Mycobacterium tuberculosis Phe-tRNA synthetase

Michalska, Karolina; Jedrzejczak, Robert; Wower, Jacek; Chang, Changsoo; Baragaña, Beatriz; Gilbert, Ian H.

Published in:
Nucleic Acids Research

DOI:
10.1093/nar/gkab272

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Nucleic Acids Research, 2021 **1**
doi: 10.1093/nar/gkab272

Mycobacterium tuberculosis Phe-tRNA synthetase: structural insights into tRNA recognition and aminoacylation

Karolina Michalska[1,2], Robert Jedrzejczak[1,2], Jacek Wower[3], Changsoo Chang[1,2], Beatriz Baragaña[4], Ian H. Gilbert[4], Barbara Forte[4] and Andrzej Joachimiak[1,2,5,*]

[1]Center for Structural Genomics of Infectious Diseases, Consortium for Advanced Science and Engineering, University of Chicago, Chicago, IL 60667, USA, [2]Structural Biology Center, X-ray Science Division, Argonne National Laboratory, Lemont, IL 60439, USA, [3]Department of Animal Sciences, Auburn University, Auburn, AL 36849, USA, [4]Drug Discovery Unit, Wellcome Centre for Anti-Infectives Research, Division of Biological Chemistry and Drug Discovery, University of Dundee, Dundee DD1 5EH, UK and [5]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60367, USA

Received December 23, 2020; Revised March 30, 2021; Editorial Decision March 31, 2021; Accepted April 19, 2021

ABSTRACT

Tuberculosis, caused by *Mycobacterium tuberculosis*, responsible for ∼1.5 million fatalities in 2018, is the deadliest infectious disease. Global spread of multidrug resistant strains is a public health threat, requiring new treatments. Aminoacyl-tRNA synthetases are plausible candidates as potential drug targets, because they play an essential role in translating the DNA code into protein sequence by attaching a specific amino acid to their cognate tRNAs. We report structures of *M. tuberculosis* Phe-tRNA synthetase complexed with an unmodified $tRNA^{Phe}$ transcript and either L-Phe or a non-hydrolyzable phenylalanine adenylate analog. High-resolution models reveal details of two modes of tRNA interaction with the enzyme: an initial recognition via indirect readout of anticodon stem-loop and aminoacylation ready state involving interactions of the 3′ end of $tRNA^{Phe}$ with the adenylate site. For the first time, we observe the protein gate controlling access to the active site and detailed geometry of the acyl donor and tRNA acceptor consistent with accepted mechanism. We biochemically validated the inhibitory potency of the adenylate analog and provide the most complete view of the Phe-tRNA synthetase/$tRNA^{Phe}$ system to date. The presented topography of amino adenylate-binding and editing sites at different stages of tRNA binding to the enzyme provide insights for the rational design of anti-tuberculosis drugs.

INTRODUCTION

With ∼1.5 million people killed by tuberculosis (TB) in 2018, *Mycobacterium tuberculosis* remains the deadliest pathogen worldwide (1) and the leading cause of morbidity and mortality among HIV-infected patients. It is estimated that one third of the world's population is at risk of developing active TB disease, especially in low- and middle-income countries. Extensively drug-resistant TB (XDR-TB) and multidrug-resistant TB (MDR-TB) constitute a growing global public health threat (2). The existing TB multidrug treatments are long, particularly for XDR-TB and MDR-TB infections. Discovery of drugs with novel mechanisms of action is critical for defeating the existing drug resistant strains and shortening the TB treatment (3).

Lynezolid, an FDA-approved antibiotic, and several other anti-TB compounds presently tested in clinical trials, target protein translation, an essential process in all cellular organisms (4,5). One of the crucial steps in this process is carried out by aminoacyl-tRNA synthetases (aaRS) that catalyze the attachment of a specific amino acid to the cognate tRNA (6,7). Some aaRSs are equipped with proofreading mechanisms to ensure correct pairing of amino acid and tRNA. The synthesis of aminoacyl-tRNA is accomplished by a two-step reaction. Formation of the aminoacyl-adenylate from an amino acid and ATP, sometimes carried out in a tRNA-dependent manner (8), is followed by ligation of the aminoacyl group to the 3′-terminal adenosine of tRNA.

FRS, which is specific for L-phenylalanine and $tRNA^{Phe}$, has been studied extensively biochemically, structurally and computationally for the past 50 years (9–12). Several FRS structures at moderate resolutions are available in the Protein Data Bank (PDB) (Table 1). All of them belong to the

*To whom correspondence should be addressed. Tel: +1 630 252 3926; Fax: +1 630 252 6126; Email: andrzejj@anl.gov
Present address: Andrzej Joachimiak, Structural Biology Center, X-ray Science Division, Argonne National Laboratory, Lemont, IL 60439, USA.

class IIC aaRS and show similarity to class IIA SRS as well as more distant homology to ancient GRS and ARS class IID (6,7). Typically, class II aaRSs aminoacylate tRNA on 3′OH of ribose, but FRS does so on 2′OH. The bacterial, archaeal and human cytoplasmic enzymes consist of α (PheS) and β subunits (PheT), that via domain swapping, form highly intertwined $(\alpha\beta)_2$ heterotetramers (11,13–17). The α subunit recognizes L-Phe and 3′ terminal adenosine of tRNA and carries out key catalytic activities (18). The β subunit recognizes $tRNA^{Phe}$ and provides an editing function assuring that only the L-Phe is attached to $tRNA^{Phe}$ (19).

FRS, an essential component of the protein synthesis apparatus, has been recognized as a high-priority antimicrobial drug target (19,20). Phenyl-thiazolylurea-sulfonamides are potent inhibitors of both Gram-negative and Gram-positive bacterial FRS orthologs (9). Ethanolamine derivatives inhibit FRS from *Staphylococcus aureus* (21), gossypol blocks *Pseudomonas aeruginosa* and *Streptococcus pneumoniae* (22). Some natural products are able to act on *M. tuberculosis* FRS (*Mt*FRS) (23). However, no FRS inhibitor has reached advanced therapeutic development stages.

Here, to address a major knowledge gap in *M. tuberculosis* protein synthesis, we present four crystal structures of *Mt*FRS in complex with a full-length unmodified *M. tuberculosis* $tRNA^{Phe}$ transcript (abbreviated as *Mt*$tRNA^{Phe}$ throughout the text). For the first time we show for FRS that the tRNA binding and recognition are occurring in two distinct stages. The 'initial tRNA recognition' state was determined either with bound L-Phe (in two crystal forms), or with an inhibitor – a nonhydrolyzable phenylalanine adenylate analog 5′-*O*-(*N*-phenylalanyl)sulfamoyl-adenosine (F-AMS). The 'aminoacylation ready' state, in which tRNA is predisposed for accepting an activated amino acid, was determined in a complex with F-AMS. These structures offer the most complete picture of the FRS-$tRNA^{Phe}$ complex to date and combined with biochemical characterization of the enzyme and F-AMS binding, provide insights that are essential for rational structure-based drug design of selective inhibitors of *Mt*FRS.

MATERIAL AND METHODS

Cloning and expression of *M. tuberculosis* FRS subunits, PheS and PheT

The *pheS* and *pheT* genes were amplified from genomic DNA of *M. tuberculosis* H37Rv using the following pairs of primers: 5′-TACTTCCAATCCAATGCCATGTTGT CGCCGGAGGCATTGA, TTATCCACTTCCAATG TTAGGCACCCACCCCGAACGG and 5′-TACTTCC AATCCAATGCCATGCGGCTACCCTACAGCTGG, TTATCCACTTCCAATGTTAGCCACGCAGCACGGC ACC. PCR products were purified and treated with T4 polymerase in presence of dCTP according to Kim *et al.* (24). *PheS* gene was cloned to pMCSG53, while *pheT* was cloned to pMCSG120. Both vectors contain identical cloning sites. pMCSG53 carries ampicillin resistance and ColE1 origin of replication. pMCSG120 retains resistance to kanamycin and has RSF origin of replication. In the final construct the PheS sequence is amended on N-terminus with His_6 affinity tag and Tobacco Etch Virus (TEV) protease cleavage site.

Escherichia coli BL21-Gold (DE3) was co-transformed with plasmids PheS-pMCSG53 and PheT-pMCSG120 and grown overnight in LB medium supplemented with 40 mM K_2HPO_4 at 37°C/220 rpm with selection against ampicillin (150 μg/ml) and kanamycin (100 μg/ml). The following morning the culture was diluted 1:100 into LB media with the same concentration of antibiotics and grown at 37°C/190 rpm till OD_{600} reached about 1.0 followed by cooling of media to about 20°C. At this point the media were supplemented with 0.25 mM IPTG, 40 mM K_2HPO_4 and 0.5% glucose. The culture was grown for another 20 h at 18°C/190 rpm followed by cells harvest and resuspension in buffer A (50 mM HEPES/NaOH, pH 8.0, 0.5 M NaCl, 20 mM imidazole, pH 8.0, 5% glycerol, 1 mM TCEP). The cell suspension was frozen and stored at −80°C.

Synthesis of a full-length unmodified *M. tuberculosis* $tRNA^{Phe}$ transcript (GAA)

The isoacceptor $tRNA^{Phe}$(GAA) recognizing specifically UUC codon and its truncated derivatives lacking the 3′-terminal adenosines were prepared as described previously by Sherlin *et al.* (25). Oligonucleotides used for construction of the $tRNA^{Phe}$ gene are: 5′-AATTCCTGCAG**TAATACGACTCACTATA**GGCC AGGTAGCTCAGTCGGTATGAGC-3′, 5′-mUmGGTG GCCAGGGGCGGGATCGAACCGCCGACCTTCCG CTTTTCAGGC-GGACGCTCATACCGAC-3′ and 5′-m GmGTGGCCAGGGGCGGGATCGAACCGCCGA-C CTTCCGCTTTTCAGGCGGACGCTCATACCGAC-3′. mU and mG represent the 2′-O-methyl nucleotides used to reduce the proportion of runover transcripts and increase their yields. The underlined portions represent the overlapped segment and bold type indicates the T7 RNA polymerase promoter. To synthesize these transcripts, 10 μg/ml duplex DNA template in a solution containing 200 mM HEPES/KOH (pH 7.5), 30 mM $MgCl_2$, 2 mM spermidine, 40 mM DTT, 6 mM each ATP, CTP, GTP and UTP was incubated with 100 μg/ml T7 RNA polymerase, 5 U/ml inorganic pyrophosphatase, 50 U/ml SUPERase•In RNase Inhibitor (Invitrogen) in an 8 ml reaction for 5 h at 37°C. The procedure yields full length *Mt*$tRNA^{Phe}$ transcript. DNA template was removed by addition of 50 U/ml RQ1 RNase-free DNase (Promega) and incubation for 1 h at 37°C, followed by phenol-chloroform extraction. The tRNA was recovered from the aqueous phase by ethanol precipitation and then fractionated on DE-52 (Whatman) column as described before (25). Integrity of the purified tRNA was confirmed by electrophoresis on a 10% denaturing polyacrylamide gel. The tRNA was dialyzed into 20 mM HEPES/KOH (pH 7.5), 150 mM KCl and 5 mM $MgCl_2$ buffer prior to further use. The fraction of tRNA capable of being aminoacylated by *Mt*FRS was determined by assays under conditions giving maximal amino acid incorporation (approximately ten-fold molar excess of enzyme over tRNA).

Table 1. FRS structures in the Protein Data Bank

PDB ID	Source	Structure	Res (Å)	% Sequence identity to *Mt* (α/β)[a]	reference
1B70	*T. thermophilus*	$(\alpha/\beta)_2$ complex with L-Phe in α	2.70	39/34	Reshetnikova *et al.* (18)
1B7Y	*T. thermophilus*	$(\alpha/\beta)_2$ complex with adenosine-5′-[phenylalaninol-phosphate] in α	2.50		Reshetnikova *et al.* (18)
1EIY	*T. thermophilus*	$(\alpha/\beta)_2$ complex with $tRNA^{Phe}$	3.30		Goldur *et al.* (55)
1JJC	*T. thermophilus*	$(\alpha/\beta)_2$ complex with adenosine-5′-[phenylalaninyl-phosphate] in α	2.60		Fishman *et al.* (66)
1PYS	*T. thermophilus*	$(\alpha/\beta)_2$ complex	2.90		Mosyak *et al.* (13)
2AKW	*T. thermophilus*	$(\alpha/\beta)_2$ complex with p-Cl-Phe in α	2.80		Kotik-Kogan *et al.* (19)
2ALY	*T. thermophilus*	$(\alpha/\beta)_2$ complex with 5′-*O*-[*N*-L-tyrosyl)sulfamoyl]adenosine in α	2.60		Kotik-Kogan *et al.* (19)
2AMC	*T. thermophilus*	$(\alpha/\beta)_2$ complex with L-Tyr in β	2.70		Kotik-Kogan *et al.* (19)
2IY5	*T. thermophilus*	$(\alpha/\beta)_2$ complex with $tRNA^{Phe}$ and adenosine-5′-[phenylalaninol-phosphate] in α	3.10		Moor *et al.* (67)
3HFZ	*T. thermophilus*	$(\alpha/\beta)_2$ complex with m-L-Tyr in α	2.90		Klipcan *et al.* (68)
3TEH	*T. thermophilus*	$(\alpha/\beta)_2$ complex with 3,4-dihydroxyphenylalanine in α	2.85		Moor *et al.* (69)
4TVA	*T. thermophilus*	$(\alpha/\beta)_2$ complex with puromycin in β and L-Phe in α	2.60		Tworowski *et al.* (70)
2RHQ	*S. haemolyticus*	$(\alpha/\beta)_2$ complex with 1-{3-[(4-pyridin-2-ylpiperazin-1-yl)sulfonyl]phenyl}-3-(1,3-thiazol-2-yl)urea in α, 6 mutations in β	2.20	41/31	Evdokimov *et al.* (15)
2RHS	*S. haemolyticus*	$(\alpha/\beta)_2$ complex with 1-{3-[(4-pyridin-2-ylpiperazin-1-yl)sulfonyl]phenyl}-3-(1,3-thiazol-2-yl)urea in α, 5 mutations in β	2.20		Evdokimov *et al.* (15)
3PCO	*E. coli*	$(\alpha/\beta)_2$ complex with L-Phe and AMP in α	3.02	45/31	Mermershtain *et al.* (17)
4P71	*P. aeruginosa*	$(\alpha/\beta)_2$ complex	2.79	43/32	Abibi *et al.* (71)
4P72	*P. aeruginosa*	$(\alpha/\beta)_2$ complex with compound 2-{3-[(4-chloropyridin-2-yl)amino]phenoxy}-N-methylacetamide in α	2.62		Abibi *et al.* (71)
4P73	*P. aeruginosa*	$(\alpha/\beta)_2$ complex with compound 1-{3-[(4-pyridin-2-ylpiperazin-1-yl)sulfonyl]phenyl}-3-(1,3-thiazol-2-yl)urea in α	3.03		Abibi *et al.* (71)
4P74	*P. aeruginosa*	$(\alpha/\beta)_2$ complex with compound N-[(3S)-1, 1-dioxidotetrahydrothiophen-3-yl]-2-[(4-methylphenoxy) methyl]-1,3-thiazole-4-carboxamide in α	2.70		Abibi *et al.* (71)
4P75	*P. aeruginosa*	$(\alpha/\beta)_2$ complex with compound 3-(3-methoxyphenyl)-5-(trifluoromethyl)-1H-pyrazole in α	2.96		Abibi *et al.* (71)
3ICA	*P. gingivalis*	β subunit fragment	2.44	36/27	Fan *et al.*, unpublished
3IG2	*B. fragilis*	β subunit fragment	2.09	38/28	Stein *et al.*, unpublished
2CXI	*P. horikoshii*	β subunit fragment, 9 mutations	1.94	19/13	Sasaki *et al.* (14)
3CMQ	*H. sapiens*	Mitochondrial monomer with adenosine-5′-[phenylalaninyl-phosphate]	2.20	27, 28[b]	Klipcan *et al.* (72)
3HFV	*H. sapiens*	Mitochondrial monomer with m-L-Tyr	2.60		Klipcan *et al.* (68)
3TEG	*H. sapiens*	Mitochondrial monomer with 3,4-dihydroxyphenylalanine	2.20		Moor *et al.* (69)
3TUP	*H. sapiens*	Mitochondrial monomer with *Tt*$tRNA^{Phe}$	3.05		Klipcan *et al.* (11)
5MGH	*H. sapiens*	Mitochondrial monomer, point mutant with L-Phe	1.87		Kartvelishvili *et al.* (73)
5MGU	*H. sapiens*	Mitochondrial monomer, point mutant with L-Phe	1.89		Kartvelishvili *et al.* (73)
5MGW	*H. sapiens*	Mitochondrial monomer, point. mutant with L-Phe	1.46		Kartvelishvili *et al.* (73)
5MGV	*H. sapiens*	Mitochondrial monomer, point mutant	2.05		Kartvelishvili *et al.* (73)
3L4G	*H. sapiens*	Cytosolic $(\alpha/\beta)_2$ complex with L-Phe in α	3.30	34/18	Finarov *et al.* (16)

