Supplementary material:

Crystal structures of the tRNA:m^{2}G6 methyltransferase Trm14/TrmN from two domains of life

Marcus Fislage, Martine Roovers, Irina Tuszynska, Janusz M. Bujnicki, Louis Droogmans, Wim Versées

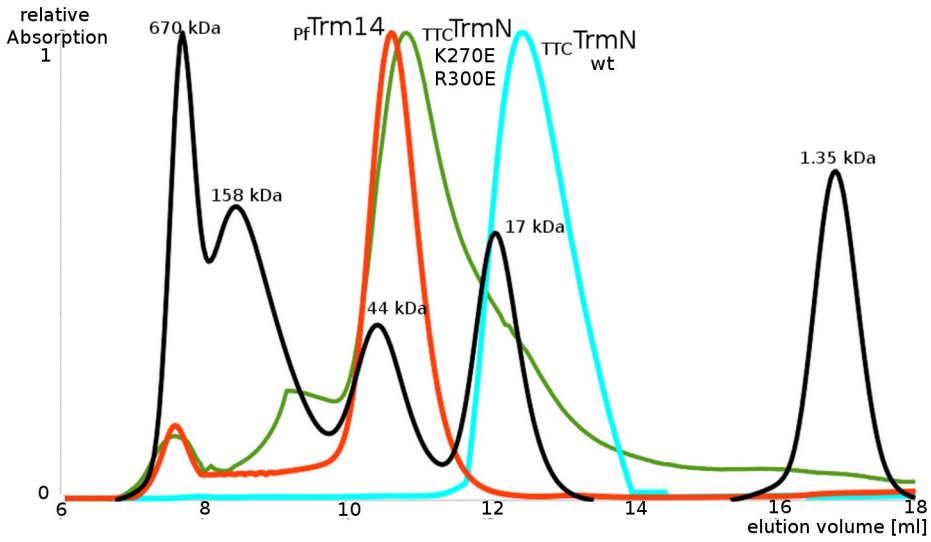

Figure S1: Size exclusion chromatography of $_{Pf}$Trm14 (red), $_{TTC}$TrmN wt (blue) and $_{TTC}$TrmN K270E/R300E (green) as compared to a molecular weight marker (black). 0.5 mg of $_{Pf}$Trm14 or $_{TTC}$TrmN were loaded on a Superdex 75 10/300 GL column run at 0.5 ml/min in a buffer containing 50 mM Tris pH8, 500 mM NaCl, 1 mM DTT. $_{Pf}$Trm14 eluted at a volume corresponding to a monomer (expected molecular weight: 42 kDa). However, $_{TTC}$TrmN wt elutes delayed (calculated molecular weight: 15 kDa; expected molecular weight: 39 kDa). The delay of the wt $_{TTC}$TrmN on the Superdex column might be due to the highly positive charged character of the protein, leading to electrostatic interactions with the column matrix. Indeed, the variation of two positively charged residues to negatively charged residues in the TTCTrmN K270E/R300E variant led to an elution volume corresponding to a monomer (calculated molecular weight: 41 kDa). This observation, in combination with the analogy with $_{Pf}$Trm14 and analysis of the crystal structure strongly suggests that also $_{TTC}$TrmN is a monomer in solution.

