conformational properties of the hmASL$^{Met}_{f5CAU}$ are best described from the perspective of the cognate codon-bound structure, whereas conclusions about the AUA-bound structure will be addressed from comparisons of its highly resolved codon•anticodon interaction and 3′-stacked nucleosides. In the canonical U-turn motif of tRNA's anticodon loop, a sharp backbone curvature occurs between U_{33} and the wobble nucleoside at position 34 (30). Despite the presence of a C_{33}, the hmASL$^{Met}_{f5CAU}$ had a U-turn–like architecture. This result confirms NMR (18) and biophysical (17) characterizations. At physiological pH, protonation of U_{33}N3 facilitates hydrogen bond formation between U_{33}N3 and the phosphate of nucleoside 36, stabilizing the characteristic U-turn (30, 31). Although C_{33}N3 is not protonated, the solution structure of hmASL$^{Met}_{f5CAU}$ by NMR suggested a weak C_{33}•U_{36} interaction (18). In the crystal structures, the backbone at C_{33} was found to be distorted in such a way that the position of C_{33} was shifted from that expected of a U_{33}. A superimposition of our structure with that of the *E. coli* elongator ASL$^{Lys}_{UUU}$ (32) and the human elongator ASL$^{Lys3}_{UUU}$ (22), both of which contain a U_{33} and a U_{36} (Fig. 2*B*), confirmed that the position of C_{33} was slightly altered from the geometry of a U_{33}. This positioning allows for a hydrogen bond between C_{33}N4 and the phosphate of U_{36} (Fig. 2*A*), thus stabilizing the backbone turn and the base pairing of the wobble position f^5C_{34}.

Two other properties of the canonical U-turn were investigated: the positioning of A_{37} over the first codon•anticodon base pair (22, 32, 33) and the formation of a noncanonical