[a]As determined by Emboss Needle (74).
[b]The numbers correspond to *Mt*FRS α subunit residues 108–341 aligned with human mitochondrial FRS residues 82–347 and *Mt*FRS β subunit residues 739–831 with 360–451).

FRS purification and crystallization

Frozen cells were thawed and sonicated (5 min total time, 130 W power output) and spun at 30 000 × g at 4°C for 1 h. The initial Ni^{2+} affinity purification step was performed using a 2.5-cm × 10-cm Flex-Column connected to a Van-Man vacuum manifold (Promega) (26). Briefly: supernatant was loaded on 3 ml Ni^{2+} Sepharose (GE Healthcare Life Sciences) equilibrated with buffer A and mixed thoroughly with the resin. Vacuum of 15 psi was used to speed removal of supernatant as well as wash out of unbound proteins (150 ml buffer A supplemented to 50 mM imidazole). The FRS complex was eluted with 15 ml buffer A supplemented with 500 mM imidazole, pH 8.0. The eluate was concentrated to about 2 ml and loaded on a Superdex 200 16/70 size exclusion column (GE Healthcare Life Sciences) equilibrated with buffer A. Fractions containing the FRS complex were collected, and buffer A was replaced with crystallization buffer (20 mM HEPES/NaOH, pH 8.0, 150 mM KCl) on Amicon 30-kDa-cutoff concentra-

tors (Millipore) followed by overnight incubation at 4°C with *Mt*tRNA^{Phe}. The molar ratio of tRNA^{Phe} to FRS heterotetramer $(\alpha\beta)_2$ was 2:1. Protein-tRNA complex was concentrated to ∼20 mg/ml. The ligands were added at molar ratios 4:1. Crystallization experiments were performed at 16°C using the sitting-drop vapor-diffusion method in 96-well CrystalQuick plates (Greiner Bio-One, Monroe, NC) with the help of the Mosquito liquid dispenser (TTP LabTech, Cambridge, MA, USA). Crystallizations trials were performed using protein-tRNA to matrix ratio 1:1 (800 nl drops). Crystallization and cryocooling conditions are given in Table 2.

X-ray data collection, structure determination, refinement and analysis

Prior to data collection at 100 K, all cryoprotected crystals were flash-cooled in liquid nitrogen. The X-ray diffraction experiments were carried out at the Structural Biology Center 19-ID beamline at the Advanced Photon Source, Argonne National Laboratory. The diffraction images were recorded on the PILATUS3 × 6M detector. The data sets were processed with the HKL3000 suite (27). Intensities were converted to structure factor amplitudes in the Ctruncate program (28,29) from the CCP4 package (30). The data collection and processing statistics are given in Table 2.

The initial FRS structure was solved by molecular replacement using monomeric units of *E. coli* FRS as a search template (PDB id 3PCO) against the dataset from low resolution apo crystals of *Mt*FRS. The FRSt/F1 structure was determined using the partially built apo model. The search of tRNA molecule from *T. thermophilus* FRS complex structure (PDB id 1EIY) failed so the entire nucleic acid molecule has been built manually from scratch. The subsequent structures were determined using the FRSt/F1 model. The structures were refined by manual corrections in Coot (31) and crystallographic refinement in Refmac (32) and Phenix (33). The electron density in composite omit map for protein, tRNA, L-Phe, F-AMS is excellent (Supplementary Figure S1). The refinement statistics are given in Table 2. The structures were analyzed using Chimera (34), SNAP (for protein-RNA interactions, snap.x3DNA.org), and the RNA-align server (for RNA superposition) (35). Some of the atoms that appeared to be cations could not be confidently identified and have been modeled as unknown atoms UNX. The atomic coordinates and structure factors have been deposited in the PDB under accession codes 7KA0, 7KAB, 7K9M, 7K98.

Aminoacylation of *Mt*tRNA^{Phe}

Purified *M. tuberculosis* tRNA transcripts lacking the 3′-terminal adenosine were ^{32}P-labeled with yeast ATP/CTP tRNA nucleotidyl transferase in the presence of [α-^{32}P]ATP as described in Wower *et al.* (36). Aminoacylations of the 3′-^{32}P-labeled *Mt*tRNA^{Phe} were carried out in 50 μl of 50 mM HEPES/KOH, pH 7.5, 20 mM KCl, 10 mM $MgCl_2$, 100 μg/ml BSA, 5 mM DTT, 4 mM ATP, 0.2 mM L-Phe and 50 nM tRNA transcript. Reactions were initiated by addition of a purified *M. tuberculosis* phenylalanyl-tRNA synthetase to final concentration of 0.7 μM and incubated for 30 min at 37°C. The extent of aminoacylation was analyzed by gel-shift electrophoresis according to Varshney *et al.* (37). Prior to electrophoresis, on a 10% acid/7 M urea polyacrylamide gel, 10 pmoles of uncharged and aminoacylated tRNA (Phe-tRNA^{Phe}) were completely digested with 2 U of RNase T1 (Worthington) in 10 mM sodium acetate (pH 4.5). The amount of radioactivity in the bands corresponding to uncharged and aminoacylated CCACCA oligonucleotides was quantitated by phosphorimaging and/or liquid scintillation counting of the excised strips of the polyacrylamide gel.

The value of IC_{50} was determined at 1.5 μM concentration of phenylalanine over a range of inhibitor concentration from zero to 200 nM. According to the Cheng-Prusoff equation, IC_{50} values approximate K_i when substrate concentration used in the assay is lower than K_m (38).

Aminoacylation of *E. coli* and yeast tRNA^{Phe}

Aminoacylation reactions were performed as described above with the following modifications: 3′-^{32}P-labeled *Mt*tRNA^{Phe} were replaced either by non-radioactive *Mt*tRNA^{Phe} or fully modified *E. coli* and yeast tRNA^{Phe} preparations from Sigma. [α-^{32}P]ATP from PerkinElmer was used instead of non-radioactive ATP. All reactions were quenched in 400 mM sodium acetate (pH 5.2) with 0.1% SDS. Aliquots were spotted on prewashed polyethyleneimine-cellulose TLC plates (Macherey and Nagel) and developed in 100 mM ammonium acetate, 5% (v/v) acetic acid.

*Mt*FRS inhibition by F-AMS

Aminoacylation assays were carried out by a very sensitive assay that has been described by Bullock *et al.* (39). The assay was performed by 10 min incubation of L-Phe with purified enzyme in 50 mM HEPES/KOH (pH 7.5), 20 mM KCl, 10 mM $MgCl_2$, 4 mM ATP/20 μM [α-^{32}P]ATP, 0.1% dimethyl sulfoxide in a final volume of 20 μl at 37°C. K_m for L-Phe and K_i were determined by varying concentrations of amino acid and inhibitor from 0.5 to 25 mM and from 15 to 125 nM, respectively. The concentration of FRS was 6 nM. The reactions were stopped by adding 2.0 μl of the reaction mixtures to 3.0 μl of a solution containing 400 mM sodium acetate (pH 5.2) and 0.1% SDS. 2.0 μl aliquots of this mixture were spotted on dry polyethyleneimine (PEI) cellulose plates that have been prewashed in water. Phe-[^{32}P]AMP were separated from [α-^{32}P]ATP using 0.1 M ammonium acetate/5% acetic acid as developing solution, visualized by autoradiography and quantified by phosphorimaging analysis.

Synthesis of phenylalanine adenylate analog

5′-*O*-[*N*-(phenylalanine)sulfamoyl] adenosine, Phe-AMS, was been prepared in six steps as previously described (40).

Table 2. Data processing and refinement statistics

	Data processing			
Structure	FRSt/F1	FRSt/F2	FRSt/F-AMS1	FRSt/F-AMS2
Crystallization conditions, cryoprotectant supplement	0.2 M Li_2SO_4, 0.1 M HEPES pH 7.5, 25% PEG3350; 10% glycerol	14.3% PEG20K, 150 mM TAPS pH 8.5, 5% 1,2-propanediol; 20% glycerol	0.2 M Na/K tartrate, 20% PEG3350; 15% glycerol	30% PEG400, 0.2 M $MgCl_2$, 0.1 M HEPES pH 7.5; No cryoprotectant
Wavelength (Å)	0.9793	0.9793	0.9793	0.9793
Resolution range (Å)[a]	2.40–50.00 (2.40–2.44)	2.50–30.00 (2.50–2.54)	2.50–50.0 (2.50–2.54)	2.20–30.0 (2.20–2.24)
Space group	*C*2	*C*2	*C*2	$P2_1$
Unit cell (Å, °)	$a = 292$, $b = 110$, $c = 128$, $\beta = 100$	$a = 127$, $b = 110$, $c = 148$, $\beta = 105$	$a = 128$, $b = 110$, $c = 147$, $\beta = 103$	$a = 147$, $b = 65$, $c = 193$, $\beta = 109$
Unique reflections (merged)	155 793 (7677)	68 179 (3404)	68 621 (3274)	173 141 (8263)
Multiplicity	6.4 (5.5)	7.0 (5.9)	6.2 (3.9)	6.4 (6.0)
Completeness (%)	98.9 (98.2)	99.3 (99.4)	99.0 (96.2)	96.9 (93.6)
Mean I/sigma(I)	14.2 (1.78)	17.9 (1.10)	13.8 (1.60)	21.5 (1.56)
Wilson *B*-factor (Å^2)	32.4	52.5	31.9	45.8
R-merge[b]	0.147 (1.207)	0.141 (1.374)	0.171 (0.832)	0.094 (1.223)
CC1/2[c]	0.992 (0.673)	(0.487)	0.990 (0.522)	(0.596)
Refinement				
Resolution range (Å)	2.40–48.00	2.50–29.69	2.50–47.92	2.19–29.85
Reflections work/test	135 987/7094	66 452/1665	63 320/3264	169 565/3498
R_{work}/R_{free}[d]	0.176/0.215	0.187/0.213	0.179/0.215	0.178/0.224
Number of non-hydrogen atoms	21 651	10 440	10 722	22 017
Macromolecules	20 563	10 237	10 292	21 241
Ligands/solvent	184/904	53/150	108/322	101/675
Amino acid residues	339+835+336+835	336+835	339+834	343+834+343+836
Nucleic acid bases	66+66	66	66	77+77
RMSD(bonds) (Å)	0.003	0.002	0.002	0.007
RMSD(angles) (°)	0.599	0.478	0.504	0.915
Ramachandran favored[e] (%)	96.65	96.39	96.83	96.85
Ramachandran allowed (%)	3.09	3.26	3.08	3.02
Ramachandran outliers (%)	0.26	0.34	0.09	0.13
Rotamer outliers (%)	2.03	1.88	0.98	2.79
Clashscore	3.45	3.93	3.04	2.70
Average *B*-factor (Å^2)	65.58	99.46	56.72	61.11
macromolecules	66.51	100.05	57.36	61.46
ligands	67.43	92.47	54.38	49.86
solvent	44.14	61.43	36.99	51.67
Number of TLS groups	39	16	11	34
PDB ID	7KA0	7KAB	7K9M	7K98

[a]Values in parentheses correspond to the highest resolution shell.
[b]$R_{merge} = \Sigma h \Sigma j |Ihj - <Ih>| / \Sigma h \Sigma j Ihj$, where *Ihj* is the intensity of observation *j* of reflection *h*.
[c]As defined by Karplus and Diederichs (75).
[d]$R = \Sigma h ||Fo| - |Fc|| / \Sigma h |Fo|$ for all reflections, where *F*o and *F*c are observed and calculated structure factors, respectively. R_{free} is calculated analogously for the test reflections, randomly selected and excluded from the refinement.
[e]As defined by Molprobity (76).