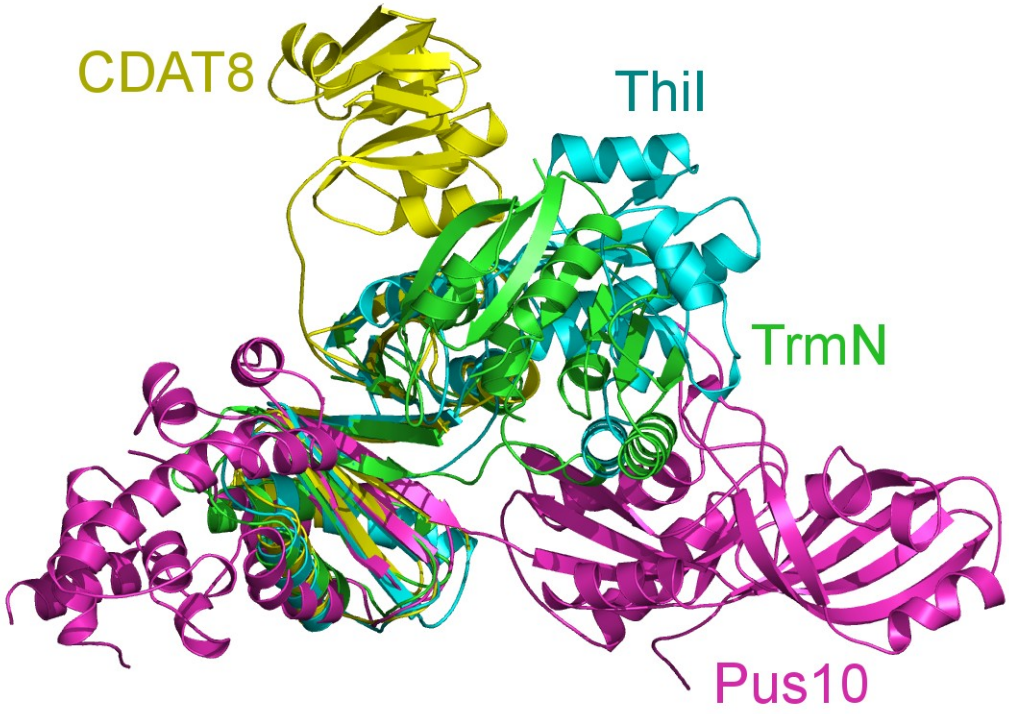

Figure S2: Structural alignment of $_{TTC}$TrmN (green), based on its core-THUMP subdomain, with the three THUMP domain-containing tRNA modifying enzymes: the deaminase CDAT8 (yellow), the uridine synthase Pus10 (pink) and the s^4U sulfurtransferase ThiI (blue). It should be noted that CDAT8 and ThiI contain a core-THUMP subdomain fused to a NFLD subdomain (comparable to $_{TTC}$TrmN), while Pus10 lacks the NFLD subdomain. The orientation of the catalytic domains of CDAT8 and Pus10 toward their THUMP domains is unrelated to the relative orientation of the RFM and THUMP domains of $_{TTC}$TrmN. In contrast, in ThiI the relative orientation of the catalytic domain toward the THUMP domain is similar to $_{TTC}$TrmN.