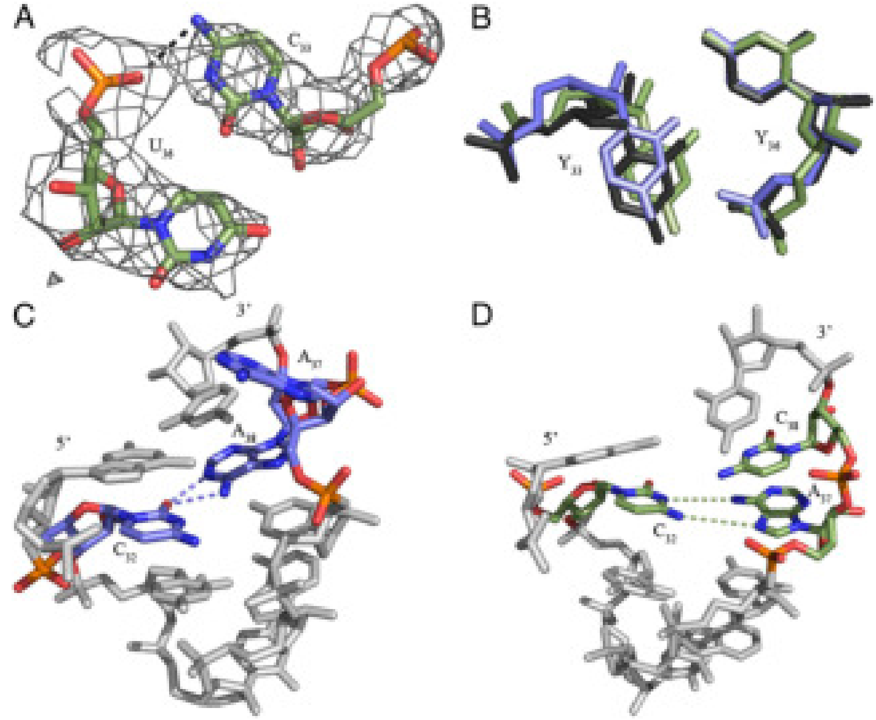

Fig. 2. Unusual features of the hmASL$^{Met}_{f5CAU}$ structure. (*A*) A hydrogen bond is able to form between C_{33}N4 and the phosphate of U_{36}. ASL carbons are colored green, and mRNA carbons are blue (m2F$_O$-dF$_C$ contoured at 1.5 σ). (*B*) hmASL$^{Met}_{f5CAU}$ (green) is superimposed with *E. coli* elongator tRNA$^{Lys}_{UUU}$ (blue) and human elongator tRNA$^{Lys3}_{UUU}$ (black), showing a slightly distorted conformation. (*C*) In *E. coli* tRNA$^{Metf}_{CAU}$, the cross-loop interaction involves the familiar C_{32}•A_{38} noncanonical base pair with A_{37} displaced from the loop, where it cannot participate in base stacking. (*D*) A cross-loop interaction in hmASL$^{Met}_{f5CAU}$ involves a unique interaction between C_{32} and A_{37} that consists of a base pair between the Watson–Crick edge of C_{32} and the Hoogsteen edge of A_{37}.

base pair between positions 32 and 38 (32–34). In ASLs responding to codons beginning with A, the conserved A_{37} is modified and positioned directly above the U_{36}•A_1 base pair (Fig. 1*B*) (22, 32, 33). In most organisms, the cytoplasmic methionyl-tRNAs contain an *N6*-threonylcarbamoyladenosine-37 (t^6A_{37}) (23) stacked with, and thus stabilizing, the U_{36}•A_1 base pair. However, A_{37} of hmtRNA$^{Met}_{f5CAU}$ is unmodified, yet appears to satisfactorily promote a stable U_{36}•A_1 base pair. The geometry of the noncanonical base pair between residues 32 and 38 was also examined because of its importance in wobble decoding (34). Although the *E. coli* initiator tRNA is closed by a common C_{32}•A_{38} noncanonical base pair (Fig. 2*C*), the primary sequence and secondary structure folding of hmtRNA$^{Met}_{f5CAU}$ would allow for a C_{32}•C_{38} cross-loop interaction. Interestingly, A_{37} and C_{38} adopt positions within the anticodon loop where C_{32} is planar with A_{37} rather than C_{38}. C_{32} is base paired with A_{37} (Fig. 2*D*).

f^5C_{34}•A_3 Base Pair Adopts a Watson–Crick-Like Geometry. The f^5C_{34}•A_3 and f^5C_{34}•A_3 base pairs are clearly defined in our crystal structures, and both are in Watson–Crick geometry (Fig. 3*A*). A superposition of the two crystal structures revealed nearly identical codon•anticodon base pair conformations (Fig. 3*B*). More importantly, this geometry indicates that N4 of the f^5C_{34} must be in the imino-oxo form rather than the common amino-oxo form (Fig. 4). If N4 of f^5C_{34} was in the universal amino form, the close proximity of f^5C_{34}N4 and A_3N6 (2.8 Å) would cause a very high electronic and steric repulsion. We hypothesized that

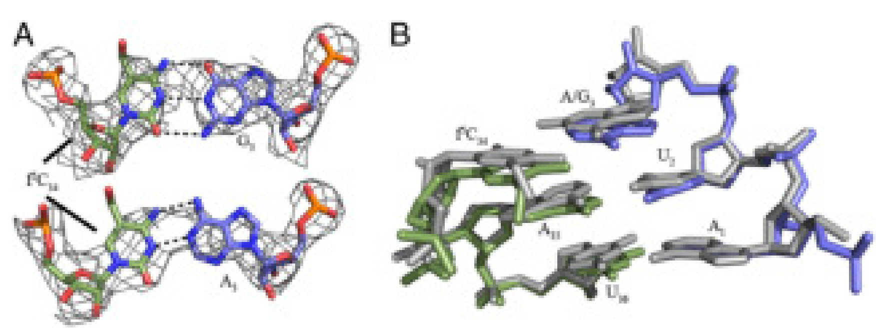

Fig. 3. Geometry of the wobble base pair and codon•anticodon interaction. (*A*) The electron density shows that both the f^5C_{34}•G_3 and f^5C_{34}•A_3 base pairs are in Watson–Crick geometry. ASL carbons are colored green, and mRNA carbons are blue (m2F$_O$-dF$_C$ contoured at 1.5 σ). (*B*) Superposition of the hmASL$^{Met}_{f5CAU}$•AUG complex with the hmASL$^{Met}_{f5CAU}$•AUA structure aligned with respect to the mRNA residues. The overlay reveals a nearly identical orientation of the codon•anticodon interaction between the AUA-bound and AUG-bound structures. The ASL and mRNA residues for the AUA-bound structures are green and blue, respectively. The ASL and mRNA residues of the AUG-bound structure are both gray.