RESULTS AND DISCUSSION

tRNA^{Phe} and F-AMS binding

To investigate structure-function relationships in *Mt*FRS, we have co-expressed PheS (α) and PheT (β) subunits of *Mt*FRS and purified the functional heterotetrameric complex. In addition, we have prepared *Mt*tRNA^{Phe} specific for the 5′UUC codon. This *Mt*tRNA^{Phe} lacks post-transcriptional modifications. The genus-specific modifications are unknown but in other organisms tRNA^{Phe} has been shown to carry varying sets of m1A, ms2i6A, m1C, m3C, m5C, m1G, m2G, m3G, Ψ, acp3U, D, T and wybutosine (41,42). The majority of these modified nucleotides have little effect on aminoacylation by aaRSs but are impor-

tant for codon – anticodon interactions, translational rates, reading frame maintenance, co-translational protein folding dynamics and other cellular processes (43,44).

To confirm enzymatic activity of the protein, we have independently tested both L-Phe adenylation and tRNA aminoacylation in the presence of three substrates, *Mt*tRNA^{Phe} as well as *E. coli* and yeast tRNAs^{Phe}. *Mt*tRNA^{Phe} can be almost completely (>99%) charged by *Mt*FRS with L-Phe (Figure 1). *Mt*FRS can also aminoacylate mature tRNA^{Phe} molecules isolated from *E. coli* and yeasts (Figure 1). *Mt*FRS, like all known FRS, is able to activate its amino acid substrate in the absence of tRNA (Figure 1A–C), though the addition of tRNA stimulates this reaction.

Aminoacyl-sulfamoyl adenosine compounds have been previously reported as inhibitors of bacterial aaRSs (45–47). Here, we show strong inhibition of *Mt*FRS by the non-hydrolyzable phenylalanine adenylate analog, 5′-*O*-[*N*-(phenylalanine)sulfamoyl] adenosine (F-AMS). Earlier studies have demonstrated that aminoacyl-sulfamoyl-adenosines efficiently inhibit *E. coli* ERS (K_i 2.8 nM) (45), murine liver ERS (K_i ∼ 70 nM) (45)), *E. coli* QRS (K_i 1.3 mM), *S. aureus* IRS (IC_{50} 4 nM) (47), RRS (IC_{50} 4.5 nM), HRS (IC_{50} 130 nM) and TRS (IC_{50} 15 nM) (46). Consistent with the earlier studies, F-AMS acts as a competitive inhibitor with a K_i of 33 nM and IC_{50} value of ∼36 nM (Figure 2).

Overall structure of *Mt*FRS

To understand structural basis of aminoacylation reaction and tRNA recognition, we have crystallized *Mt*FRS with *Mt*tRNA^{Phe} and either L-Phe or F-AMS. Four structures have been determined - two representing different crystal forms of the *Mt*FRS-*Mt*tRNA^{Phe}/L-Phe complex (*Mt*FRSt/F1, and *Mt*FRSt/F2) and two corresponding to different forms of the *Mt*FRS-*Mt*tRNA^{Phe}/F-AMS complex (*Mt*FRSt/F-AMS1 and *Mt*FRSt/F-AMS2) (Figure 3, Table 2). The former three structures illustrate the initial tRNA recognition state, where the enzyme binds primarily to the anticodon region, with only partial contacts with D- and T-loops of tRNA. In these structures the electron density for the acceptor end of tRNA is poorly defined and cannot be modeled. The *Mt*FRSt/F-AMS2 structure corresponds to the tRNA aminoacylation ready state, where the complex of *Mt*FRS with tRNA^{Phe} is predisposed for acceptance of the activated amino acid. *Mt*FRSt/F2 and *Mt*FRSt/F-AMS1 are isomorphous and contain a single αβ unit in the asymmetric unit (ASU), while *Mt*FRSt/F1 and *Mt*FRSt/F-AMS2 have an entire $(\alpha\beta)_2$ heterotetramer in the ASU of different unit cells. The crystal lattice or the tRNA binding status do not seem to affect the protein interactions with L-Phe, F-AMS and tRNA ligands.

Our findings are consistent with earlier observations, which indicate that the presence of adenylate improves order of the CCA3′ end in crystal structures of *Tt*SRS/tRNA^{Ser} complexes and the binding specificity of *E. coli* tRNA^{His} to cognate *Ec*HRS is increased when adenylate analog 5′-*O*-[*N*-(*L*-histidyl)sulfamoyl]-adenosine is bound to the active site (48,49). Interestingly, the *Mt*FRSt/F1 complex is reminiscent of other aaRS structures where the tRNA anticodon is properly recognized, but the acceptor stem remains disordered, in particular KRS–tRNA^{Lys} complex (50) and the heterologous complex composed of yeast tRNA^{Asp} and *E. coli* DRS (51).

The structures of enzyme with specific *Mt*tRNA^{Phe} show the previously described tetrameric assembly (Figure 3). The electron density for protein, tRNA, L-Phe, F-AMS is excellent and with a few exceptions, as discussed earlier and below, the entire structure can be modeled. The $(\alpha\beta)_2$ assembly has a complex shape that can fit into a rectangular prism with overall dimensions 70 Å × 110 Å × 170 Å. The individual subunits are composed of several domains, which structurally and functionally make sense only when assembled into highly intertwined heterotetramer of this intricate molecular machine. Each tRNA molecule interacts with all four polypeptide chains.

Our high-resolution structures allow us to improve domain assignment and their boundaries. The smaller α subunit (modeled residues 3–341) is very extended (∼130 Å) and has two distinct domains, α1 and α2, connected by a 14-residue linker (Figure 4, Supplementary Figure S3). The N-terminal all helical α1 domain binds the tRNA D-loop region and the C-terminal α2 domain performs aminoacylation. The larger β subunit (modeled residues 1–831) can be subdivided into three globular domains β1–3, two of which have subdomains (Figure 4, Supplementary Figure S3). β1 has four distinct subdomains: β1s1, β1s2 and β1s3 corresponding to the editing domain that shares fold with α2 catalytic domain (CATH domain 3.30.930.10), and β1s4. The latter, a small α/β subdomain together with β1s1, interacts with the α2 aminoacylation domain via direct protein-protein contacts and a conserved magnesium ion. The β1s1, β1s3 and β1s4 subdomains form a core of β1 and β1s2 is on its periphery. β1s4 is followed by a 13-residue linker connecting to the β2 domain, which consists of two subdomains – larger β2s1 and smaller β2s2. β2 interacts extensively with aminoacylation domain α2 (surface area ∼2150 Å^2) contributing significantly to the complex interface. β2, via a 15-residue linker, leads to the C-terminal domain β3 that extends out of the main body of the protein and is responsible for tRNA anticodon binding.

*Mt*tRNA^{Phe} structure

The complete *Mt*tRNA^{Phe} molecules are modeled only in the *Mt*FRSt/F-AMS2 complex, where the only atoms missing from the model correspond to the linkage between U21 and G22. Notably, this structure has been determined at high magnesium concentration and several Mg^{2+} ions have been identified—some of them are highly hydrated and bind exclusively to the nucleic acid via electrostatic interactions rather than direct metal ion coordination. The complete structure of *Mt*tRNA^{Phe} is similar to the partial structure of the *Mt*tRNA^{Phe} in the initial recognition structure (see below) and other structures: the mature yeast equivalent carrying the Y base in the anticodon region (PDB id 4TNA), unmodified *T. thermophilus* tRNA^{Phe} transcript in complex with FRS (PDB id 2IY5) and unmodified tRNA^{Phe} transcript from *E. coli* (PDB id 3L0U). Sequence identity of these tRNAs varies between 45 and 60% and their superposition indicates that root mean square deviations (RMSD)

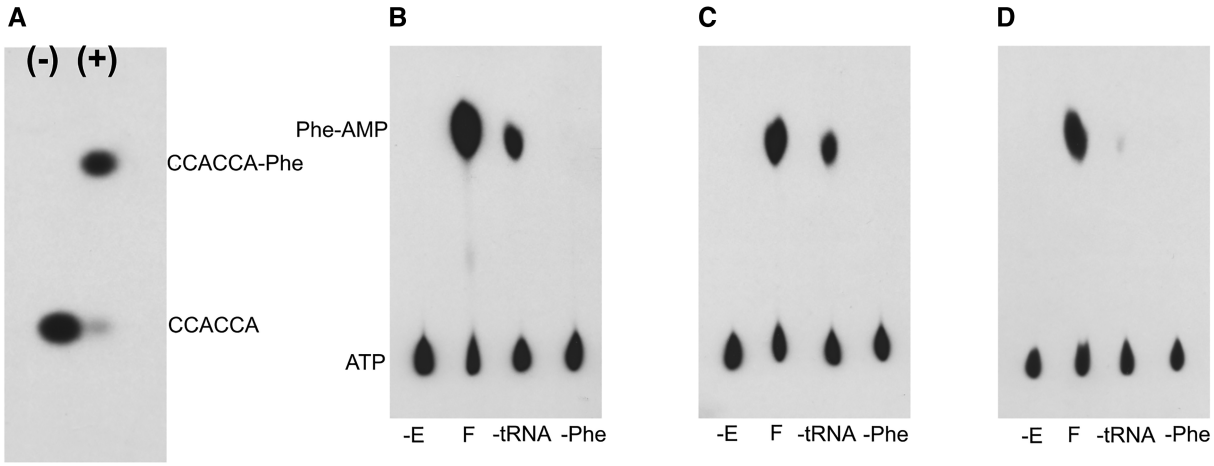

Figure 1. *Mt*FRS activity. (**A**) Aminoacylation of unmodified *M. tuberculosis* $tRNA^{Phe}$. The 3′-^{32}P-labeled tRNA transcript was charged with L-Phe in the presence of purified *Mt*FRS. The extent of aminoacylation was analyzed according to Varshney *et al.* (37). Prior to polyacrylamide gel electrophoresis, tRNA and Phe-tRNA were digested with RNase T1 to yield 3′-^{32}P-labeled CCACCA and CCACCA-Phe, respectively. (+) and (-) indicate the presence and the absence of the enzyme. (B–D) Synthesis of the Phe-AMP intermediate by *Mt*FRS in the presence of unmodified *M. tuberculosis* $tRNA^{Phe}$ (**B**), *E. coli* $tRNA^{Phe}$ (**C**) and yeast $tRNA^{Phe}$ (**D**). Phe-[^{32}P]AMP synthesis was monitored by thin layer chromatography. Phe-AMP and ATP denote ^{32}P-labeled compounds. F – full reaction; -E, -tRNA and -Phe – control reactions lacking the enzyme, tRNA or amino acid.

ranges between 2.3 and 3.2 Å, with the largest differences observed for *Mt*$tRNA^{Phe}$ and *Tt*$tRNA^{Phe}$. There is one magnesium site in the D-loop region conserved between yeast, *E. coli* and *Mt*$tRNA^{Phe}$. As expected, the largest differences are in the 3′ and 5′ ends, anticodon, T- and D-loops. Some of the differences may result from the fact that we compare free tRNA (*E. coli* and yeast) with tRNA bound to protein (*M. tuberculosis* and *T. thermophilus*). In fact, anticodon region of these last two tRNAs are more similar to each other. The anticodon loop after the U33 base forms a canonical U-turn stabilized by an internal hydrogen bond of U33 N3 to phosphoryl group of A36 and 2′OH to N7 of A35. This element is conserved in all four structures. In *Mt*$tRNA^{Phe}$, six consecutive purine bases in the anticodon region (G34 to G39) stack together and form a rigid structure important for tRNA recognition (see below), while *E. coli*, *T. thermophilus* and yeast tRNAs have a block of five stacked purine bases. Majority of bases are involved in base-pairing/stacking, but curiously a number of bases project into solvent in different tRNAs and they correspond to this same RNA regions. For example: C17, U20, U47 (*M. tuberculosis*), U20, U45 (*E. coli*), D17, U47 (yeast), U17, U47 (*T. thermophilus*). These structural features must be conserved for functional reasons.

Ligand binding

The active site of FRS adenylation domain of the α subunit contains subpockets responsible for recognition of L-Phe and ATP substrates. Our structures illustrate how L-Phe and its adenylate analog bind to their respective sites (Figure 5). In the phenylalanine-dedicated subpocket (Figure 5), αPhe255, αPhe257, αAla305, αArg201, and αThr258 interact with the aromatic ring of the ligand in both *Mt*FRSt/F1 and *Mt*FRSt/F2 structures. αArg201 and αGln215 form hydrogen bonds with the carboxyl group while the amino group is bonded to αHis175 and, via a water molecule, to αGln215 and αGlu217. The latter residue participates in an extended network of hydrogen bonds which, besides αGln215, includes αSer177 and αGln180. The organization of the amino acid binding subpocket is virtually identical in both tRNA-enzyme structures complexed with L-Phe. However, in the neighborhood that does not directly participate in the amino acid recognition, the conformations of some side chains facing the adenine binding site differ between the two structures, and *Mt*FRSt/F2 contains Mg^{2+} ion bound nearby the pocket, interacting with αGlu263, αGlu279 and a water molecule. The adenine subpocket, in the absence of a biological ligand, is occupied by a glycerol molecule.

An analogous pattern of interactions with phenylalanyl portion of the F-AMS ligand is observed in the *Mt*FRSt/F-AMS1 and *Mt*FRSt/F-AMS2 complexes. Additional bonds are formed with sulfamoyl group, ribose and adenine moieties. The former interacts with the guanidinium group of αArg201 and via a water molecule with Glu263 while the sugar ring forms hydrogen bonds with αGln215, main chain carbonyl oxygen atom of αTrp280 and carboxyl group of αGlu279 (in chain A of *Mt*FRSt/F-AMS2). The nucleotide base is anchored between αHis209 and αPhe213 and participates in several hydrogen bonds: (i) N1 with main chain nitrogen atom of αThr210, (ii) N3 via a water molecule with main chain of αArg312 and αHis214, and side chain of αGlu311, (iii) N7 with αArg201 and αAsp203 via another molecule and (iv) N6 with main chain carbonyl group of αThr210 and carboxyl group of αAsp203. Most of these interactions are conserved between the two F-AMS complexes, regardless of the tRNA binding status, with the exception of αGlu279, αGlu263 and N7-mediated contact.