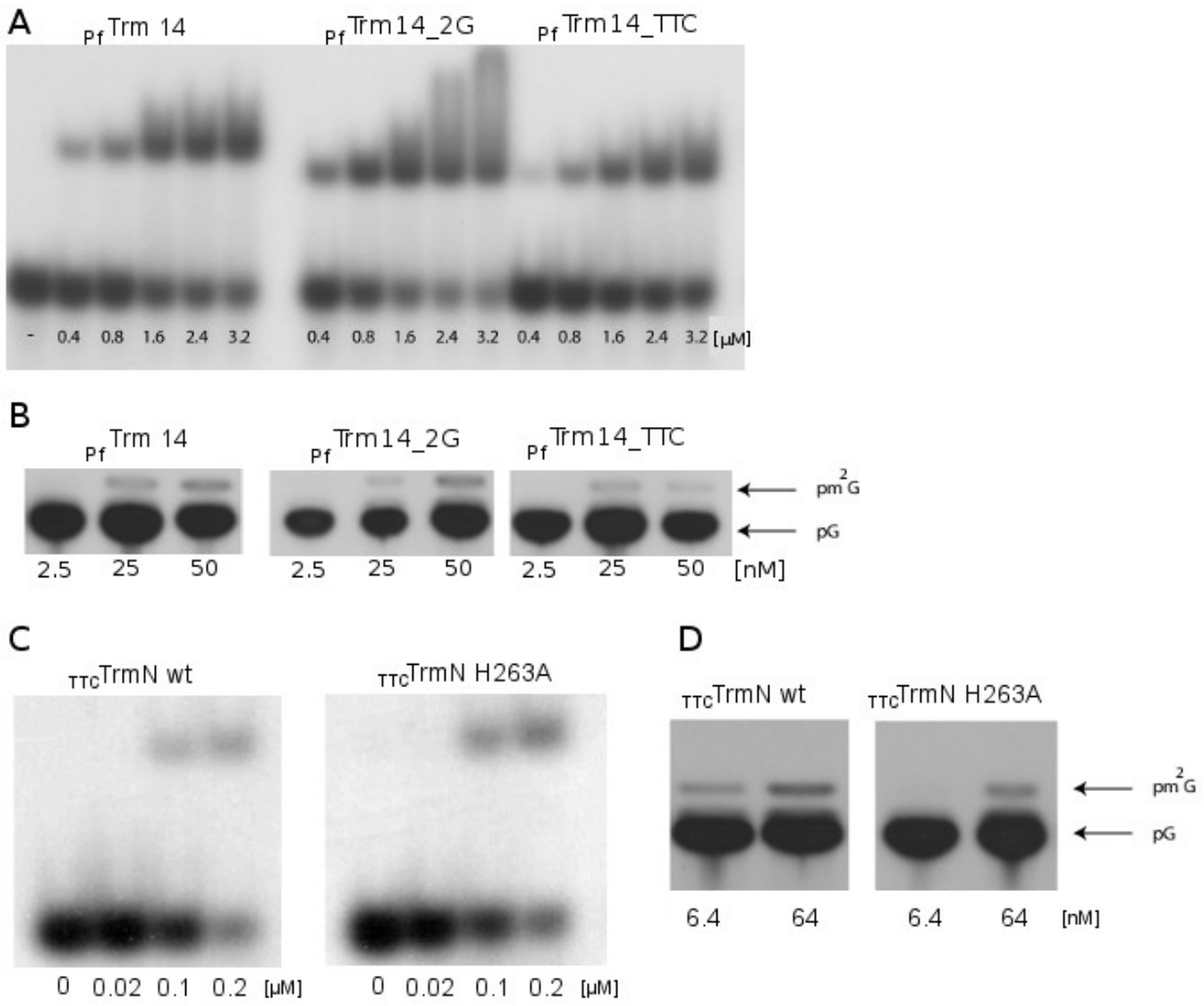

Figure S3: A. Band shift assay showing binding of [α-^{32}P] GTP-labeled *T. thermophilus* tRNA^{Phe} transcripts to increasing amounts of wt $_{Pf}$Trm14, Trm14_2G ($_{Pf}$Trm14 with the β-sheet insertion replaced by two glycines) and Trm14_TTC ($_{Pf}$Trm14 with the β-sheet insertion replaced by the corresponding loop sequence of $_{TTC}$TrmN). Bound and unbound tRNA^{Phe} were separated by 6% PAGE. **B.** MTase activity measurement of wt $_{Pf}$Trm14, Trm14_2G and Trm14_TTC variants by observing the m^{2}G production on a 1D TLC using varying concentrations of protein. **C**. Band shift assay showing binding of [α-^{32}P] GTP-labeled *T. thermophilus* tRNA^{Phe} transcripts to wt $_{TTC}$TrmN and the active site variant H263A. **D.** MTase activity measurement of wt $_{TTC}$TrmN compared to $_{TTC}$TrmN H263A.