BIOCHEMISTRY

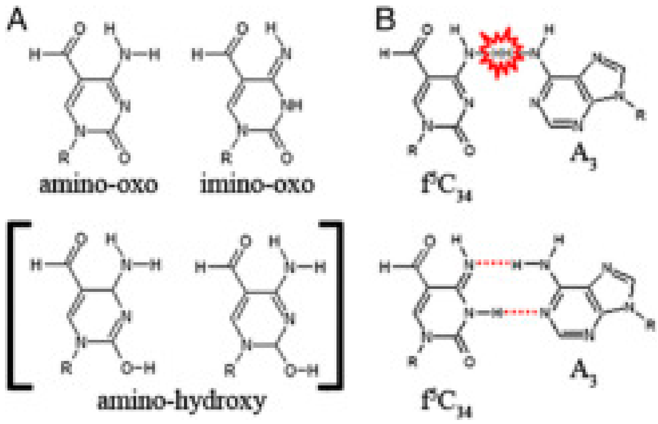

Fig. 4. Prototropic tautomerism alleviates steric repulsion and allows a Watson–Crick f^5C_{34}•A_3 base pair. (*A*) Three possible tautomeric forms of 5-formylcytidine are denoted by the chemical nature of the N4 and O2 positions. (*B*) In the common amino-oxo form (*Upper*) of f^5C_{34}, a Watson–Crick base pair with adenosine results in steric repulsion due to the proximity of f^5C_{34}N4H and A_3N6H (marked in red); however, the imino-oxo form (*Lower*) allows for a favorable U•A-like hydrogen bonding.

the formyl group at the C5 position may reduce the energy difference between the tautomeric forms, thus shifting the equilibrium toward the imino-oxo tautomer. A quantum mechanical ground-state free energy calculation was performed on the amino-oxo and imino-oxo forms of the bases cytosine and 5-formylcytosine. A free energy difference of 6.3 kcal/mol between the two tautomeric forms of cytosine favored the amino-oxo form and confirmed previous estimations between 5.4 and 6.7 kcal/mol (35). The tautomers of 5-formylcytosine differed in free energy by 7.8 kcal/mol. Therefore, the f^5C-dependent shift in tautomeric equilibrium is not caused by a simple reduction in the energetic properties of the free base.

Discussion

Cytoplasmic translation initiation at the AUG codon is distinct from elongation in which AUG is decoded in the A-site. There are two different cytoplasmic $tRNA^{Met}_{CAU}$ isoacceptors: one for initiating and one for elongating protein synthesis (17). In contrast in the human mitochondrion, a single $hmtRNA^{Met}_{f5CAU}$ initiates and elongates at either AUG or the universal isoleucine codon AUA. Eighty percent of mitochondrial mRNAs initiate translation with AUG; whereas 80% of all methionine codons within the mRNA are AUA (15, 16). The f^5C_{34}-modified $hmtRNA^{Met}_{f5CAU}$ decodes the isoleucine AUA codon as methionine in vivo and in vitro (20). The crystal structures presented here offer a structural and chemical rationale for how $hmtRNA^{Met}_{f5CAU}$ decodes both AUA and AUG and alters the anticodon domain architecture to act as both an initiator and elongator tRNA.

Three possible f^5C_{34}•A_3 base pairing strategies have been proposed. The base pair could be in a sheared geometry in which A_3N6 acts as a proton donor in a bifurcated hydrogen bond with f^5C_{34}N1 and O2. A bridging water molecule acts as a dual proton donor for hydrogen bonds with both f^5C_{34}O2 and A_3N1 (18). In another possible geometry, a single hydrogen bond is formed between f^5C_{34}N4 and A_3N1 (20). Finally, it has been proposed that A_3 could become protonated and form an f^5C_{34}•A_3^+ base pair. The resulting nonconomical pair could be in either of two different geometries: one in which f^5C_{34} is shifted such that f^5C_{34}O2 and N3 act as hydrogen bond acceptors with A_3N1 and N6, respectively (19), and one in which the bases are in the Watson–Crick orientation (20) with f^5C_{34}N3 bound to A_3N1^+.