The conformation of α2 domain is susceptible to the type of ligand, as clear rearrangements are observed when L-Phe and F-AMS complexes are compared. Specifically, in the presence of F-AMS, the loop connecting β4 and β5 el-

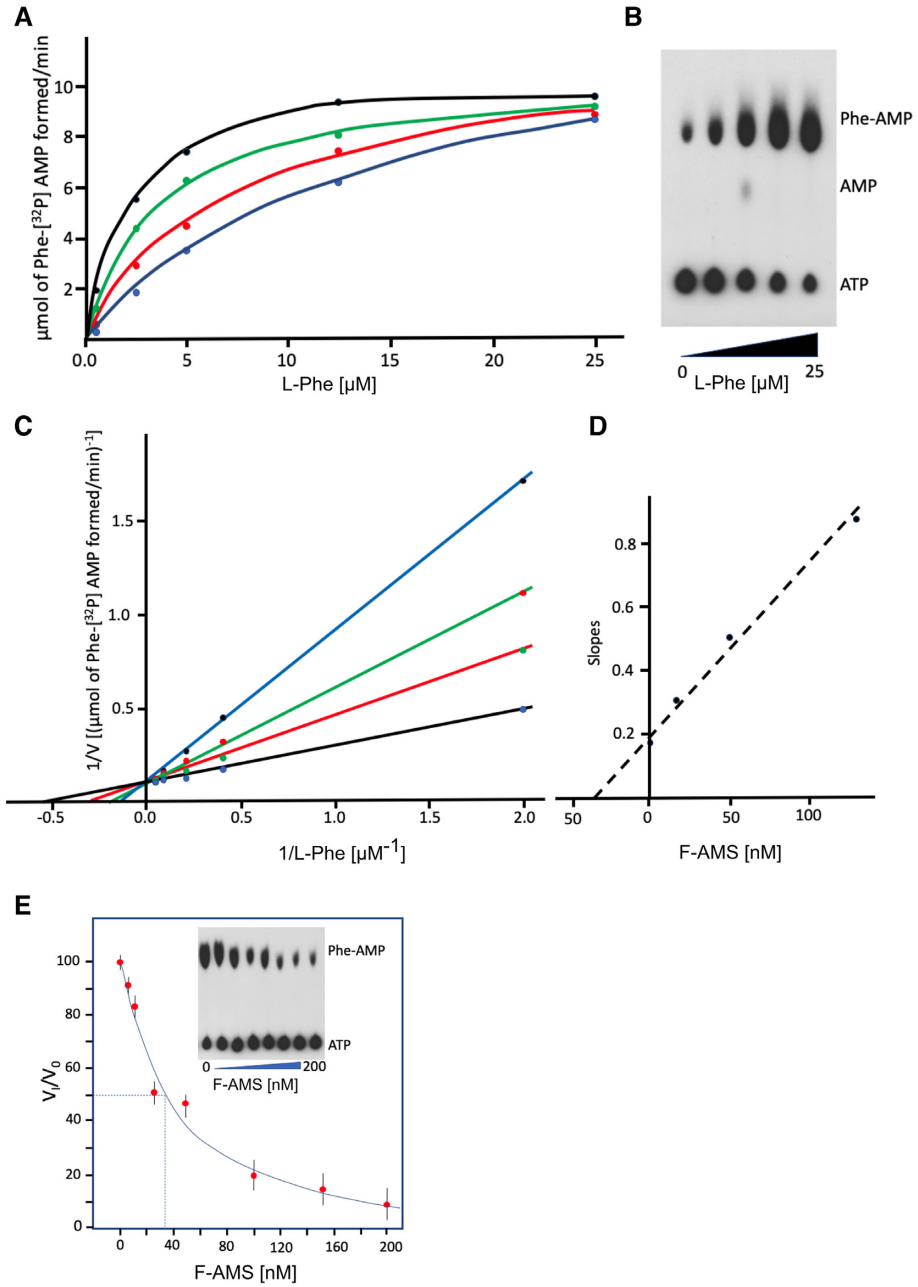

Figure 2. Inhibition of *Mt*FRS by F-AMS. (**A**) Competitive binding of inhibitor with respect to L-Phe. Inhibitor concentrations were as follows: 0 μM (black line), 15 nM (green line), 50 nM (red line), and 125 nM. (**B**) Phe-[^{32}P]AMP synthesis in the presence of 15 nM inhibitor (corresponding to green line) monitored by thin-layer chromatography. The rate of Phe-[^{32}P]AMP formation was determined at 37°C over 10 min. (C, D) Competitive inhibition of *Mt*FRS by the F-AMS shown in the form of a Lineweaver–Burk plot. K_m for Phe is 1.78 μM (**C**). The secondary diagram was used to determine the inhibition constant K_i, which is 33 nM (**D**). (**E**) Plot of the dose-response curve shows the percentage of initial velocity reduced by various concentrations of the F-AMS. The insert illustrates the inhibition of Phe-[^{32}P]AMP synthesis monitored by thin layer chromatography. Dotted lines indicate that the inhibitor has IC_{50} value of ∼36 nM.

ements from the β4↑β5↓β9↑ β8↓β7↑β6↓ sheet moves towards the β8↓β7↑β6↓ half, closing over the nucleotide moiety. This movement enables αHis209 and αAsp203 to approach the F-AMS base ring (Figure 5). The β8↓β7↑β6↓ section, especially the β7-β8 hairpin, also comes closer to the β4–β5 loop. Similar conformational changes upon binding of an adenylate analog were observed for other aaRS, for example *T. thermophilus* KRS (50).

Initial tRNA^{Phe} recognition

In the assembly, each of the tRNA molecules contacts all four subunits through mostly α1′and β3′ domains and to a lesser extent α2, β1 and β2. Remarkably, tRNA recognition is accomplished by one αβ (purple/grey) unit while aminoacylation and editing are completed by the complementary α'β' (coral/teal) dimer (Figure 3). In three structures, two with L-Phe and one with F-FMS (four indepen-

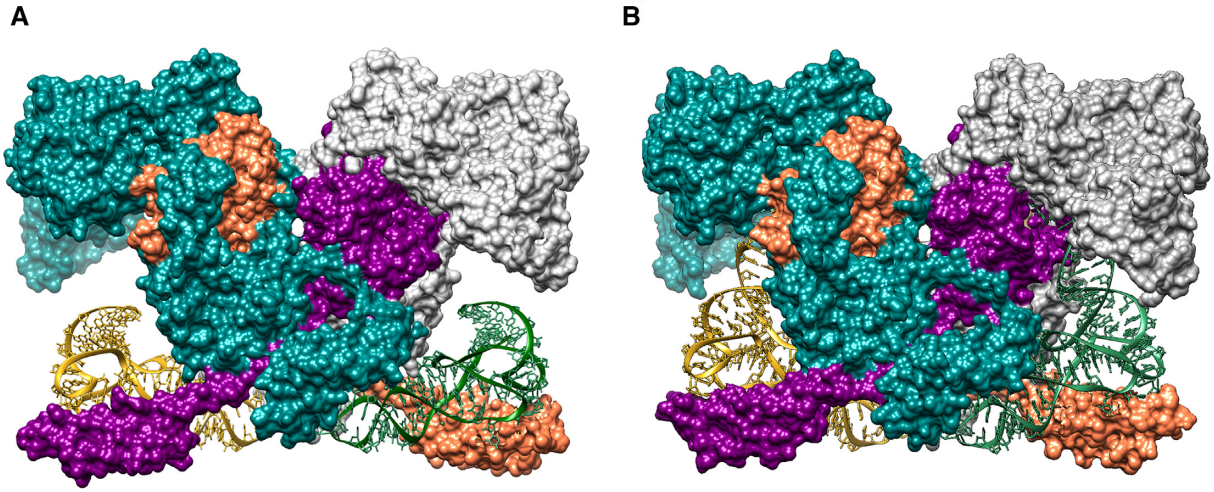

Figure 3. Overall structure of the *Mt*FRS-tRNA^{Phe} complex. (**A**) *Mt*FRS-tRNA^{Phe} heterotetramer with initial tRNA binding (*Mt*FRSt/F1) (**B**) *Mt*FRS-tRNA^{Phe} heterotetramer with the tRNA in aminoacylation ready state (*Mt*FRSt/F-AMS2). Subunits α shown in purple and coral, subunits β shown in gray and teal, tRNA represented as yellow and green ribbon models.

dent αβ/tRNA structural representations), we observe the initial recognition state. Here the anticodon-, D- and T-stem and loop regions of tRNA are well-ordered. On the other hand, the 3′ and 5′ ends of tRNA are disordered and consequently only 5–69 tRNA bases could be modeled. In these structures, access to the aminoacylation site is blocked by the 3_{10}-helix region of the α2 domain (residues 152–157, discussed below).

The tRNA selectivity is accomplished by enzyme interacting with the anticodon region of tRNA^{Phe} via anticodon binding domain β3′ (Figure 6). In *Mt*tRNA^{Phe} there is a rigid six-purine stack (involving G34 to G39, three of which constitute anticodon triplet (an identity element)) seems to be the key structural feature recognized by protein that provides a shallow surface with several specific contact points. The conserved β'Phe780 (FRSs have either phenylalanine or tyrosine in this position) interact via π–π stacking with the G34 guanine wobble anticodon base, extending the GAAAAG RNA stack. In addition, G34 forms several direct hydrogen bonds. Specifically, β'Asp778 hydrogen bonds to G34 N1 and N2, β'Arg830 to O6 and N7, and β'Gln784 interacts with G34 phosphoryl group or ribose ring oxygen atom (observed in two structures of the initial recognition state). Both β'Asp778 and β'Arg830 are highly conserved residues. There are two direct hydrogen bonds to A35, one from β'Ser747 to N6 and second from β'Thr793 to N1. A36 interacts with the carboxylate group of conserved β'Asp745 via ribose 2′OH. The A37 ribose 2′OH also interacts with β'Asp745.

The purine stack may represent a new recognition element in tRNA^{Phe} (*E. coli* and yeast tRNAs^{Phe} have an equivalent five-purine stack G34-A38) that is necessary for enhancing its binding specificity to cognate aaRSs. Because position 39 is occupied by uridine in almost all known tRNA^{Phe}, which has much lower propensity for stacking than purines, the role of the motif formed by stacked nucleotides 34 through 38–39 has had been overlooked until now. The stacking of the nucleotides 34–39 in tRNA^{Phe} molecules may play pivotal role in indirect readout of anticodon stem-loop structure (52). This possibility was first recognized when the crystal structure of the *E. coli* QRS–tRNA^{Gln} complex was determined (53,54).

The interactions with D-loop vary across all three structures in the initial tRNA recognition state, indicating the dynamic nature of these contacts. Here, the α1′ domain participates in hydrogen bonds to G19 and/or U20 via α'Arg45 and/or α'Arg56, α'Asp37. Inconsistent pattern of interactions is also present in binding with the T-loop. This region primarily implicates α1′Asn64 residues and, in some instances, α1′Ala57 and/or α1′Arg45 that form hydrogen bonds with C56 and G57.

tRNA^{Phe} binding in preaminoacylation complex

In the *Mt*FRSt/F-AMS2 structure, corresponding to the tRNA aminoacylation ready state, the entire tRNA molecule is ordered with the anticodon region firmly bound and the acceptor arm approaching the active site containing the F-AMS–a phenylalanine adenylate analog (Figure 6). In this state, the enzyme continues to interact with the anticodon region via domain β3′, maintaining all interactions described above for the initial tRNA recognition state. D-loop G19 and U20 bind to the α1′ domain with somewhat different patterns. For example, G19 contacts a coiled-coil extension in the enzyme's β-subunit, but contacts with the D20 identity determinant are not visible (55). Interestingly, photoaffinity cross-linking studies by Moor *et al.* (56) demonstrated that contacts between D20 and β-subunit take place in *Tt*FRS/tRNA^{Phe} complex. Therefore, our present studies of *Mt*FRS and earlier studies of *Tt*FRS suggest that subtle structural perturbation occur in these enzymes during aminoacylation. The T-loop is engaged as described for the initial tRNA recognition state and variable loop contributes one stabilizing interaction via the G45-α'Thr32 hydrogen bond.

Additional new features are observed as well. For example, the phosphoryl group of A37 interacts with two cations. One of them links to β'Glu807 and two water molecules,

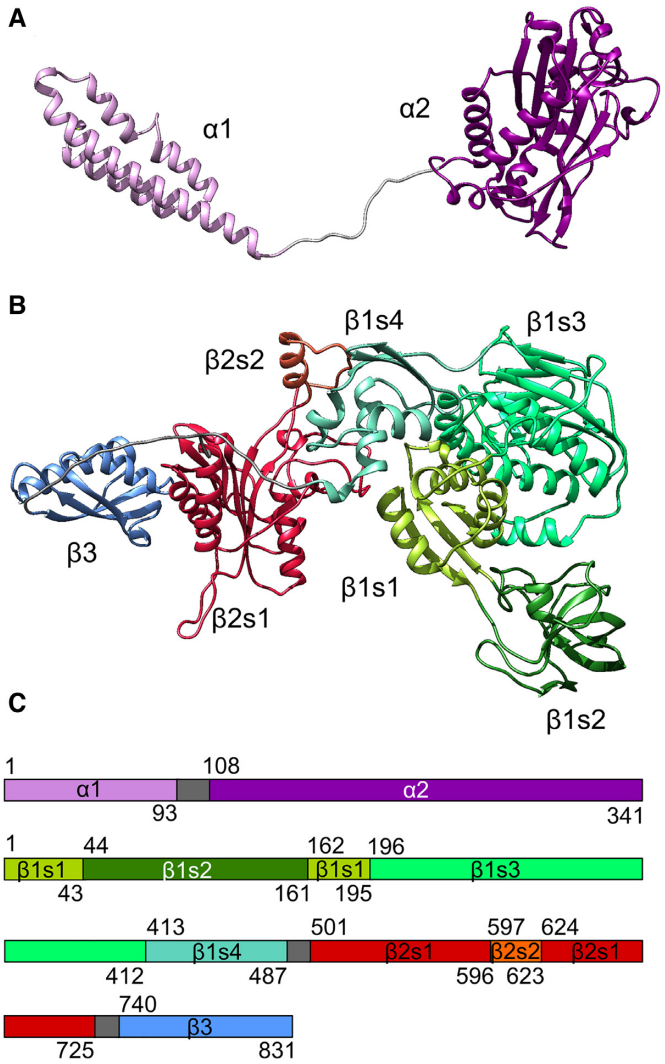

Figure 4. *Mt*FRS domains. (**A**) Subunit α. (**B**) Subunit β. (**C**) Domain boundaries with starting residues numbered above the sequence-representing bars and ending residues shown under the bars. Gray sections depict linkers.

the other binds to β'Phe743 and two water molecules. This latter region bridges the anticodon loop with the D-loop arm, as the magnesium-coordinating water molecules and β'Phe743 main chain amide group bind to phosphoryl group of G26. The β3′Glu807 main chain interacts with the phosphoryl group of C25, and C25 with G24 2′OH groups interact with main chain carbonyl oxygen atoms of β'Ala741 and β'Leu805. Additional hydrogen bonds link β'Pro738 to 2′OH of C11 and β'Thr804 to G24, with the latter OG1 – N2 interaction being the only one involving a D-loop base. Then, three hydrogen bonds connect the backbone of D-loop (G10, C11 and U12) to the β2s1 subdomain (βSer578 and βArg579). All the contacts with the D-loop arm (G10, C11, U12, G24, C25) exist only in the tRNA aminoacylation ready state.