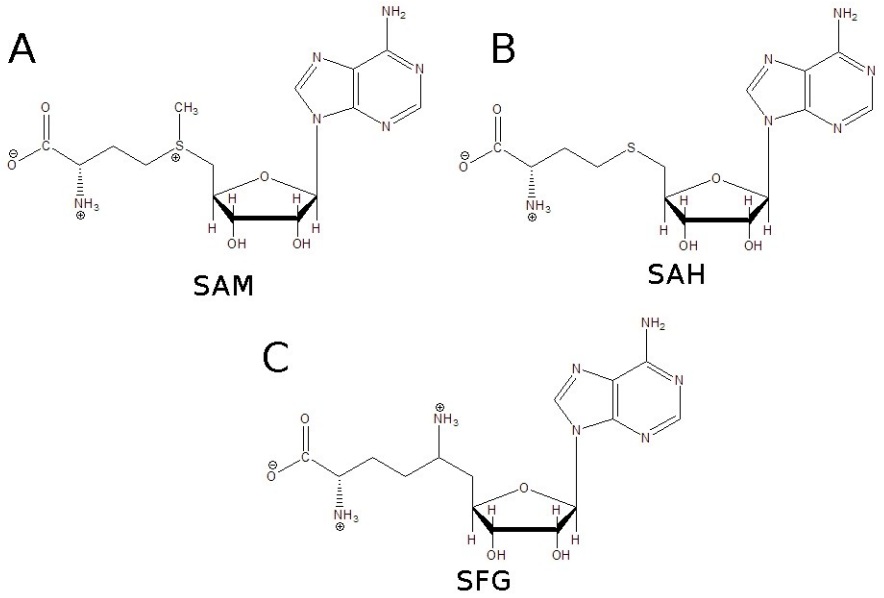

Figure S4: Chemical structure of the ligands of $_{Pf}$Trm14 shown in Figure 3. In the SAM (A) and SAH (B) bound state the loop region near the active site, connecting β-strand 4 and helix 5, is highly flexible and could not be observed in the electron density of the crystal structures. Only in the SFG (C) bound state became this loop visible in the density map. Although SAM and SFG are expected to have the positive charge at roughly the same position, only SFG seems to evoke this conformational change. This might be due to the position of the positive charge, which is located on the S^{γ} in SAM and on the N^{ε} in SFG.

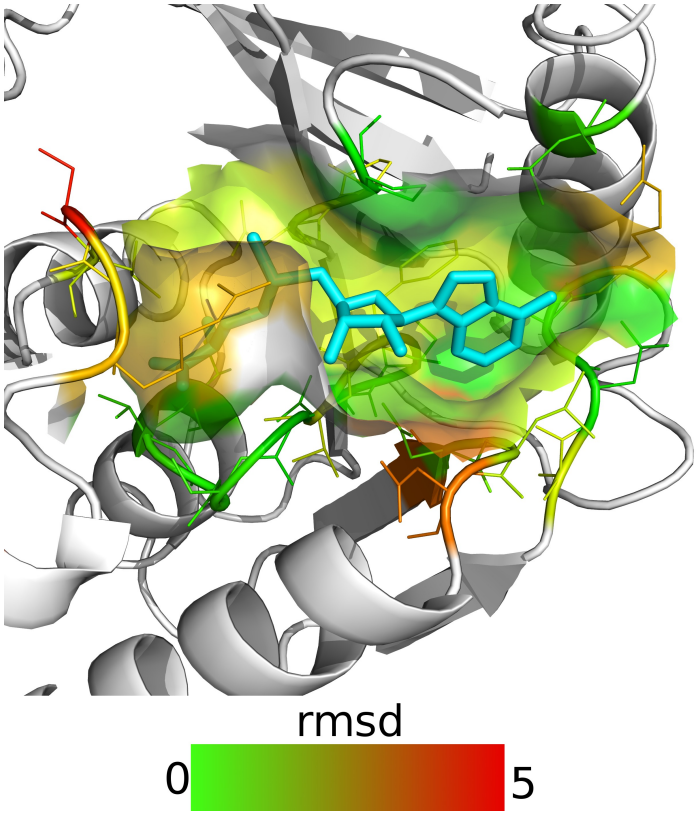

Figure S5: Modeling (epitope matching) of SAM into the active site (SAM-binding) pocket of $_{TTC}$TrmN, using the program Epitope Match. The structure of $_{Pf}$Trm14 in complex with SAM was used as the epitope for the SAM-binding pocket.
The accessible surface area around the active site is shown and colored according to the rmsd between the epitope residues of the model ($_{Pf}$Trm14-SAM) and the corresponding $_{TTC}$TrmN epitope residues. The epitope residues are shown in line representation and the modeled SAM is shown as blue sticks. In the $_{TTC}$TrmN structure the SAM pocket is clearly visible and chemically related residues compared to $_{Pf}$Trm14 are involved in the binding of SAM.

