Alteration in H-bond strength affects the stability of codon-anticodon interaction at in-frame UAG stop codon during *in vitro* translation

Purnima Mala and Ishu Saraogi*

Department of Chemistry, Indian Institute of Science Education and Research Bhopal, M.P 462066, INDIA

ABSTRACT

We have studied the decoding ability of a non-standard nucleobase modified tRNA for non-natural amino acid mutagenesis. The insertion of 2, 6-diaminopurine (D) base at the 3rd position of a tRNA anticodon enabled us to evaluate the effect of an additional hydrogen bond during translation. The presence of D at the tRNA anticodon led to stabilization of the codon-anticodon interaction due to an additional H-bond between the N^2-exocyclic amine of D and the C2 carbonyl group of uracil during protein translation. While decoding UAG codons using stop codon suppression methodology, the enhanced codon-anticodon interaction improved codon readthrough and synthesis of modified protein with a non-natural amino acid at multiple sites. Our findings imply that the number of hydrogen bonds at the tRNA-mRNA duplex interface is an important criterion during mRNA decoding and improves protein translation at multiple UAG stop sites. This work provides valuable inputs towards improved non-natural amino acid mutagenesis for creating functional proteins.

KEYWORDS

Codon-anticodon interaction**,** amber suppression, release factor 1, non-canonical amino acid, suppressor tRNA, 2, 6-diaminopurine

INTRODUCTION

Protein translation is a highly dynamic process that decodes genetic information with the help of transfer RNA (tRNA) that act as adapter molecules[1]. A key step during translation is the codon-anticodon pairing which occurs in the 30S subunit of the ribosome[2]. The strength of such codon-anticodon interaction is dependent on composition of base pairs and identity of neighboring nucleotides of tRNA anticodon stem loop (ASL) and mRNA codons (mRNA context). The presence of an array of posttranscriptional modifications in tRNA contributes to the accuracy and speed of translation by stabilizing the interaction between codon and anticodon[3,4,5]. The three nucleotide codon-anticodon duplex is stabilized by ionic interactions, base stacking and hydrogen bonds[6]. Codon–anticodon recognition between the first and second bases of mRNA codons is dominated by Watson-Crick base pairing[7], while the interaction between the third base of mRNA codon occurs by flexible non-Watson–Crick (wobble) base pairing[8].

The hydrogen bonding interaction at the codon-anticodon interface is an important criterion for efficiency and accuracy during translation. Experiments suggest that the loss of a single hydrogen at the duplex interface provides a difference in the free energy value of 20 kJ/mol or more which is sufficient to distinguish between correct and errant codon reading[6]. The hydrogen bond between the N1 of purines and the N3 of pyrimidines (Fig. 1) is important for decoding the first two codons, but sufficient stacking between the RNA bases is critical at the wobble position[9]. Loss or gain of H-bond at codon-anticodon interface results in alteration of codon-anticodon pairing rules in ribosome. There are efforts made by altering codon-

anticodon pairing and utilizing the decoding property of tRNA to reprogram the genetic code for non-natural amino acid (NAA) mutagenesis[10,11].

NAA mutagenesis allows incorporation of NAAs into proteins during ribosomal translation using engineered aminoacyl tRNA synthetase (aaRS)-tRNA pair. One of the most successful methods for NAA mutagenesis is stop codon reassignment[12]. In this method, NAAs are incorporated into proteins at an in frame UAG stop codon using orthogonal *Mj*tRNA$^{Tyr}_{CUA}$/aaRS pair *(Methanocaldococcus jannaschii* tRNA$^{Tyr}_{CUA}$/aminoacyl tRNA synthetase)[12–15]. This method requires an orthogonal tRNA and synthetase pair that charges the orthogonal tRNA with the desired NAA, and the acylated tRNA incorporates the attached NAA into proteins in response to the unique codon by utilizing the endogenous translational machinery. However, this method suffers from low yields of modified protein of interest (POI) and inability to add multiple NAAs at defined UAG sites. This limitation is a result of competition between orthogonal tRNA and release factor (RF1) for binding to the UAG stop codon, resulting in reduced yield of modified proteins[14–16]. Also, variability in modified protein yield has been reported from protein to protein, and at different sites on the same protein during stop codon reassignment. These have been rationalized by factors like expression system composition (concentration of suppressor tRNA and RF1) and mRNA context effects [17,18].