The most significant structural change, besides tighter contacts with the D-loop arm, is the stabilization of the acceptor arm and 3′ end. This region becomes ordered and the CCA3′ motif is inserted into acylation site, as observed for other aaRS/tRNA complexes (57,58). This is possible because the α2 domain rearranges (Figure 6). Specifically, the 3_{10}-helix η2 and its short flanking sections (residues 151–159) unfold into a loop-like structure. Residues αGln158, αGlu157 and αAsp159 in the CCA3′ binding site dramatically change their positions away from the pocket providing access to the otherwise blocked active site. Thus, this section of the protein serves as a 3′ end gate. Another residue that moves significantly is αHis152, which relocates closer to the binding pocket. There are two consequences of these shifts—the first movement, particularly that of αAsp159, creates a new magnesium binding site and interacts with A76, while the second positions αHis152 for the interaction with the tRNA backbone.

The very well-ordered magnesium ion is hexagonally coordinated by five water molecules and the carboxyl group of αAsp203. The water molecules bridge the protein and the tRNA 3′ end, as they interact with αThr202, αGln158 (main chain), αAsp159 (main chain and side chain) as well as all three CCA bases. This rich network of water-mediated contacts appears to be crucial for tRNA anchoring in the active site as there are only two direct hydrogen bonds contributing to the binding: A76 phosphoryl group interacts with αHis152 and its purine N6 atom binds to the αAsp159 main chain. The magnesium ion might be therefore essential for proper positioning of the A76 ribose, which approaches the imino moiety of the ligand within the 3.0 Å distance of nitrogen atom to 2′OH. The observed geometry between the acyl donor (F-AMS) and acceptor (tRNA), seen for the first time for any aaRS-tRNA complex, mimics the reaction stage where the 2′OH attacks the carbonyl carbon atom and is fully consistent with the accepted mechanism for FRS (59,60).

*Mt*FRS vs other bacterial homologs

FRS homologs from *E. coli* (*Ec*FRS), *Staphylococcus haemoliticus* (*Sh*FRS), *P. aeruginosa* (*Pa*FRS), and the most extensively studied from *T. thermophilus* (*Tt*FRS) have been characterized structurally as functional heterotetramers (Table 1). Overall organization of these proteins is identical to that of *Mt*FRS, however superpositions of the entire chains lead to high RMSDs due to the movements of the individual subdomains.

Several features of the adenylation active site are common between the four proteins. In particular, the residues interacting with the L-Phe or adenylate ligands are mostly conserved and superpose well, especially in the vicinity of the amino acid moiety. The exceptions include αHis175 substitution by αGln169 in *Ec*FRS associated with a different conformation of the L-Phe ligand and inhibitor-induced adjustments to the equivalents of Gln180 in *Sh*FRS and *Pa*FRS. The region binding the nucleotide is less aligned. Interestingly, the reorganization of the α2 domain upon adenylate binding vs L-Phe binding varies between the homologs. While in *Ec*FRS, *Pa*FRS, and *Sh*FRS the conformation of ligand binding fragment of the α2 domain appears to follow the *Mt*FRS pattern, being a more compact in the presence of a nucleotide (*E. coli*) and an open without it (*P. aeruginosa*, *S. haemoliticus*), such a change is less pro-

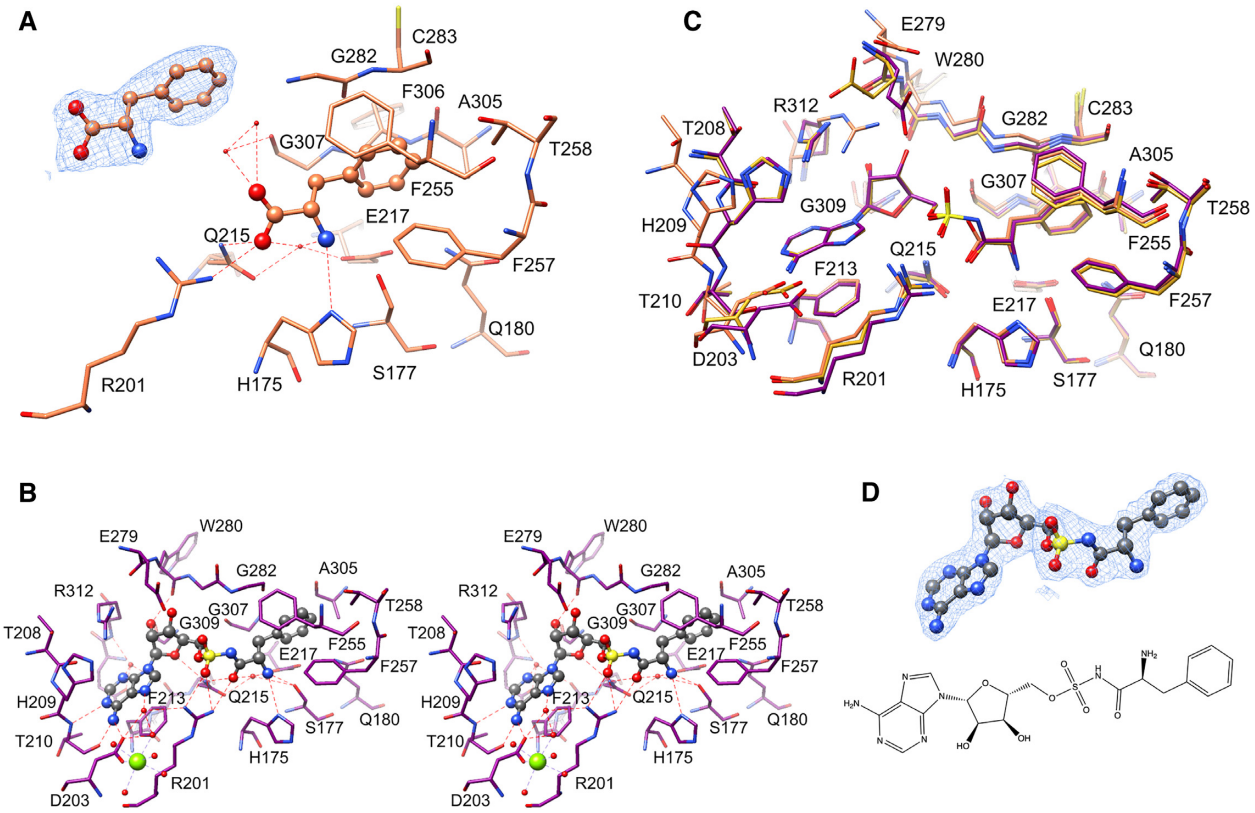

Figure 5. *Mt*FRS interactions with ligands. (**A**) L-Phe binding to the aminoacylation site of FRSt/F1. The upper-left insert shows 2mFo-DFc electron density map contoured around L-Phe at 1.2 σ. (**B**) Stereoview of F-AMS binding to the aminoacylation site of FRSt/F-AMS2. (**C**) Superposition of the aminoacylation site of FRSt/F-AMS2 (purple), FRSt/F-AMS1 (yellow), and FRSt/F1 (coral). (**D**) F-AMS ligand and its 2mFo-DFc electron density map from the FRSt/F-AMS2 complex contoured at 1.2 σ. Red spheres represent water molecules, a green sphere marks magnesium ion. Hydrogen bonds are shown as dashed lines.

nounced in *Tt*FRS. A small shift is only observed in the β4–β5 loop, the β8↓β7↑β6↓ subsection of the central β-sheet remains intact (Figure 7). The structure is not much different when *Tt*FRS is in complex with adenylate and tRNA.

The editing domain, encompassed within the β1s4 subdomain, also bears high level of similarity (Figure 6). The residues lining the amino acid binding pocket, *Mt*βHis275, *Mt*βLeu300, *Mt*βGly325, *Mt*βGlu344 and *Mt*βPro273 are conserved in *Tt*FRS, *Ec*FRS, *Pa*FRS, and *Sh*FRS. *Mt*βGly324, *Mt*βMet327, *Mt*βAla366 and *Mt*Tyr370 are sometimes replaced by similar residues (Supplementary Figure S2). The most striking difference is the presence of *Mt*βTrp348. This residue is unique to the *M. tuberculosis* homolog, all other FRS contain Phe in this position. The indole nitrogen atom is perfectly positioned to form a hydrogen bond with the ligand, possibly switching the amino acid recognition from repulsive to attractive interactions and providing distinct opportunities for selective inhibitors targeting the editing domain.

T. thermophilus is the only homolog allowing for comparisons of the tRNA binding. The anticodon recognition is achieved by binding to the β3′ domain, but the interactions with tRNA vary between the two proteins due to sequence variations. *Mt*β'Ser747 and *Mt*β'Thr793 are replaced by alanine residues (*Tt*β'Ala698 and *Tt*β'Ala744, respectively), leading to the loss of direct hydrogen bonds to A35 and a potential (water molecules are not modeled in the *Tt*FRSt structures) water-mediated contact with G34. Alanine residues in these positions are more common in homologs than Thr and Ser (Supplementary Figure S2). Additional hydrogen bond bridging to the RNA backbone is eliminated by another non-conserved substitution *Mt*β'Gln784/*Tt*β'Pro735. Residues *Mt*β'Asp745, *Mt*β'Phe777, *Mt*β'Asp778 and *Mt*β'Arg830 are conserved (*Tt*β'Asp696, *Tt*β'Phe728, *Tt*β'Asp729 and *Tt*β'Arg780) and bind the tRNA molecule, albeit in somewhat different fashion, and *Tt*β'Tyr731 stacks against G34 similarly to *Mt*β'Phe780.

In the two *Tt*FRS structures of complexes with tRNA, the 3′ end appears to be inserted into the active site of the acylation domain. However, neither of the modeled states illustrate the reaction-ready position of the tRNA molecule observed in *Mt*FRSt/F-AMS2. In the *Tt*FRS/tRNA^{Phe} complex, the 3′ end adenine moiety occupies the adenylate binding pocket, it is of note, however, that the electron density map for the acceptor arm is poor and therefore the accuracy of the model raises some concerns. In the *Tt*FRS/tRNA^{Phe}/adenylate structure, the 3′ end adenosine nucleotide faces the solvent. In neither of the *Tt*FRS structures, the region equivalent to the 3′ end gate is significantly reorganized to facilitate tRNA binding. It thus suggests that in the observed states, the *Tt*FRS 3′ end gate remains closed,

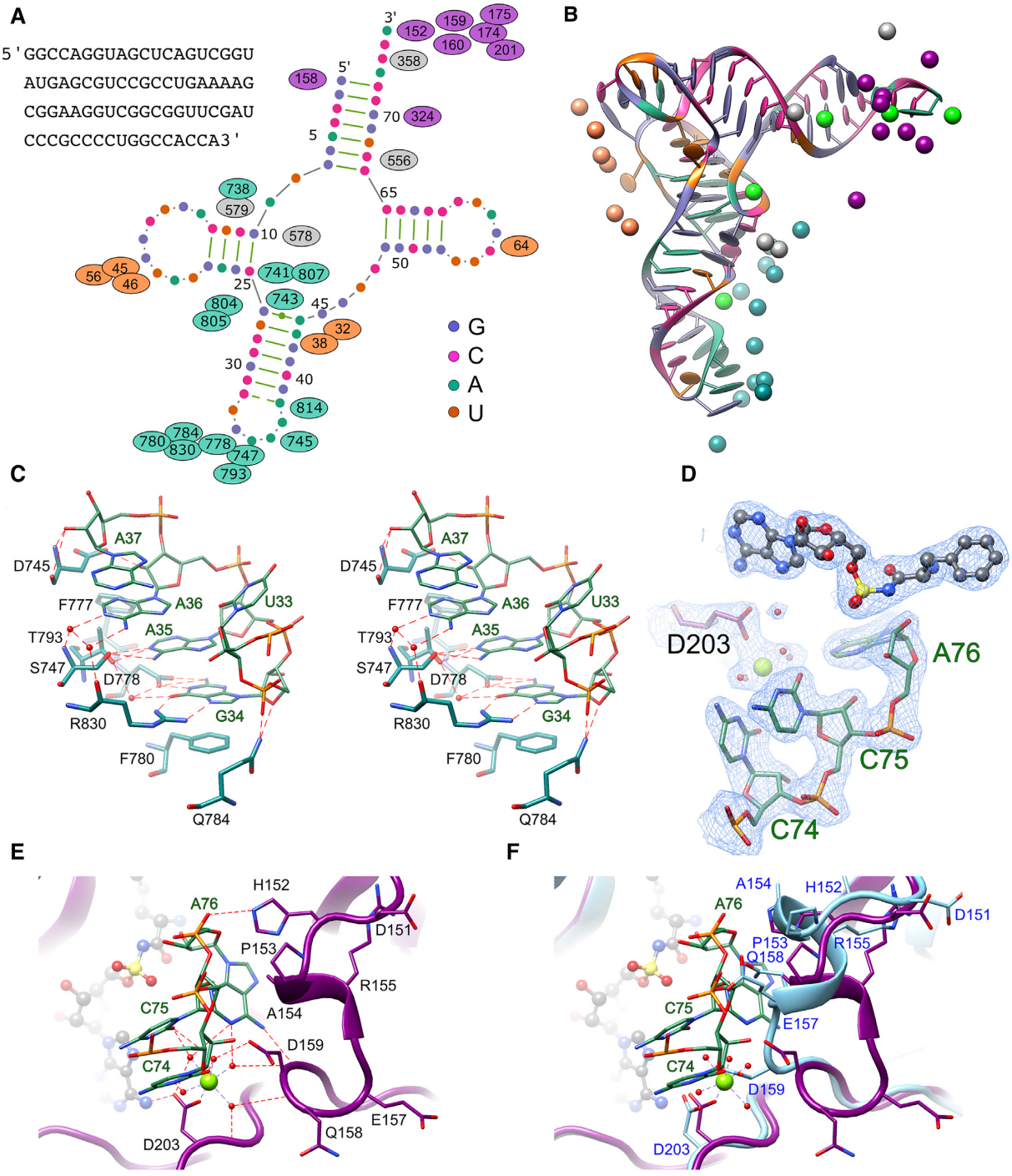

Figure 6. *Mt*FRS interactions with tRNA. (**A**) tRNA sequence and molecule with interacting residues in *Mt*FRSt/F-AMS2. (**B**) tRNA molecule and its contacts with the polypeptide chain as seen in *Mt*FRSt/F-AMS2 (chain C). The interacting residues include those forming hydrogen bonds, participating in base/amino acid pairs and base/amino acid stacks, as defined by the SNAP program (http://forum.x3dna.org/dna-protein-interactions/). The Cα atoms of the interacting residues are shown, colored as in A. Magnesium ions are colored green. Nucleotides have the following color-coding: G-purple, C-pink, A-green, U-orange. (**C**) Stereoview of *Mt*FRS interactions with the anticodon loop. (**D**) F-AMS (gray), tRNA 3′ end (green) and magnesium ion (green sphere) binding in the *Mt*FRSt/F-AMS2 active site shown in 2mFo – DFc electron density map contoured at 1.2 σ. (**E**) tRNA 3′ end interactions in *Mt*FRSt/F-AMS2. (**F**) Superposition of the tRNA 3′ end binding region of *Mt*FRSt/F-AMS2 (purple) and *Mt*FRSt/F-AMS1 (blue).