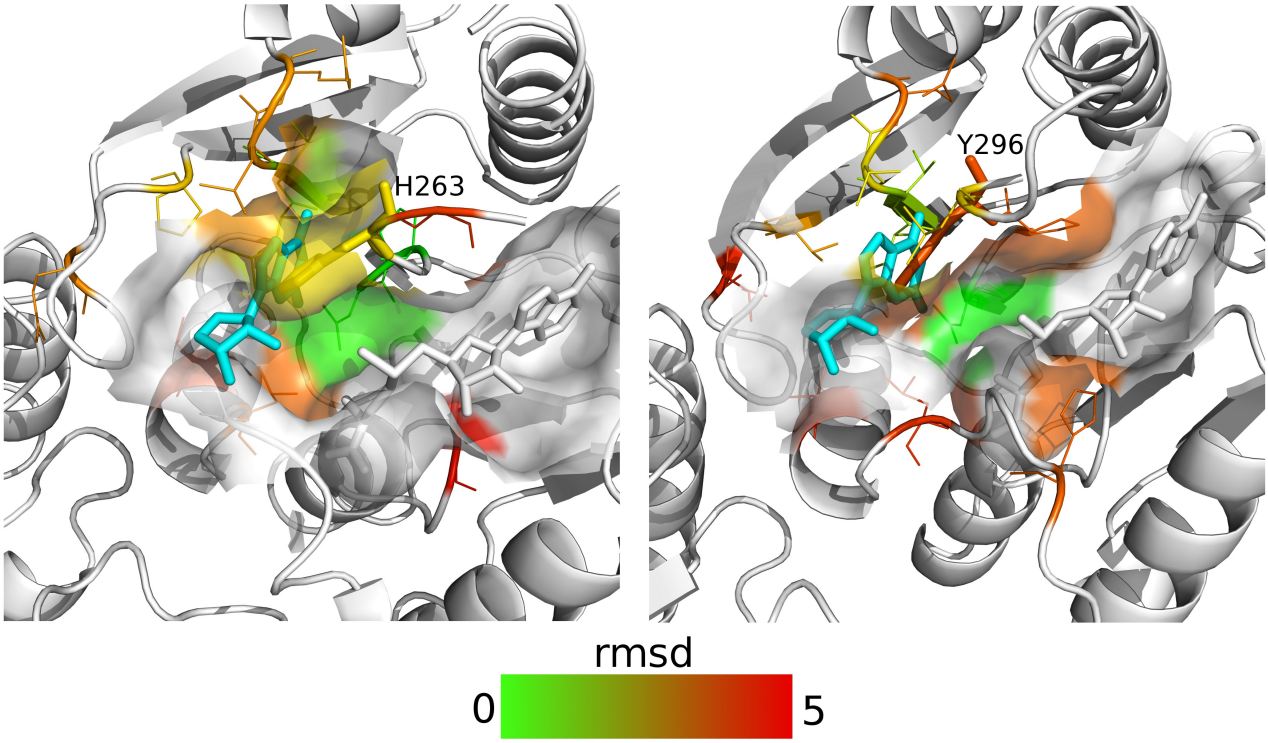

Figure S6: Modeling of a guanosine (corresponding to G6) into the active site pockets of $_{TTC}$TrmN and $_{Pf}$Trm14, using the program Epitope Match. The structure of RsmC in complex with guanosine and SAM (PDB 3DMH) was used as the epitope for the guanosine-binding pocket. **A.** Modeling (epitope matching) of the guanosine binding pocket of RsmC onto the structure of $_{TTC}$TrmN, modeled with bound SAM. **B.** Modeling of the guanosine-binding pocket of RsmC onto the structure of $_{Pf}$Trm14-SAM.
The accessible surface area around the active site is shown and colored according to the rmsd between the epitope residues of the model (RsmC) and the corresponding $_{TTC}$TrmN/$_{Pf}$Trm14 epitope residues. The epitope residues are shown in line representation, SAM is shown as white sticks and the modeled guanosine as blue sticks. In the $_{TTC}$TrmN structure the guanosine pocket is clearly visible and only a small movement of H263 is necessary for accommodating the guanosine of the tRNA. In the case of $_{Pf}$Trm14 the guanosine-binding pocket is not immediately visible since Y296 is blocking the pocket. A movement of the latter residue upon tRNA binding would be necessary to free this pocket.