We reasoned that modifying the tRNA anticodon bases with exocyclic functional groups to enhance codon-anticodon pairing might improve the decoding efficiency of *Mj*tRNA$^{Tyr}_{CUA}$ as it might be able to more effective compete with endogenous components (like RF1 or endogenous tRNA) during NAA mutagenesis. To test this idea, we created a tRNA analog from *Mj*tRNA$^{Tyr}_{CUA}$ by replacing adenine with 2, 6-diaminopurine (D) (Fig. 1) at the anticodon stem loop (ASL). Earlier reports suggest that all positions of tRNA anticodon (bases 34, 35 and 36) show equal proportion of all the four nucleotides, and very limited nucleotide modification at positions 35 and 36[19,20]. It has been suggested that RF1 recognizes the first base (U) of the UAG stop codon by more stringent hydrogen bonding than second (A) and third (G) nucleotide[21]. However, the efficiency of RF1 competition with anticodon modified tRNA has not been reported. Since a large fraction of the codon-anticodon binding energy is provided by the first two codon bases (non-wobble positions), we expected a substantial contribution from the presence of D[22] at the third position of tRNA which pairs with the first mRNA codon (U).

We created a modified *Mj*tRNA$^{Tyr}_{CUA}$ (called tRNA analog or tRNA$_{CUD}$) containing D at position 36 of tRNA ASL to check its ability to decode the UAG stop codon. The tRNA was assembled through three-way RNA fragment ligation. The synthesized tRNA, its cognate aaRS, and non-natural amino acid (7-hydroxycoumarin-4-yl)ethyl glycine or 7-HMC) were introduced into an *E. coli* based cell-free translation system[2324–27,28] to express 7-HMC modified chloramphenicol acetyl transferase (CAT) or GFP. We compared the decoding efficiency of unmodified and modified tRNA in the form of expressed protein yields by ^{14}C autoradiography. Our results indicate that the tRNA analog is specifically recognized at the ribosomal decoding site, resulting in improved efficiency of NAA mutagenesis, with most pronounced effect at constructs containing more than one UAG codon in the protein of interest. Thus, our synthesized tRNA analog enables a better understanding of (a) importance of an additional H-bond at codon-anticodon interface during translation (b) RF1 out-competition.

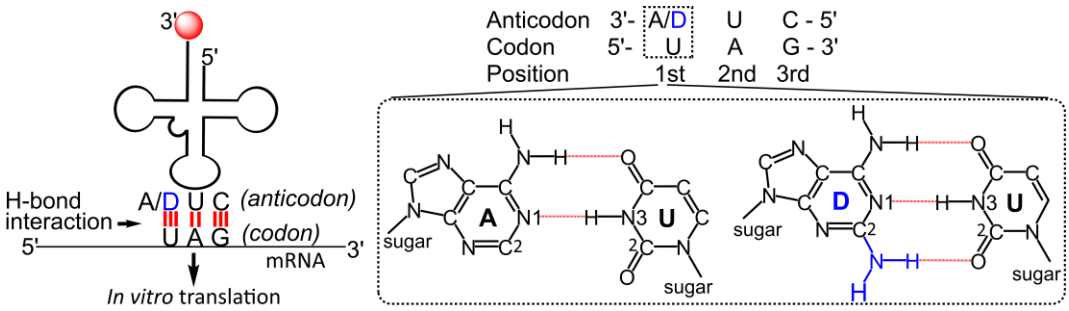

Fig. 1. H-bond mediated interaction between tRNA and mRNA at codon-anticodon interface. The right panel shows the hydrogen bonding pattern (red) of A-U (left) and D-U (right) base pairs. (D = 2, 6-diaminopurine)

MATERIALS AND METHODS

Design and synthesis of tRNA fragments containing non-natural base (D)

All DNA primers were obtained from IDT. All RNAs are synthesized in-house and concentrations were measured using NanoDrop spectrophotometer (Thermo). The tRNA analog ($tRNA_{CUD}$) is an anticodon modified version of o-$tRNA^{opt}$ (from *Mj*$tRNA^{Tyr}_{CUA}$)[29].The *Mj*$tRNA^{Tyr}_{CUA}$ sequence shown in bold was fused to hammerhead ribozyme sequence shown as plain text below[18]. The base position to be modified is underlined. As a result of the design, position 31 of the tRNA was also modified to D.

5'GGGAGACCGGCUGAUGAGUCCGUGAGGACGAAACGGUACCCGGUACCGUC**CCGGCGGU AGUUCAGCAGGGCAGAACGGCGGACUCUAAAUCCGCAUGGCAGGGGUUCAAAUCCCC UCCGCCGGACCA**3'[18].