Of the three possible base pairing schemes, only the Watson–Crick geometry with the protonated adenosine on the codon fits with our refined structure. The importance of Watson–Crick base pairing for codon recognition on the ribosome was first proposed by Crick (7), and more recently reiterated with the support of extensive crystallographic data (21, 22). However, a Watson–Crick geometry for an f^5C_{34}•A_3^+ base pair should be highly unstable, if not energetically impossible, due to the close proximity of the amine protons of f^5C_{34}N4 and A_3N6 (Fig. 4*B*). The most likely scenario for resolving the clear geometric restraints observed in the electron density with the physicochemical properties of the two residues involves a novel shift in the tautomeric equilibrium of modified f^5C_{34}. Of the three possible prototropic tautomers of cytidine (Fig. 4*A*), equilibrium must exist between the most common amino-oxo form for base pairing with G_3 and the less common imino-oxo tautomer for interacting with A_3. Although theoretical calculations coupled with empirical photoemission data have identified a rarely populated existence of the imino-oxo tautomer (36), the C5-formyl modification may perturb the electronic structure of the heterocycle, reducing the energy of this tautomer. Prototropic tautomerism is not unprecedented at the wobble position. C5-modified uridines at position 34 in *E. coli* $tRNA^{Val}_{cmo5UAC}$ (21) and human $tRNA^{Lys3}_{mcm5s2UUU}$ (22) also make use of tautomeric shifts to enable wobble decoding in the ribosomal A-site. Although the different tautomeric forms of uridine have been shown experimentally, this property has not been suggested for cytidine due to significant differences in the electronic structures of the two pyrimidines.

The apparent commonality of using prototropic tautomerization in tRNA's decoding of bacterial, metazoan, and mitochondrial mRNA compelled us to explore how the C5 position modifications facilitate tautomeric shifts. It has been estimated that common tautomers of nucleobases outnumber their rarer counterparts by $\sim10^4$–10^5 (35), constituting an energy difference between the two states of a mere 5.4–6.7 kcal/mol. Such a relatively small energetic difference resulting in a large shift in the populations of the different tautomeric states indicates that these modifications must be finely tuned for their function in protein synthesis. We hypothesized that the addition of the formyl group at C5 may contribute to the preference for a proton shift from the N4 amine to N3. However, our computational investigation of the effects of the weak electron withdrawing formyl group shows that it increases the energy difference between the tautomers of the free base from 6.3 kcal/mol for cytosine to 7.8 kcal/mol for 5-formylcytosine. However, the tautomeric equilibrium may be affected by interactions of the modification with the environment. This proposition is justified by a study of the interaction between glycine and uridine using density functional theory and showing that the enol form of uridine can be stabilized by up to five orders of magnitude through intermolecular interactions (37). Also, the potential for a pseudobicyclic ring system resulting from a hydrogen bond between the formyl oxygen and the N4 imine proton may contribute energy to the system in the form of improved base stacking interactions and ring current effects.

In addition to providing clear evidence of mechanism by which $hmtRNA^{Met}_{f5CAU}$ decodes AUA, the crystal structures provide insight into the ability of this single isoaccepting species to act as both an initiator and an elongator tRNA. Many other characteristics of the tRNA may be responsible for this dual role; however, architectural features of the ASL may also contribute. Because the mitochondrial translation system is more similar to prokaryotic translation (38, 39), we compared $hmtRNA^{Met}_{f5CAU}$ with prokaryotic initiator tRNAs. Although many examples of insect, plant, and vertebrate initiator tRNAs contain C_{33} (40–45), all known prokaryotic initiator tRNAs contain the universal U_{33}.