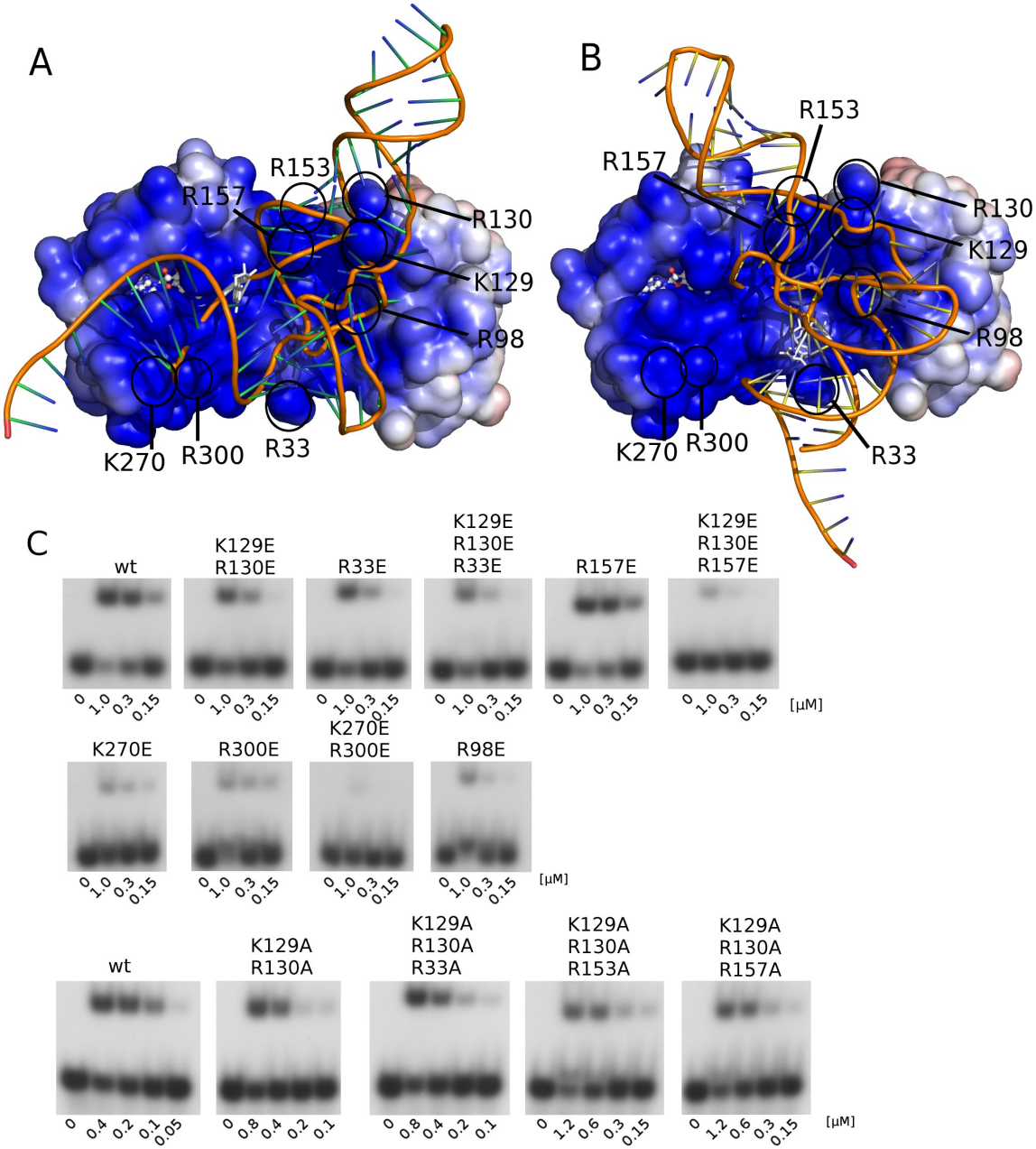

Figure S7: A&B. Results of computational docking of the $_{TTC}$TrmN-tRNA^{Phe} complex. $_{TTC}$TrmN is shown with the electrostatics drawn on the solvent accessibility surface, tRNA (a homology model based on the orthologous tRNA^{Phe} from *E. coli*) is shown in orange, the SAM molecule is shown in ball and stick representation and the target nucleoside G6 is shown in grey. Residues that upon mutation (either as a single or as part of a double/triple substitution) result in protein variants with lowered affinity for tRNA^{Phe} are indicated by black circles. **A.** Highest scoring docking model (DARS-RNP score = -3.7M) of the best-scored (and functionally more convincing) cluster (mean score = -2.7M for 15 models). The target nucleotide (G6) is positioned close to the active site, suitable for a flip out mechanism to bind in the active site of $_{TTC}$TrmN. **B.** Best-scored docking model (DARS-RNP score = -3.5M) of a cluster with an alternative binding mode (mean score = -2.3M for 25 models). In this cluster the target nucleotide is less appropriately oriented toward the active site. **C.** Band shift assay of $_{TTC}$TrmN variants that showed an impaired binding of tRNA^{Phe}. The K270E, R300E and K270E/R300E variants show a clear decreased affinity towards tRNA^{Phe}. These residues are only in proximity of the tRNA in the docking model shown in (A), supporting this model as the biological relevant one.

Table S1: Oligonucleotides used in this study

Oligo	Sequence	Purpose
Gly-1	CTC ATT AAG GTA AGT TGA AAT GCT CAA CCC ACC AGC TTT TTC AGC TTC AAG TAG C	Variant of the β-sheet in $_{Pf}$Trm14_2G
Gly-2	GCT ACT TGA AGC TGA AAA AGC TGG TGG GTT GAG CAT TTC AAC TTA CCT TAA TGA G	