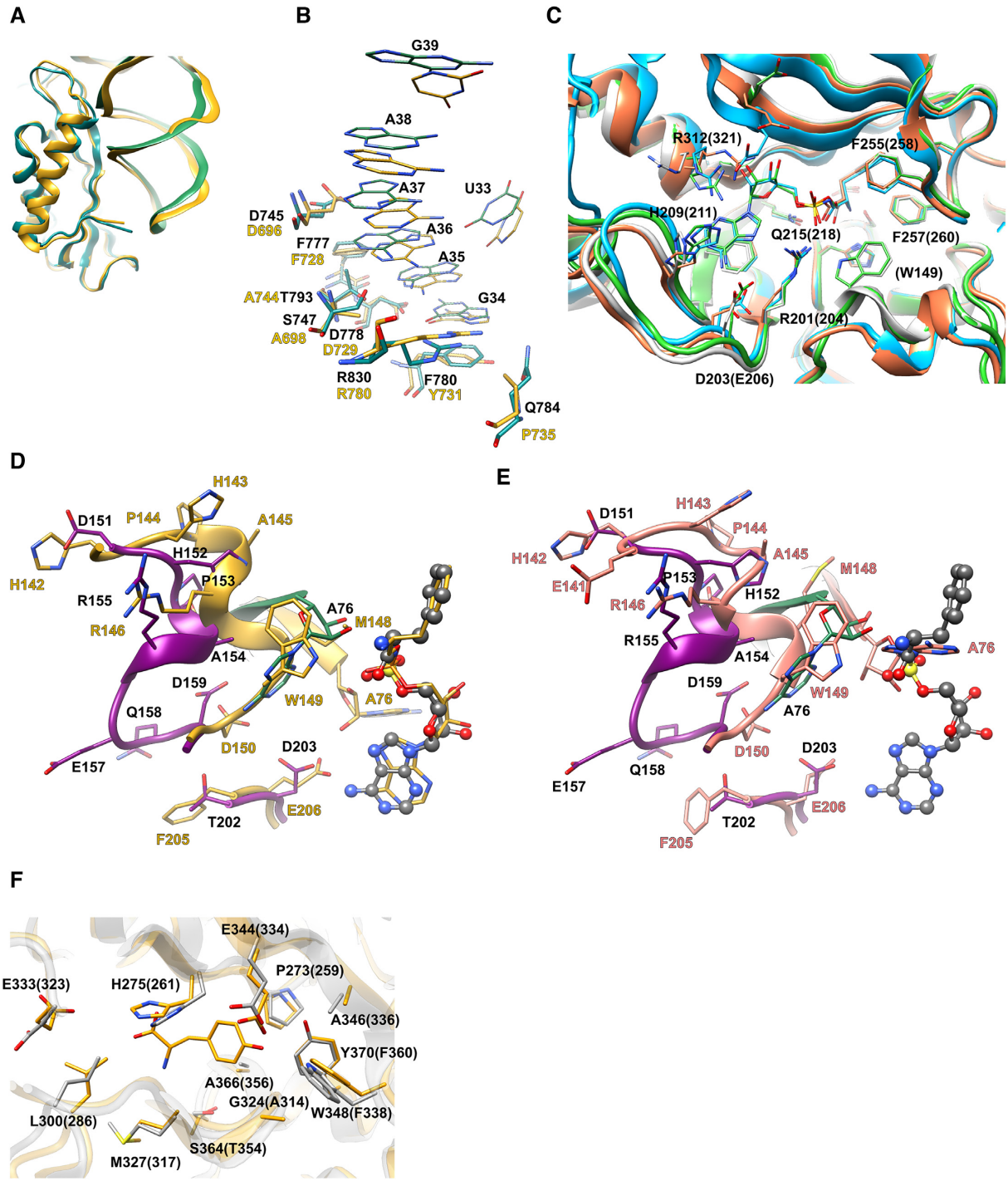

Figure 7. Comparison of *Mt*FRS with *T. thermophilus* FRS. (**A, B**) Comparison of the anticodon binding between *Mt*FRSt/F-AMS2 (green, chain C/E) and *Tt*FRSt/adenylate (yellow, PDB 2IY5). In (A) ribbon representation of the β3′ domain and anticodon loop. In (B) atomic representation of the anticodon loop and interacting residues. For superposition only domain β3′ was used. Labels in black correspond to *Mt*FRS, in yellow to *Tt*FRS. (**C**) Superposition of the aminoacylation domains from *Mt*FRSt/F (coral), *Mt*FRSt/F-AMS1 (blue), *Tt*FRS/F (grey, 1B70), and *Tt*FRS/adenylate (green, 1B7Y). Only select residues are labeled. Parentheses indicate *Tt*FRS. (**D**) Superposition of the 3′ end binding region in *Mt*FRSt/F-AMS2 and *Tt*FRSt/adenylate. *Mt*FRS is shown in purple (protein), green (tRNA) and grey (F-AMS). *Tt*FRS is shown in yellow. (**E**) As in (D) but with the *Tt*FRSt complex (pink). Labels in black correspond to *Mt*FRS, in yellow or pink to *Tt*FRS. (**F**) Superposition of the editing domains (β1s4 subdomain) from *Mt*FRSt/F-AMS2 (grey) and *Tt*FRS/L-Tyr complex (gold, 2AMC).

preventing the proper positioning of CCA3′ end. Of note, while the η2 sequence is conserved in FRSs, the following three amino acids are not: *Mt*FRS contains the GEQ segment (residues 156–158) while *Tt*FRS has DMW. It is not only folded differently, but also is positioned closer to the ligand binding pocket, with αTrp149 facing the lumen and overlapping with the 3′ end adenine of tRNA. Another feature distinct for *Tt*FRS is the replacement of *Mt*αAsp203 with *Tt*Glu206, indicating that the potential magnesium binding site created during the aminoacylation reaction may look somewhat different. In other bacterial homologs the sequence of 3′ end gate region varies and adopts either the *Tt*FRS state (*Ec*FRS) or *Mt*FRSt/F-AMS2 state (*St*FRS, *Pa*FRS). While they preserve Asp corresponding to *Mt*αAsp203, neither binds a metal cation in the magnesium site. Given the lack of any obvious common patterns guiding the conformation of the 3′ end gate, such as the tRNA presence, magnesium coordination, ligand binding or sequence motifs, it appears that its state might be governed by cellular conditions these enzymes must function in.

*Mt*FRS versus human FRSs

Humans have two FRSs, cytoplasmic (hcFRS) and mitochondrial (hmFRS), both of which have had their structures determined (Table 1). The hcFRS resembles *Mt*FRS $(\alpha\beta)_2$ heterotetramer but it lacks the anticodon binding domain and contains an extra domain in the α subunit. The hmFRS variant, characterized in several states, including heterologous complex with *T. thermophilus* $tRNA^{Phe}$, is a monomeric chimera of the bacterial α2 catalytic module and anticodon binding domain β3. HmFRS exists in two states: closed in the absence of tRNA, where the β3 domain blocks the access to α2, and open, which mimics the domains arrangement seen in bacterial $(\alpha\beta)_2$ complexes (11). While overall organization and sequences of the human enzymes are distinct from *Mt*FRS, the corresponding domains superpose well. The aminoacylation domain of hmFRS adopts conformation similar to that of *Mt*FRSt/F-AMS2, with the β7-β8 hairpin equivalent closed over the pocket and the 3′ end gate open regardless of the ligand/tRNA binding status. The cytosolic homolog, characterized in the presence of L-Phe, remains more open, as seen in the initial tRNA recognition complexes, with its 3′ end gate open as well.

The aminoacylation sites of the human enzymes, especially hmFRS, share high similarity with *Mt*FRS (Figure 8). Several residues lining the adenylate binding site are also conserved. In hmFRS the only exception includes replacement of *Mt*αVal286, sitting deep in the L-Phe pocket, by hmMet258. HcFRS is more variable, with several substitutions in the L-Phe binding site: *Mt*αPhe257/hcαTyr412, *Mt*αPhe255/hcαAsn410, *Mt*αVal286/hcαPhe438, *Mt*αGly282/hcαAsn434, and *Mt*αGln180/hcαSer332. These residues are located in the L-Phe binding cavity, which is generally better aligned between the homologs than the adenine-dedicated pocket.

The editing domain of hcFRS is differently organized in terms of the secondary structure composition, and consequently some of the residues lining the cavity come from different structural elements than in *Mt*FRS, making the architecture of this site substantially different. In particular, the side chain of *Mt*βTyr370 is in three-dimensional context replaced by hcβArg118 and the role of *Mt*βTrp348 is played by hcβThr256 and the surface created by *Mt*βAla366 is lined with hcβArg154. Also, some residues contributed by analogous secondary structures differ: *Mt*bHis275 is replaced by hcβGly160 while *Mt*βGly328 is mutated to hcβAsn238.

The tRNA anticodon interactions with hmFRS mostly resemble those found in *Mt*FRS including recognition of the wobble base and the key residues are conserved or replaced by similar side chains (*Mt*β'Thr793 to hmCys377 or *Mt*β'Phe777 to hmIle363). The hydrogen bond linking the protein with the RNA backbone via *Mt*β'Gln784 (replaced by hmLys370) is eliminated due the conformational rearrangement of this protein segment.

Functional implications

All tRNAs must have similar shape to fit into ribosomal translation machinery and have identical CCA3′ end to deliver amino acid to the ribosome active site. The tRNAs seem to have common origin and have significant sequence conservation, as they were generated from repeating sequences and inverted repeats (6). During evolution these tRNAs differentiated into over 20 amino acid specific groups. It is believed that aaRSs emerged together with tRNAs and they coevolved, making aaRSs one of the oldest enzymes to arise. Today's aaRS utilize different scaffolds, tRNA recognition strategies and catalytic mechanisms to catalyze aminoacylation reaction. In bacteria, each enzyme must specifically recognize tRNA(s) specific for a single amino acid from a pool of some 30 plus tRNAs.

So how do aaRSs enzymes make the connection? The amino acid specific isoacceptor tRNAs may have up to six different codons and these tRNAs are aminoacylated by a single enzyme. At the same time, amino acid must be delivered to adenosine ribose at 3′ end, about 65 Å away from the anticodon. Therefore, for aaRSs codon recognition may not be the easiest solution to accomplish this.

It was proposed that a code exists equivalent to the 'genetic code' in the interactions between aaRS and tRNAs. This code was linked to amino acid properties. One hypothesis proposed that aaRS emerged first as ancestor urzyme-like synthetase that initially recognized only the acceptor stem and later evolved to include other parts of tRNA to increase specificity. Urzymes correspond to conserved cores of both classes of aaRSs lacking their anticodon-binding domains. Surprisingly, urzymes catalyze amino acid activation and acyl transfer with affinities similar to full length aaRSs (61). The argument for first aaRS to recognize tRNA acceptor stems rather than anticodons is in agreement with the idea that originally proteins were coded by an 'operational RNA code' located in the acceptor stem (62,63).