The *Mj*$tRNA^{Tyr}_{CUA}$ DNA template for transcription was designed by fusing T7 promoter to ribozyme and t-DNA sequences to construct 'transzyme' template by overlapping PCR. The transzyme sequence encoding *Mj*$tRNA^{Tyr}_{CUA}$ was ligated into pUC19 vector using Roche rapid DNA ligation kit. The transzyme sequence containing plasmid was digested using BstNI enzyme that provides the correct CCA 3′ terminus. From this linearized plasmid, the T7 RNA polymerase transcribed the transzyme which subsequently cleaved itself to release the tRNA.

The o-$tRNA^{opt}$ transcript was produced from *in vitro* transcription using T7 RNA polymerase and subsequent ribozyme cleavage. The D-containing o-$tRNA^{opt}$ was constructed by splint-mediated ligation of RNA fragments named F1, F2 and F3. F1 (26nt) and F3 (40nt) templates were designed by fusing T7 promoter followed by hammerhead ribozyme and either F1 or F3 DNA sequences using overlapping PCR. The F1 and F3 DNA templates were transcribed *in vitro* in a solution containing 40mM Tris-HCl pH 7.9, 50mM DTT, 4mM NTPs, 30mM $MgCl_2$, 0.02% v/v Triton X-100, and 50µg/ml T7 RNAP. The transcription reaction was incubated at 37°C for 3 hr. The reaction was diluted 5-fold, $MgCl_2$ (30mM) and EDTA (0.2mM) were added and incubated at 60°C for 3 hours for ribozyme cleavage. Samples were analyzed on a 20% denaturing polyacrylamide gel (Fig. 2. A, B, E, F).

F2 RNA fragment (11nt) contained non-natural nucleobase D in the ASL. The F2 RNA fragment was synthesized from annealed double stranded T7 promoter strand and template strand. F2 RNA was synthesized *in vitro* in a solution that contained 40mM Tris-HCl pH 7.9, 50mM DTT, 0.5mM GTP, 0.5mM CTP, 0.5mM UTP and 0.5mM DTP (2,6-diaminopurine triphosphate, Jena Bioscience) or 0.5mM ATP (in control $tRNA_{CUA}$), 4mM GMP, 30mM $MgCl_2$, 0.02% v/v Triton X-100, and 50µg/ml T7 RNAP (Fig. 2. C, D). Modified F2 RNA (11nt) was analyzed on a 20% 7M denaturing polyacrylamide gel. The four transcripts produced by T7 RNA polymerase mediated transcription were F1, F2 (D-containing), F2 (unmodified) and F3. The RNA fragments were loaded on a 20% 7M denaturing polyacrylamide gels for extraction and purification. The gel purified RNA samples were characterized by mass spectrometry.

Processing of RNA fragments for mass spectrometry characterization

RNA fragments used for mass spectrometry were precipitated with ammonium acetate. All the RNAs (F1, F2 modified, F2 unmodified and F3) were used at 100ng/µL in 1:1 ratio with solution of 3-hydroxy picolinic acid (3-HPA) saturated in H_2O:ACN 50:50 (v:v), containing 10mg/ml diammonium citrate (DAC). 1µL of RNA was mixed with 1µL of the matrix solution, spotted on standard ground steel target and allowed to dry at room temperature. The samples were subjected to MALDI mass spectrometry using Bruker ultrafleXtreme in positive mode. For ESI-MS, the RNA samples are used at 200ng/µL in H_2O: ACN 50:50 (v:v) with 1% TFA in negative mode on Bruker MicroTOF QII instrument.

Processing of RNA fragments for *in vitro* ligation to construct tRNA analog

After ribozyme cleavage, F1 and F3 RNA transcripts have 5'-hydroxyl and 3'-hydroxyl groups. For T4 DNA ligase-based ligation each RNA fragment should bear 5'-monophosphate. Hence, F1 and F3 RNA fragments were 5'-monophosphorylated using T4 PNK (New England Biolabs).
The F2 fragment (modified or unmodified) contained either a 5'-monophosphate or 5'-triphosphate since the reaction contained GMP as an initiator of transcription. To ensure complete conversion of 5'-triphosphate to 5'-monophosphate, F2 RNA was treated with calf intestinal alkaline phosphatase (Sigma-Aldrich), and subsequently treated with T4 PNK for 5'-monophosphorylation.