TTC-1	GAA TTC TCA TTA AGG TAA GTT GAA ATG CTA CCT ACC CAT GCA CGT TCA GCT TCA AGT AGC AGC CTT C	Variant of the β-sheet in $_{Pf}$Trm14_TTC
TTC-2	GAA GGC TGC TAC TTG AAG CTG AAC GTG CAT GGG TAG GTA GCA TTT CAA CTT ACC TTA ATG AGA ATT C	
R33A-1	GCG GAG GTG GAC GCC GCG AAG GGC CGG GTC CGC	R33A mutation in $_{TTC}$TrmN
R33A-2	GCG GAC CCG GCC CTT CGC GGC GTC CAC CTC CGC	
R33E-1	GCG GAG GTG GAC GCC GAG AAG GGC CGG GTC CGC	R33E mutation in $_{TTC}$TrmN
R33E-2	GCG GAC CCG GCC CTT CTC GGC GTC CAC CTC CGC	
R98E-1	CCG CGT GGA GGC CGA GAG GGA GGG GGA ACA CC	R98E mutation in $_{TTC}$TrmN
R98E-2	GGT GTT CCC CCT CCC TCT CGG CCT CCA CGC GG	
KR-AA-1	CGG CGT GCC CGT GGA CCT GGC GGC CCC GGC CGT GCG GGT CCG G	K129A/R130A mutation in $_{TTC}$TrmN
KR-AA-2	CCG GAC CCG CAC GGC CGG GGC CGC CAG GTC CAC GGG CAC GCC G	
KR-EE-1	CGG CGT GCC CGT GGA CCT GGA GGA GCC GGC CGT GCG GGT CCG G	K129E/R130E mutation in $_{TTC}$TrmN
KR-EE-2	CCG GAC CCG CAC GGC CGG CTC CTC CAG GTC CAC GGG CAC GCC G	
R153A-1	GTC CAG CTC ACG GAA GCG CCC CTC TCC CGC CGC	R153A mutation in $_{TTC}$TrmN
R153A-2	GCG GCG GGA GAG GGG CGC TTC CGT GAG CTG GAC	
R157A-1	GAA AGG CCC CTC TCC GCC CGC TTC CCC AAG GCG	R157A mutation in $_{TTC}$TrmN
R157A-2	CGC CTT GGG GAA GCG GGC GGA GAG GGG CCT TTC	
R157E-1	GAA AGG CCC CTC TCC GAG CGC TTC CCC AAG GCG	R157E mutation in $_{TTC}$TrmN
R157E-2	CGC CTT GGG GAA GCG CTC GGA GAG GGG CCT TTC	
H263A-1	GAT CCT CGC CAA CCC GCC CGC CGG CCT CCG CCT GGG CCG G	H263A mutation in $_{TTC}$TrmN
H263A-2	CCG GCC CAG GCG GAG GCC GGC GGG CGG GTT GGC GAG GAT C	
K270E-1	CCG CCT GGG CCG GGA GGA GGG GCT TTT CC	K270E mutation in $_{TTC}$TrmN
K270E-2	GGA AAA GCC CCT CCT CCC GGC CCA GGC GG	
R300E-1	GCC CTC CTC ACC CTG GAG CCC GCC CTC CTC AAG CGG GC	R300E mutation in $_{TTC}$TrmN
R300E-2	GCC CGC TTG AGG AGG GCG GGC TCC AGG GTG AGG AGG GC	