Structural studies of aaRS complexes with tRNA showed, that there are two groups of enzymes, one that recognizes both anticodon loop and stem to discriminate between cognate and non-cognate tRNA and the other that does not. Earlier studies of the yeast FRS binding to $tRNA^{Phe}$ identified five bases that were important for se-

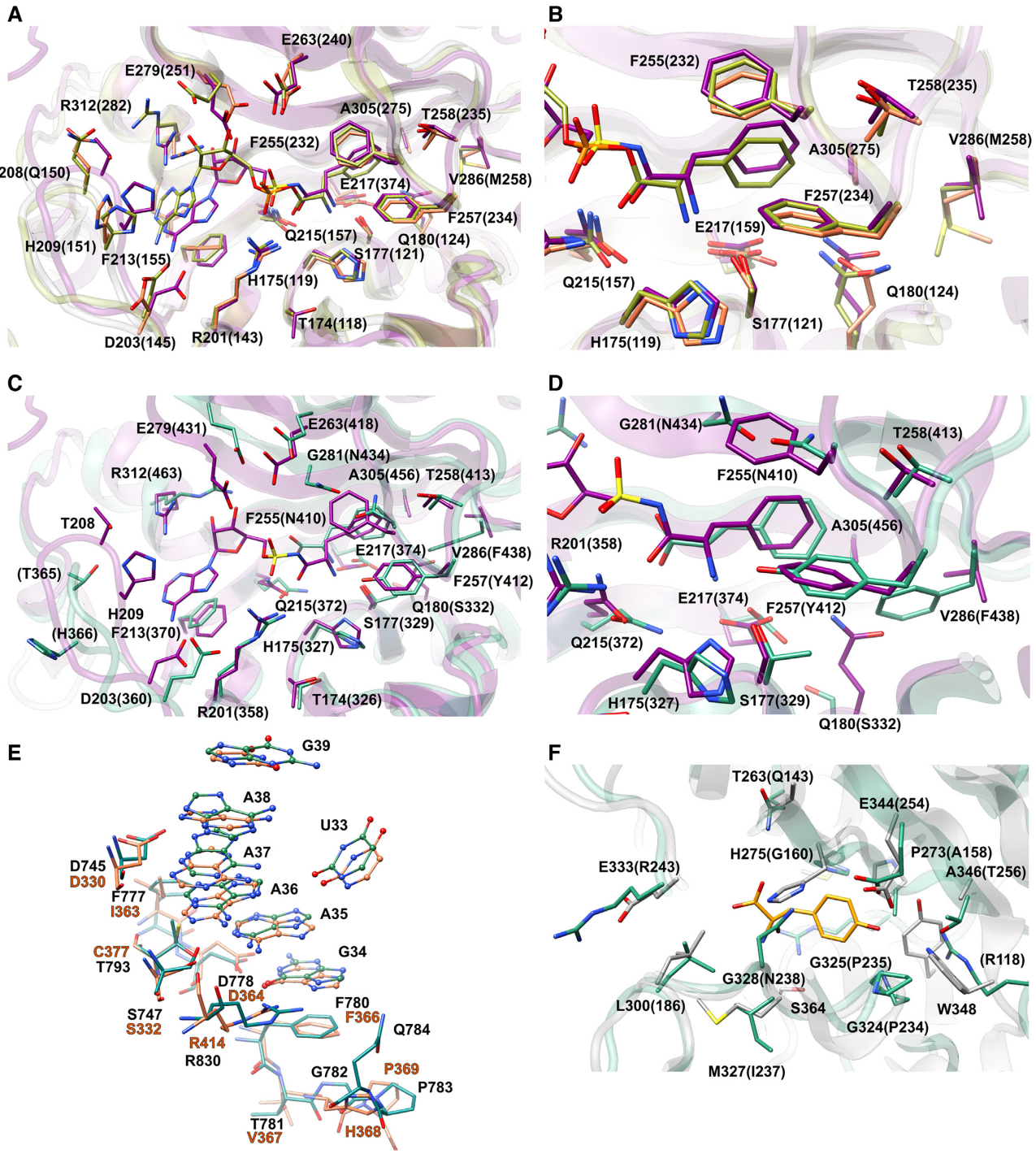

Figure 8. Comparison of *Mt*FRS with human FRSs. (**A**, **B**) Superposition of the aminoacylation domains from *Mt*FRSt/F-AMS2 (purple) with human mitochondrial (hm) enzyme in complex with adenosine-5′-[phenylalaninyl-phosphate] (3CMQ, olive green) or with *T. termophilus* $tRNA^{Phe}$, hmFRSt (3TUP, coral). In (B), close-up view of the L-Phe subpocket is shown. (C, D). Superposition of the aminoacylation domains from *Mt*FRSt/F-AMS2 (purple) with human cytosolic enzyme (hcFRS) in complex with L-Phe (3L4G, green). In (**D**), close-up view of the L-Phe subpocket is shown. (**E**) Superposition of the anticodon binding region of *Mt*FRSt/F-AMS2 (gray) and hmFRSt (coral). Labels in parentheses correspond to the respective human homologs. (**F**) Superposition of the editing domains (β1s4 subdomain) from *Mt*FRSt/F-AMS2 (gray) and hcFRS (3L4G, green). Gold L-Tyr from the *Tt*FRS/L-Tyr complex marks the amino acid binding site.

lectivity in positions 20, 34, 35, 36 and 72 (64), implying that FRS belongs to the former category. Two codons code for L-Phe, 3′CUU and 3′UUU, however typically only one isoacceptor $tRNA^{Phe}$ with 5′GAA anticodon is required to decode both of these two codons, as the wobble G base can pair with both C and U. This is the case for *M. tuberculosis*, as it has only one $tRNA^{Phe}$. Our structures demonstrated direct FRS interactions only with U20, G34, A35 and A36, with U20 and A36 having very minor contributions. Therefore, the role of A36 and A72 must be indirectly affecting the structure of tRNA that allows it to make the other interactions. Additional interactions are observed through phosphate and sugar moieties. Interestingly, this pattern is in good agreement with nucleotides characterized as identity determinants for aminoacylation of $tRNA^{Phe}$ by *E. coli* FRS (65) where ten out of the eleven nucleotides (G10, U20, G25, A26, G34, A35, A36, G44, G45, U59 and A73) are involved in interactions.

Our data confirm the interaction of FRS with the anticodon loop and stem, yet no anticodon sequence is recognized. The key feature of the initial detection is binding to the G34 wobble base, the least important in codon-anticodon pairing. With the most extensive interactions with β3′ domain residues G34 appears to be the primary discriminator with some additional selectivity provided by A35, A36, G19 and U20. We believe that the recognition of the anticodon loop and stem is governed by the structural motif of stacked five-six purine bases (six for *M. tuberculosis* $tRNA^{Phe}$, five for *T. thermophilus* $tRNA^{Phe}$), which positions tRNA and FRS to make additional contacts, including D- and T-loops. This interface is preserved across the FRS family, including much simplified version of the enzyme found in human mitochondria. FRS does not interact with the acceptor stem at all, in fact in the initial recognition step the acceptor stem is disordered and the access to active site is often blocked by the enzyme itself. Completion of the recognition process allows for conformational changes in the 3′ gate permitting binding of the CCA3′ region into the active site.

It seems FRS lost its need to read the CCA stem sequence. This is because other parts of $tRNA^{Phe}$ provide unique signatures sufficient to allow FRS to distinguish it from other tRNAs. FRS is a large, multidomain enzyme that provides many opportunities for interactions and specificity. tRNA molecules also offer a variety of surfaces for interactions. Based on our results and earlier published data, we hypothesize that recognition of the CCA stem might have been essential for ancestral aaRSs to discriminate between tRNAs, but later in the evolution of the protein genetic code, new specificity determinants were added while previous were lost. This indirect recognition in protein-nucleic acids interactions is one of the most important highlights of biological systems.

Therapeutics development

With a multidomain structure rich in pockets, equally complex reaction and essentiality, FRS represents an important target for drug discovery. Highlighted by our study structural differences between the active sites of bacterial and human FRS could be exploited to design narrow spectrum inhibitors that would have low toxicity for humans, other animal hosts and their microbiomes. Admittedly, close similarity of the adenylate binding cavity between *Mt*FRS and mitochondrial homolog presents a formidable challenge for medicinal chemistry. Therefore, targeting the editing domain, substantially distinct from the cytoplasmic variant and absent in the mitochondrial enzyme may offer an untapped opportunity. We will pursue a drug discovery program of *Mt*FRS and plan to present our results in the future manuscript.

DATA AVAILABILITY

The structural datasets generated during the current study are available in the PDB repository (https://www.rcsb.org/) under accession codes: 7KA0 (https://www4.rcsb.org/structure/7KA0), 7KAB (https://www4.rcsb.org/structure/7KAB), 7K9M (https://www4.rcsb.org/structure/7K9M) and 7K98 (https://www4.rcsb.org/structure/7K98). Diffraction images are available on server in Dr. W. Minor laboratory https://proteindiffraction.org. All other data generated during the current study including the raw kinetic and biophysical data are available upon request.

ENDNOTES

Abbreviations used: aaRS, aminoacyl-tRNA synthetase; F-AMS, 5′-O-(N-phenylalanyl)sulfamoyl-adenosine; FRS, phenylalanyl-tRNA synthetase; *Mt*$tRNA^{Phe}$, full length unmodified *M. tuberculosis* $tRNA^{Phe}$ transcript; *Mt*FRS, *Mycobacterium tuberculosis* phenylalanyl-tRNA synthetase; *Mt*FRSt/F1, *Mt*$tRNA^{Phe}$/L-Phe complex crystal form 1; *Mt*FRSt/F2, *Mt*$tRNA^{Phe}$/L-Phe complex crystal form 2; *Mt*FRSt/F-AMS1, *Mt*$tRNA^{Phe}$/F-AMS1 complex crystal form 1; *Mt*FRSt/F-AMS1, *Mt*$tRNA^{Phe}$/F-AMS2 complex crystal form 2; hmFRS, human mitochondrial FRS; hcFRS, human cytosolic FRS; ALS, alanyl-tRNA synthetase; DRS, aspartyl-tRNA synthetase; ERS, glutamyl-tRNA synthetase; GRS, glycyl-tRNA synthase; QRS, glutaminyl-tRNA synthetase; HRS, histidyl-tRNA synthetase; IRS, isoleucyl-tRNA synthetase; LRS, leucyl-tRNA synthetase; KRS, lysyl-tRNA synthetase; TRS, threonyl-tRNA synthetase; SRS, seryl-tRNA synthase, TB, tuberculosis, XDR-TB, extensively drug resistant tuberculosis; MDR-TB, multi-drug resistant tuberculosis.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We truthfully thank the members of the SBC at Argonne National Laboratory for their help with setting beamline and data collection at beamline 19-ID. We also acknowledge support from the Structure-guided Drug Discovery Coalition funded by the Bill & Melinda Gates Foundation.

FUNDING

National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of

Health and Human Services [HHSN272201200026C, HHSN272201700060C, in part]; U.S. Department of Energy (DOE) Office of Science and operated for the DOE Office of Science by Argonne National Laboratory [DE-AC02-06CH11357]; J.W. was supported by the Hatch program of the National Institute of Food and Agriculture, U.S. Department of Agriculture. Funding for open access charge: National Institutes of Health.
Conflict of interest statement. None declared.

REFERENCES

1. World Health Organization. (2019) In: Global tuberculosis report 2019.
2. Mabhula,A. and Singh,V. (2019) Drug-resistance in Mycobacterium tuberculosis: where we stand. *Medchemcomm*, **10**, 1342–1360.
3. Shetye,G.S., Franzblau,S.G. and Cho,S. (2020) New tuberculosis drug targets, their inhibitors, and potential therapeutic impact. *Transl Res*, **220**, 68–97.
4. Li,X., Hernandez,V., Rock,F.L., Choi,W., Mak,Y.S.L., Mohan,M., Mao,W., Zhou,Y., Easom,E.E., Plattner,J.J. *et al.* (2017) Discovery of a potent and specific *M. tuberculosis* leucyl-tRNA synthetase inhibitor: (*S*)-3-(aminomethyl)-4-chloro-7-(2-hydroxyethoxy)benzo[*c*][1,2]oxaborol-1(3H)-ol (GSK656). *J. Med. Chem.*, **60**, 8011–8026.
5. Wallis,R.S., Dawson,R., Friedrich,S.O., Venter,A., Paige,D., Zhu,T., Silvia,A., Gobey,J., Ellery,C., Zhang,Y. *et al.* (2014) Mycobactericidal activity of sutezolid (PNU-100480) in sputum (EBA) and blood (WBA) of patients with pulmonary tuberculosis. *PLoS One*, **9**, e94462.
6. Kim,Y., Opron,K. and Burton,Z.F. (2019) A tRNA- and anticodon-centric view of the evolution of aminoacyl-tRNA synthetases, tRNAomes, and the genetic code. *Life (Basel)*, **9**, 37.
7. Rubio Gomez,M.A. and Ibba,M. (2020) Aminoacyl-tRNA synthetases. *RNA*, **26**, 910–936.
8. Ibba,M., Losey,H.C., Kawarabayasi,Y., Kikuchi,H., Bunjun,S. and Soll,D. (1999) Substrate recognition by class I lysyl-tRNA synthetases: a molecular basis for gene displacement. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 418–423.
9. Beyer,D., Kroll,H.P., Endermann,R., Schiffer,G., Siegel,S., Bauser,M., Pohlmann,J., Brands,M., Ziegelbauer,K., Haebich,D. *et al.* (2004) New class of bacterial phenylalanyl-tRNA synthetase inhibitors with high potency and broad-spectrum activity. *Antimicrob. Agents Chemother.*, **48**, 525–532.
10. Diaz-Lazcoz,Y., Aude,J.C., Nitschke,P., Chiapello,H., Landes-Devauchelle,C. and Risler,J.L. (1998) Evolution of genes, evolution of species: the case of aminoacyl-tRNA synthetases. *Mol. Biol. Evol.*, **15**, 1548–1561.
11. Klipcan,L., Moor,N., Finarov,I., Kessler,N., Sukhanova,M. and Safro,M.G. (2012) Crystal structure of human mitochondrial PheRS complexed with tRNA(Phe) in the active "open" state. *J. Mol. Biol.*, **415**, 527–537.
12. Lee,E.Y., Kim,S. and Kim,M.H. (2018) Aminoacyl-tRNA synthetases, therapeutic targets for infectious diseases. *Biochem. Pharmacol.*, **154**, 424–434.
13. Mosyak,L., Reshetnikova,L., Goldgur,Y., Delarue,M. and Safro,M.G. (1995) Structure of phenylalanyl-tRNA synthetase from Thermus thermophilus. *Nat. Struct. Biol.*, **2**, 537–547.
14. Sasaki,H.M., Sekine,S., Sengoku,T., Fukunaga,R., Hattori,M., Utsunomiya,Y., Kuroishi,C., Kuramitsu,S., Shirouzu,M. and Yokoyama,S. (2006) Structural and mutational studies of the amino acid-editing domain from archaeal/eukaryal phenylalanyl-tRNA synthetase. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 14744–14749.
15. Evdokimov,A.G., Mekel,M., Hutchings,K., Narasimhan,L., Holler,T., McGrath,T., Beattie,B., Fauman,E., Yan,C., Heaslet,H. *et al.* (2008) Rational protein engineering in action: the first crystal structure of a phenylalanine tRNA synthetase from Staphylococcus haemolyticus. *J. Struct. Biol.*, **162**, 152–169.
16. Finarov,I., Moor,N., Kessler,N., Klipcan,L. and Safro,M.G. (2010) Structure of human cytosolic phenylalanyl-tRNA synthetase: evidence for kingdom-specific design of the active sites and tRNA binding patterns. *Structure*, **18**, 343–353.
17. Mermershtain,I., Finarov,I., Klipcan,L., Kessler,N., Rozenberg,H. and Safro,M.G. (2011) Idiosyncrasy and identity in the prokaryotic Phe-system: crystal structure of E. coli phenylalanyl-tRNA synthetase complexed with phenylalanine and AMP. *Protein Sci.*, **20**, 160–167.
18. Reshetnikova,L., Moor,N., Lavrik,O. and Vassylyev,D.G. (1999) Crystal structures of phenylalanyl-tRNA synthetase complexed with phenylalanine and a phenylalanyl-adenylate analogue. *J. Mol. Biol.*, **287**, 555–568.
19. Kotik-Kogan,O., Moor,N., Tworowski,D. and Safro,M. (2005) Structural basis for discrimination of L-phenylalanine from L-tyrosine by phenylalanyl-tRNA synthetase. *Structure*, **13**, 1799–1807.
20. Kwon,N.H., Fox,P.L. and Kim,S. (2019) Aminoacyl-tRNA synthetases as therapeutic targets. *Nat. Rev. Drug Discov.*, **18**, 629–650.
21. Jarvest,R.L., Erskine,S.G., Forrest,A.K., Fosberry,A.P., Hibbs,M.J., Jones,J.J., O'Hanlon,P.J., Sheppard,R.J. and Worby,A. (2005) Discovery and optimisation of potent, selective, ethanolamine inhibitors of bacterial phenylalanyl tRNA synthetase. *Bioorg. Med. Chem. Lett.*, **15**, 2305–2309.
22. Hu,Y., Palmer,S.O., Munoz,H. and Bullard,J.M. (2014) High throughput screen identifies natural product inhibitor of phenylalanyl-tRNA synthetase from *Pseudomonas aeruginosa* and *Streptococcus pneumoniae*. *Curr. Drug Discov. Technol.*, **11**, 279–292.
23. Wang,L.-N., Di,W.-J., Zhang,Z.-M., Zhao,L.-l., Zhang,T., Deng,Y.-R. and Yu,L.-Y. (2016) Small-molecule inhibitors of the tuberculosis target, phenylalanyl-tRNA synthetase from Penicillium griseofulvum CPCC-400528. *Cogent Chem.*, **2**, 1181536.
24. Kim,Y., Babnigg,G., Jedrzejczak,R., Eschenfeldt,W.H., Li,H., Maltseva,N., Hatzos-Skintges,C., Gu,M., Makowska-Grzyska,M., Wu,R. *et al.* (2011) High-throughput protein purification and quality assessment for crystallization. *Methods*, **55**, 12–28.
25. Sherlin,L.D., Bullock,T.L., Nissan,T.A., Perona,J.J., Lariviere,F.J., Uhlenbeck,O.C. and Scaringe,S.A. (2001) Chemical and enzymatic synthesis of tRNAs for high-throughput crystallization. *RNA*, **7**, 1671–1678.
26. Wellington,S., Nag,P.P., Michalska,K., Johnston,S.E., Jedrzejczak,R.P., Kaushik,V.K., Clatworthy,A.E., Siddiqi,N., McCarren,P., Bajrami,B. *et al.* (2017) A small-molecule allosteric inhibitor of Mycobacterium tuberculosis tryptophan synthase. *Nat. Chem. Biol.*, **13**, 943–950.
27. Minor,W., Cymborowski,M., Otwinowski,Z. and Chruszcz,M. (2006) HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes. *Acta Crystallogr. D-Biol. Crystallogr.*, **62**, 859–866.
28. French,S.W.K. (1977) On the treatment of negative intensity observations. *Acta Cryst.*, **A34**, 517–525.
29. Padilla,J.E. and Yeates,T.O. (2003) A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta Crystallogr. D-Biol. Crystallogr.*, **59**, 1124–1130.
30. Winn,M.D., Ballard,C.C., Cowtan,K.D., Dodson,E.J., Emsley,P., Evans,P.R., Keegan,R.M., Krissinel,E.B., Leslie,A.G., McCoy,A. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D-Biol. Crystallogr.*, **67**, 235–242.
31. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D-Biol. Crystallogr.*, **60**, 2126–2132.
32. Winn,M.D., Murshudov,G.N. and Papiz,M.Z. (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.*, **374**, 300–321.
33. Afonine,P.V., Grosse-Kunstleve,R.W., Echols,N., Headd,J.J., Moriarty,N.W., Mustyakimov,M., Terwilliger,T.C., Urzhumtsev,A., Zwart,P.H. and Adams,P.D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D. Biol. Crystallogr.*, **68**, 352–367.
34. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
35. Gong,S., Zhang,C. and Zhang,Y. (2019) RNA-align: quick and accurate alignment of RNA 3D structures based on size-independent TM-scoreRNA. *Bioinformatics*, **35**, 4459–4461.