The ASL of the *E. coli* initiator $tRNA^{Met,f}_{CAU}$ has a peculiar characteristic in which U_{33} is positioned outside of the anticodon loop rather than interacting with the phosphate of residue 36 in the canonical U-turn (46). Although this is at odds with a solution structure indicating that U_{33} is positioned in the canonical U-turn geometry (47), it raises the possibility that the positioning of residue 33 is important for discriminating A-site and P-site binding. In comparison, the *E. coli* $tRNA^{Ile}_{k2AU}$ and archaeal $tRNA^{Ile}_{agm2AU}$ would decode AUG was it not for the restrictive recognition of A_3 by the modifications lysidine and agmatidine, respectively (12, 13). The contrast with the $hmtRNA^{Met}_{f5CAU}$ is dramatic. The $hmtRNA^{Met}_{f5CAU}$ in participating in both A-site and P-site decoding should have the flexibility to adopt different conformations depending on the present function. Indeed, a solution structure of the modified $hmASL^{Met}_{f5CAU}$ illustrates very little interaction between C_{33} and U_{36} (18), and the present crystal structure shows a clear hydrogen bonding between the two residues, albeit in an uncommon geometry. Therefore, our results suggest an induced-fit model that results in the A-site–bound conformation. The resulting conformation favors the canonical U-turn in which C_{33} interacts with U_{36} and may also fit into a slightly different conformation for the P-site. Additionally, the unique $C_{32}\bullet A_{37}$ base pair may further highlight the dual character of this tRNA. In elongator tRNAs, especially where a $U_{36}\bullet A_1$ base pair is present, A_{37} is typically modified to promote stacking that stabilizes this first base pair of the codon•anticodon interaction. Here the lack of an A_{37} modification may allow the $hmASL^{Met}_{f5CAU}$ to adopt the more "initiator-like" conformation for P-site binding, whereas the unusual $C_{32}\bullet A_{37}$ base pair promotes architecture more similar to the canonical elongator conformation for A-site codon recognition. The unusual $C_{32}\bullet A_{37}$ base pairing does not preclude the possibility of a change in the C_{32} base-pairing partner between A_{37} and C_{38} (Fig. 2*C*). Such a change in base pairing could correspond to an architectural switch of the elongator $hmtRNA^{Met}_{f5CAU}$ to that of the initiator function. Thus, we showed that tRNA's codon recognition is expanded through a modification-dependent f^5C_{34} amine-imine tautomerization of cytidine. Overall, the architecture afforded by the natural modifications and sequence change to C_{33} allow for a significantly altered mode of codon recognition, one in which a single tRNA acts as both an initiator and elongator. The single $tRNA^{Met}$ contributes to the mitochondrion's challenge of maintaining a minimal genome.

Materials and Methods

Crystallization. *T. thermophilus* 30S ribosomal subunits were purified, crystallized, and cryoprotected in 26% (vol/vol) 2-methyl-2,4-pentanediol (MPD), 100 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) (pH 6.5), 200 mM KCl, 75 mM NH_4Cl, and 15 mM $MgCl_2$ (48). Hexameric mRNA oligonucleotides were chemically synthesized and gel-purified using preparative PAGE (Thermo Fisher) with the sequences 5′-AU(A/G)AAA-3′ (codons underlined). After cryoprotection, the empty 30S ribosome crystals were soaked in cryoprotection solution supplemented with 80 μM paromomycin, 300 μM ASL, and 300 μM mRNA (>48 h) (26, 27). Crystals were flash-frozen using liquid nitrogen and stored for data collection. Paromomycin was used for its ability to enhance ASL density and resolution (26) by inducing a closed form of the 30S ribosomal subunit without altering the ASL conformation (27). A recent study has shown that lack of covalent attachment between P- and A-site codons in 30S ribosome structures negates some restraint naturally imposed on the first codon•anticodon base pair (corresponding to $U_{36}\bullet A_1$ in the current structures), allowing them to adopt a wobble base pairing geometry (49). However, currently known structures of cognate codon recognition in the 30S ribosomal subunit show this base pair in the canonical Watson–Crick geometry (21, 22), thereby imposing the same geometric restraints on the remaining codon residues as that seen in the 70S structures. Indeed, structures of cognate codon recognition on the 30S ribosomal subunit superimpose almost exactly with those of the 70S structures [all atom rmsd of codon•anticodon residues from Protein Data Bank entries 1IBL (30S) (26) and 2J00 (70S) (50) = 0.345 Å].