Table S2: Results of structural similarity searches by PDBeFold using the default minimal number of overlapping secondary structure elements (40%). Matches are sorted by Z-score. The two last rows describe pairwise comparisons of RNA-modifying enzymes known to contain the THUMP, which were not detected by the search of the PDB database using the default threshold values.

match number	PDB id	organism	protein	PFAM Classification	Z-score	RMSD [Å]	alignment length	Seq id [%]
				NFLD + THUMP + RFM				
1	3k0b	*Listeria monocytogenes*	Putative N6-adenine-specific DNA methylase	Putative RNA methylase family UPF0020	12.9	2.27	286	23
2	3ldu	*Clostridium difficile*	Putative methylase	Putative RNA methylase family UPF0020	12.7	2.08	287	23
3	3ldg	*Streptococcus mutans*	Putative methyltransferase	Putative RNA methylase family UPF0020	12.7	2.24	292	20
				RFM				
1	3d2l	*Exiguobacterium sibiricum*	SAM-dependent methyltransferase	Methyltransferase domain	11.6	1.80	147	26
2	3ldg	*Streptococcus mutans*	Putative methyltransferase	Putative RNA methylase family UPF0020	11.4	1.69	163	20
3	3k0b	*Listeria monocytogenes*	Putative N6-adenine-specific DNA methylase	Putative RNA methylase family UPF0020	11.0	1.93	152	26
4	1ve3	*Pyrococcus horikoshii*	Putative PH0226 protein	Methyltransferase domain	10.5	1.93	154	18
5	3ggd	*Anabaena variabilis*	Putative SAM-dependent methyltransferase	Methyltransferase domain	10.3	2.24	144	16
6	3iv6	*Rhodobacter sphaeroides*	Putative Zn-dependent Alcohol Dehydrogenase	Methyltransferase domain	10.3	2.04	147	14
7	3g2q	*Amycolatopsis orientalis*	glycopeptide N-methyltransferase MtfA	Methyltransferase domain	10.1	1.79	145	14
				NFLD + THUMP				
1	3ldu	*Clostridium difficile*	Putative methylase	Putative RNA methylase family UPF0020	8.1	2.45	133	17
2	1vbk	*Pyrococcus horikoshii*	Putative thiamine-biosynthesis protein PH1313	THUMP domain	8.0	2.31	132	18
3	2dir	*Homo sapiens*	THUMP domain-containing protein 1	THUMP domain	6.3	1.86	75	20
4	2c5s	*Bacillus anthracis*	Putative thiamine biosynthesis protein ThiI	THUMP domain	5.7	2.76	139	17
…	3g8q	*Methanopyrus kandleri*	Putative RNA-binding protein	THUMP domain	6.5	2.62	133	23
…	2v9k	*Homo sapiens*	PUS10	Pseudouridine synthase	3.3	2.70	58	9