36. Wower,I.K., Zwieb,C.W., Guven,S.A. and Wower,J. (2000) Binding and cross-linking of tmRNA to ribosomal protein S1, on and off the Escherichia coli ribosome. *EMBO J.*, **19**, 6612–6621.
37. Varshney,U., Lee,C.P. and RajBhandary,U.L. (1991) Direct analysis of aminoacylation levels of tRNAs in vivo. Application to studying recognition of Escherichia coli initiator tRNA mutants by glutaminyl-tRNA synthetase. *J. Biol. Chem.*, **266**, 24712–24718.
38. Cheng,Y. and Prusoff,W.H. (1973) Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem. Pharmacol.*, **22**, 3099–3108.
39. Bullock,T.L., Uter,N., Nissan,T.A. and Perona,J.J. (2003) Amino acid discrimination by a class I aminoacyl-tRNA synthetase specified by negative determinants. *J. Mol. Biol.*, **328**, 395–408.
40. Ishikawa,F. and Kakeya,H. (2014) Specific enrichment of nonribosomal peptide synthetase module by an affinity probe for adenylation domains. *Bioorg. Med. Chem. Lett.*, **24**, 865–869.
41. Krutyholowa,R., Zakrzewski,K. and Glatt,S. (2019) Charging the code - tRNA modification complexes. *Curr. Opin. Struct. Biol.*, **55**, 138–146.
42. Boccaletto,P., Machnicka,M.A., Purta,E., Piatkowski,P., Baginski,B., Wirecki,T.K., de Crecy-Lagard,V., Ross,R., Limbach,P.A., Kotter,A. *et al.* (2018) MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.*, **46**, D303–D307.
43. de Crecy-Lagard,V. and Jaroch,M. (2020) Functions of Bacterial tRNA modifications: from ubiquity to diversity. *Trends Microbiol.*, **29**, 41–53.
44. Yanas,A. and Liu,K.F. (2019) RNA modifications and the link to human disease. *Methods Enzymol.*, **626**, 133–146.
45. Bernier,S., Dubois,D.Y., Habegger-Polomat,C., Gagnon,L.P., Lapointe,J. and Chenevert,R. (2005) Glutamylsulfamoyladenosine and pyroglutamylsulfamoyladenosine are competitive inhibitors of E. coli glutamyl-tRNA synthetase. *J. Enzyme Inhib. Med. Chem.*, **20**, 61–67.
46. Forrest,A.K., Jarvest,R.L., Mensah,L.M., O'Hanlon,P.J., Pope,A.J. and Sheppard,R.J. (2000) Aminoalkyl adenylate and aminoacyl sulfamate intermediate analogues differing greatly in affinity for their cognate Staphylococcus aureus aminoacyl tRNA synthetases. *Bioorg. Med. Chem. Lett.*, **10**, 1871–1874.
47. Rath,V.L., Silvian,L.F., Beijer,B., Sproat,B.S. and Steitz,T.A. (1998) How glutaminyl-tRNA synthetase selects glutamine. *Structure*, **6**, 439–449.
48. Bovee,M.L., Yan,W., Sproat,B.S. and Francklyn,C.S. (1999) tRNA discrimination at the binding step by a class II aminoacyl-tRNA synthetase. *Biochemistry*, **38**, 13725–13735.
49. Cusack,S., Yaremchuk,A. and Tukalo,M. (1996) The crystal structure of the ternary complex of *T. thermophilus* seryl-tRNA synthetase with tRNA(Ser) and a seryl-adenylate analogue reveals a conformational switch in the active site. *EMBO J.*, **15**, 2834–2842.
50. Cusack,S., Yaremchuk,A. and Tukalo,M. (1996) The crystal structures of *T. thermophilus* lysyl-tRNA synthetase complexed with *E. coli* tRNA(Lys) and a *T. thermophilus* tRNA(Lys) transcript: anticodon recognition and conformational changes upon binding of a lysyl-adenylate analogue. *EMBO J.*, **15**, 6321–6334.
51. Moulinier,L., Eiler,S., Eriani,G., Gangloff,J., Thierry,J.C., Gabriel,K., McClain,W.H. and Moras,D. (2001) The structure of an AspRS-tRNA(Asp) complex reveals a tRNA-dependent control mechanism. *EMBO J.*, **20**, 5290–5301.
52. Perona,J.J. and Hou,Y.M. (2007) Indirect readout of tRNA for aminoacylation. *Biochemistry*, **46**, 10419–10432.
53. Rould,M.A., Perona,J.J., Soll,D. and Steitz,T.A. (1989) Structure of *E. coli* glutaminyl-tRNA synthetase complexed with tRNA(Gln) and ATP at 2.8 A resolution. *Science*, **246**, 1135–1142.
54. Rould,M.A., Perona,J.J. and Steitz,T.A. (1991) Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. *Nature*, **352**, 213–218.
55. Goldgur,Y., Mosyak,L., Reshetnikova,L., Ankilova,V., Lavrik,O., Khodyreva,S. and Safro,M. (1997) The crystal structure of phenylalanyl-tRNA synthetase from thermus thermophilus complexed with cognate tRNAPhe. *Structure*, **5**, 59–68.
56. Moor,N.A., Ankilova,V.N., Lavrik,O.I. and Favre,A. (2001) Determination of tRNA(Phe) nucleotides contacting the subunits of Thermus thermophilus phenylalanyl-tRNA synthetase by photoaffinity crosslinking. *Biochim. Biophys. Acta*, **1518**, 226–236.
57. Desogus,G., Todone,F., Brick,P. and Onesti,S. (2000) Active site of lysyl-tRNA synthetase: structural studies of the adenylation reaction. *Biochemistry*, **39**, 8418–8425.
58. Sekine,S., Nureki,O., Dubois,D.Y., Bernier,S., Chenevert,R., Lapointe,J., Vassylyev,D.G. and Yokoyama,S. (2003) ATP binding by glutamyl-tRNA synthetase is switched to the productive mode by tRNA binding. *EMBO J.*, **22**, 676–688.
59. Eriani,G., Delarue,M., Poch,O., Gangloff,J. and Moras,D. (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature*, **347**, 203–206.
60. Loftfield,R.B., Eigner,E.A. and Pastuszyn,A. (1981) The role of spermine in preventing misacylation by phenylalanyl-tRNA synthetase. *J. Biol. Chem.*, **256**, 6729–6735.
61. Carter,C.W. Jr (2014) Urzymology: experimental access to a key transition in the appearance of enzymes. *J. Biol. Chem.*, **289**, 30213–30220.
62. Carter,C.W. Jr and Wills,P.R. (2019) Experimental solutions to problems defining the origin of codon-directed protein synthesis. *Biosystems*, **183**, 103979.
63. Schimmel,P., Giege,R., Moras,D. and Yokoyama,S. (1993) An operational RNA code for amino acids and possible relationship to genetic code. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 8763–8768.
64. Sampson,J.R., Behlen,L.S., DiRenzo,A.B. and Uhlenbeck,O.C. (1992) Recognition of yeast tRNA(Phe) by its cognate yeast phenylalanyl-tRNA synthetase: an analysis of specificity. *Biochemistry*, **31**, 4161–4167.
65. Peterson,E.T. and Uhlenbeck,O.C. (1992) Determination of recognition nucleotides for Escherichia coli phenylalanyl-tRNA synthetase. *Biochemistry*, **31**, 10380–10389.
66. Fishman,R., Ankilova,V., Moor,N. and Safro,M. (2001) Structure at 2.6 A resolution of phenylalanyl-tRNA synthetase complexed with phenylalanyl-adenylate in the presence of manganese. *Acta Crystallogr. D. Biol. Crystallogr.*, **57**, 1534–1544.
67. Moor,N., Kotik-Kogan,O., Tworowski,D., Sukhanova,M. and Safro,M. (2006) The crystal structure of the ternary complex of phenylalanyl-tRNA synthetase with tRNAPhe and a phenylalanyl-adenylate analogue reveals a conformational switch of the CCA end. *Biochemistry*, **45**, 10572–10583.
68. Klipcan,L., Moor,N., Kessler,N. and Safro,M.G. (2009) Eukaryotic cytosolic and mitochondrial phenylalanyl-tRNA synthetases catalyze the charging of tRNA with the meta-tyrosine. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 11045–11048.
69. Moor,N., Klipcan,L. and Safro,M.G. (2011) Bacterial and eukaryotic phenylalanyl-tRNA synthetases catalyze misaminoacylation of tRNA(Phe) with 3,4-dihydroxy-L-phenylalanine. *Chem. Biol.*, **18**, 1221–1229.
70. Tworowski,D., Klipcan,L., Peretz,M., Moor,N. and Safro,M.G. (2015) Universal pathway for posttransfer editing reactions: insights from the crystal structure of TtPheRS with puromycin. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 3967–3972.
71. Abibi,A., Ferguson,A.D., Fleming,P.R., Gao,N., Hajec,L.I., Hu,J., Laganas,V.A., McKinney,D.C., McLeod,S.M., Prince,D.B. *et al.* (2014) The role of a novel auxiliary pocket in bacterial phenylalanyl-tRNA synthetase druggability. *J. Biol. Chem.*, **289**, 21651–21662.
72. Klipcan,L., Levin,I., Kessler,N., Moor,N., Finarov,I. and Safro,M. (2008) The tRNA-induced conformational activation of human mitochondrial phenylalanyl-tRNA synthetase. *Structure*, **16**, 1095–1104.
73. Kartvelishvili,E., Tworowski,D., Vernon,H., Moor,N., Wang,J., Wong,L.J., Chrzanowska-Lightowlers,Z. and Safro,M. (2017) Kinetic and structural changes in HsmtPheRS, induced by pathogenic mutations in human FARS2. *Protein Sci.*, **26**, 1505–1516.
74. Needleman,S.B. and Wunsch,C.D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, **48**, 443–453.
75. Karplus,P.A. and Diederichs,K. (2012) Linking crystallographic model and data quality. *Science*, **336**, 1030–1033.
76. Davis,I.W., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2004) MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.*, **32**, W615–W619.