Data Collection and Refinement. Data were collected at The Northeastern Collaborative Access Team (NE-CAT) beamlines 24-ID-C and 24-ID-E of the Advanced Photon Source. Processing was performed using XDS (51), PHENIX (52) for format manipulation and refinement, Coot (53) for visualization and model building, and PyMOL (54) for figure production. Geometry restraints and dictionary files for the nonstandard residues (paromomycin and 5-formylcytidine) were generated with the eLBOW (55) module within PHENIX using the semiempirical quantum mechanical AM1 method. The two different structures were generated from data sets that were collected at different times; therefore, resolution differences in the structures were likely caused by differences in the X-ray beam rather than in the crystals themselves. The CC1/2 values are calculated using the newest version of XDS (51).

Relative Ground-State Energy Calculations. Structures corresponding to the amino-oxo tautomers of cytidine and 5-formylcytidine base were prepared using GaussView03 (56) by replacing the ribose moiety with a methyl group. All subsequent calculations were performed using Gaussian-03 (57). Additionally, the imino-oxo tautomers of each were built with the imino proton in the *trans*-orientation (facing away from the Watson–Crick face) to account for the geometry that would allow for base pairing without steric repulsion. The ground-state geometries were first optimized at the semiempirical level using the AM1 method. Further geometry optimization was achieved at the HF/6–31G(d,p) level. Single point energy calculations were then performed at the B3LYP/6–311++G(d,p) level. All HF and B3LYP calculations were performed with a Polarizable Continuum Model using the integral equation formalism variant (IEFPCM) to simulate water solvation (58).

ACKNOWLEDGMENTS. We thank R. Kaiser and M. O. Delaney for synthesis of the $hmASL^{Met}_{f5CAU}$ (17); W. D. Graham for purification of the ASL; and L. L. Spremulli, J. L. Spears, and C. J. Stark for critical manuscript reading. This work was supported by National Science Foundation Grant MCB1101859 (to P.F.A.). H.D. was supported by Grants GM019756 and GM094157 from the US National Institutes of Health. This work is based on research conducted at the Advanced Photon Source on the Northeastern Collaborative Access Team beamlines supported by grants from the National Center for Research Resources (5P41RR015301-10) and the National Institute of General Medical Sciences (8 P41 GM103403-10). Use of the Advanced Photon Source, an Office of Science User Facility operated for the US Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by US DOE Contract DE-AC02-06CH11357.

1. Crick FH (1968) The origin of the genetic code. *J Mol Biol* 38(3):367–379.
2. Koonin EV, Novozhilov AS (2009) Origin and evolution of the genetic code: The universal enigma. *IUBMB Life* 61(2):99–111.
3. Sengupta S, Yang X, Higgs PG (2007) The mechanisms of codon reassignments in mitochondrial genetic codes. *J Mol Evol* 64(6):662–688.
4. Cedergren RJ (1982) An evaluation of mitochondrial tRNA gene evolution and its relation to the genetic code. *Can J Biochem* 60(4):475–479.
5. Isaacs FJ, et al. (2011) Precise manipulation of chromosomes *in vivo* enables genome-wide codon replacement. *Science* 333(6040):348–353.
6. Watanabe K, Yokobori S (2011) tRNA modification and genetic code variations in animal mitochondria. *J Nucleic Acids* 2011:623095.
7. Crick FH (1966) Codon—anticodon pairing: The wobble hypothesis. *J Mol Biol* 19(2):548–555.
8. Agris PF (1991) Wobble position modified nucleosides evolved to select transfer RNA codon recognition: A modified-wobble hypothesis. *Biochimie* 73(11):1345–1349.
9. Cantara WA, et al. (2011) The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res* 39(Database issue):D195–D201.
10. Yokoyama S, Nishimura S (1995) Modified nucleosides and codon recognition. *tRNA: Structure, Biosynthesis and Function*, eds Schimmel PR, Söll D, RajBhandary UL (ASM Press, Washington, DC), pp 207–223.
11. Oashi Z, et al. (1972) Characterization of C^+ located in the first position of the anticodon of *Escherichia coli* $tRNA^{Met}$ as *N 4* -acetylcytidine. *Biochim Biophys Acta* 262(2):209–213.
12. Muramatsu T, et al. (1988) A novel lysine-substituted nucleoside in the first position of the anticodon of minor isoleucine tRNA from *Escherichia coli*. *J Biol Chem* 263(19):9261–9267.
13. Mandal D, et al. (2010) Agmatidine, a modified cytidine in the anticodon of archaeal $tRNA^{(Ile)}$, base pairs with adenosine but not with guanosine. *Proc Natl Acad Sci USA* 107(7):2872–2877.
14. Voorhees RM, et al. (2013) The structural basis for specific decoding of AUA by isoleucine tRNA on the ribosome. *Nat Struct Mol Biol* 20(5):641–643.

15. Jones CN, Wilkinson KA, Hung KT, Weeks KM, Spremulli LL (2008) Lack of secondary structure characterizes the 5′ ends of mammalian mitochondrial mRNAs. *RNA* 14(5): 862–871.
16. Nakamura Y, Gojobori T, Ikemura T (2000) Codon usage tabulated from international DNA sequence databases: Status for the year 2000. *Nucleic Acids Res* 28(1):292.
17. Lusic H, et al. (2008) Synthesis and investigation of the 5-formylcytidine modified, anticodon stem and loop of the human mitochondrial $tRNA^{Met}$. *Nucleic Acids Res* 36(20):6548–6557.
18. Bilbille Y, et al. (2011) The human mitochondrial $tRNA^{Met}$: Structure/function relationship of a unique modification in the decoding of unconventional codons. *J Mol Biol* 406(2):257–274.
19. Leontis NB, Stombaugh J, Westhof E (2002) The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res* 30(16):3497–3531.
20. Takemoto C, et al. (2009) Unconventional decoding of the AUA codon as methionine by mitochondrial $tRNA^{Met}$ with the anticodon f^5CAU as revealed with a mitochondrial in vitro translation system. *Nucleic Acids Res* 37(5):1616–1627.
21. Weixlbaumer A, et al. (2007) Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. *Nat Struct Mol Biol* 14(6):498–502.
22. Vendeix FA, et al. (2012) Human $tRNA^{(Lys3)}(_{UUU})$ is pre-structured by natural modifications for cognate and wobble codon binding through keto-enol tautomerism. *J Mol Biol* 416(4):467–485.
23. Jühling F, et al. (2009) tRNAdb 2009: Compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res* 37(Database issue):D159–D162.
24. Schmeing TM, et al. (2009) The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA. *Science* 326(5953):688–694.
25. Schmeing TM, Voorhees RM, Kelley AC, Ramakrishnan V (2011) How mutations in tRNA distant from the anticodon affect the fidelity of decoding. *Nat Struct Mol Biol* 18(4):432–436.
26. Ogle JM, et al. (2001) Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* 292(5518):897–902.
27. Ogle JM, Murphy FV, Tarry MJ, Ramakrishnan V (2002) Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell* 111(5):721–732.
28. Kurata S, et al. (2008) Modified uridines with C5-methylene substituents at the first position of the tRNA anticodon stabilize U.G wobble pairing during decoding. *J Biol Chem* 283(27):18801–18811.
29. Dunham CM, et al. (2007) Structures of tRNAs with an expanded anticodon loop in the decoding center of the 30S ribosomal subunit. *RNA* 13(6):817–823.
30. Quigley GJ, Rich A (1976) Structural domains of transfer RNA molecules. *Science* 194(4267):796–806.
31. Dirheimer G, Keith G, Dumas P, Westhof E (1995) Primary, secondary, and tertiary structures of tRNAs. *tRNA: Structure, Biosynthesis, and Function*, eds Söll D, RajBhandary U (ASM Press, Washington, DC), pp 93–126.
32. Murphy FV, Ramakrishnan V, Malkiewicz A, Agris PF (2004) The role of modifications in codon discrimination by $tRNA^{Lys}{}_{UUU}$. *Nat Struct Biol* 11(12):1186–1191.
33. Murphy FV, 4th, Ramakrishnan V (2004) Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nat Struct Mol Biol* 11(12):1251–1252.
34. Olejniczak M, Uhlenbeck OC (2006) tRNA residues that have coevolved with their anticodon to ensure uniform and accurate codon recognition. *Biochimie* 88(8):943–950.
35. Lippert B, Gupta D (2009) Promotion of rare nucleobase tautomers by metal binding. *Dalton Trans* (24):4619–4634.
36. Feyer V, et al. (2009) Tautomerism in cytosine and uracil: An experimental and theoretical core level spectroscopic study. *J Phys Chem A* 113(19):5736–5742.
37. Dabkowska I, Gutowski M, Rak J (2005) Interaction with glycine increases stability of a mutagenic tautomer of uracil. A density functional theory study. *J Am Chem Soc* 127(7):2238–2248.
38. Smits P, Smeitink J, van den Heuvel L (2010) Mitochondrial translation and beyond: Processes implicated in combined oxidative phosphorylation deficiencies. *J Biomed Biotechnol* 2010:737385.
39. Barraud P, Schmitt E, Mechulam Y, Dardel F, Tisné C (2008) A unique conformation of the anticodon stem-loop is associated with the capacity of tRNAfMet to initiate protein synthesis. *Nucleic Acids Res* 36(15):4894–4901.
40. Ghosh HP, Ghosh K, Simsek M, RajBhandary UL (1982) Nucleotide sequence of wheat germ cytoplasmic initiator methionine transfer ribonucleic acid. *Nucleic Acids Res* 10(10):3241–3247.
41. Gillum AM, Roe BA, Anandaraj MP, RajBhandary UL (1975) Nucleotide sequence of human placenta cytoplasmic initiator tRNA. *Cell* 6(3):407–413.
42. Gillum AM, Urquhart N, Smith M, RajBhandary UL (1975) Nucleotide sequence of salmon testes and salmon liver cytoplasmic initiator tRNA. *Cell* 6(3):395–405.
43. Piper PW, Clark BF (1974) Primary structure of a mouse myeloma cell initiator transfer RNA. *Nature* 247(442):516–518.
44. Simsek M, RajBhandary UL (1972) The primary structure of yeast initiator transfer ribonucleic acid. *Biochem Biophys Res Commun* 49(2):508–515.
45. Simsek M, RajBhandary UL, Boisnard M, Petrissant G (1974) Nucleotide sequence of rabbit liver and sheep mammary gland cytoplasmic initiatory transfer RNAs. *Nature* 247(442):518–520.
46. Woo NH, Roe BA, Rich A (1980) Three-dimensional structure of *Escherichia coli* initiator $tRNA^{fMet}$. *Nature* 286(5771):346–351.
47. Wakao H, et al. (1989) The solution structure of the *Escherichia coli* initiator tRNA and its interactions with initiation factor 2 and the ribosomal 30 S subunit. *J Biol Chem* 264(34):20363–20371.
48. Clemons WM, Jr., et al. (2001) Crystal structure of the 30 S ribosomal subunit from *Thermus thermophilus*: Purification, crystallization and structure determination. *J Mol Biol* 310(4):827–843.
49. Demeshkina N, Jenner L, Westhof E, Yusupov M, Yusupova G (2012) A new understanding of the decoding principle on the ribosome. *Nature* 484(7393):256–259.
50. Selmer M, et al. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 313(5795):1935–1942.
51. Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst* 26(6):795–800.
52. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
53. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.
54. DeLano WL (2006) *Pymol* (Delano Scientific, San Carlos, CA).
55. Moriarty NW, Grosse-Kunstleve RW, Adams PD (2009) Electronic Ligand Builder and Optimization Workbench (eLBOW): A tool for ligand coordinate and restraint generation. *Acta Crystallogr D Biol Crystallogr* 65(Pt 10):1074–1080.
56. Frisch AE, Dennington RD, Keith TA, Neilsen AB, Holder AJ (2003) *GaussView Version 3.0* (Gaussian, Pittsburgh, PA).
57. Frisch MJ, et al. (2003) *Gaussian 03, Revision C.01* (Gaussian, Pittsburgh, PA).
58. Cossi M, Rega N, Scalmani G, Barone V (2003) Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J Comput Chem* 24(6):669–681.