Assembly of RNA fragments on DNA splint for *in vitro* ligation using T4 DNA ligase

DNA splint mediated three-way RNA ligation was used to construct $tRNA_{CUD}$. Enzymatic *in vitro* ligation was carried out in two steps. First, annealing of the RNA fragments to the DNA splint was performed in 10 mM TE buffer (pH7.4), followed by nick sealing of RNA fragments assembled on DNA splint using T4 DNA ligase. The sequence of DNA splint (41nt) is
5'CTGCCATGCGGATTTAGAGTCCGCCGTTCTGCCCTGCTGAA3'. For a 100 µL ligation, 15 µM F1, 10 µM modified or unmodified F2, 5 µM F3, and 2.5 µM DNA splint were mixed with 10mM TE buffer, denatured at 90°C for 2 minutes, followed by slow cooling to 37°C. To the annealed solution were added ATP containing 1x T4 DNA ligase buffer, 20U T4 DNA ligase (New England Biolabs), 5% PEG 8000 (New England Biolabs) or 1.3% polyvinyl alcohol (Sigma-Aldrich) up to a final volume of 100µL, and incubated at 37°C for 20 mins, followed by incubation at 30°C for 12 hr. The concentrations of the RNA fragments and the DNA splint were optimized for maximal yield of the desired ligated product. After ligation, 1U DNase I was directly added to the reaction and incubated for 10 minutes at 37°C to digest the DNA splint. The ligated sample was purified on 20% 7M urea PAGE (Fig. 3C). The purified ligated tRNA samples were dissolved in 10µL nuclease free water and concentration was measured by UV spectrophotometer. Typically, the concentration was of the order of 150ng/µL. The ligated sample was characterized by MALDI-TOF (Fig. 4A). The ligated tRNA samples were concentrated as required.

Refolding and secondary structure analysis of ligated tRNA

Ligated tRNA samples were folded in refolding solution containing varying amounts of $MgCl_2$ (1mM, 5mM, 10mM and 20mM). Maximum refolding was observed at 20mM as analyzed by CD spectroscopy. CD spectra were measured at 25°C and data were collected in the wavelength range from 200-300nm. For CD, 3µM ligated tRNA samples in 10mM Tris-HCl (pH 7.4) were denatured at 90°C, followed by addition of 20mM $MgCl_2$ and incubation for 1 hr at room temperature. It is important that in *E. coli* S30 extract system, the final $MgCl_2$ concentration should not exceed 3mM after adding refolded tRNA because excess $MgCl_2$ is known to inhibit *in vitro* translation.

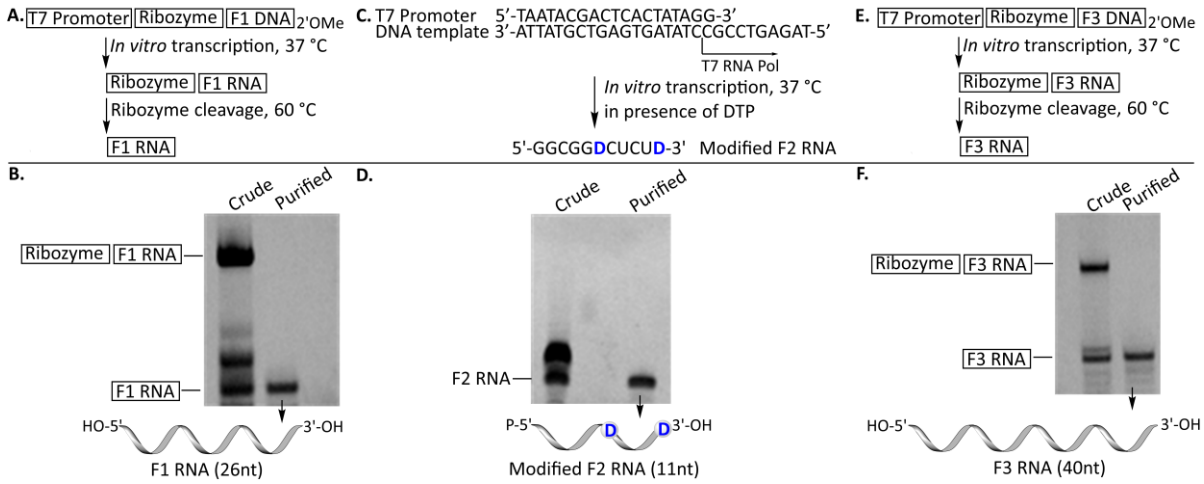

Fig. 2. Scheme showing synthesis of RNA fragments. **A.** Synthesis of F1 RNA**. B.** Analysis of purified *in vitro* transcribed F1 RNA. **C.** Design of DNA template for synthesis of F2 RNA. DTP = 2, 6-diaminopurine triphosphate **D.** Analysis of purified *in vitro* transcribed D containing F2 RNA. **E.** Synthesis of F3 RNA**. F.** Analysis of purified *in vitro* transcribed F3 RNA.

Testing of refolded tRNA samples in *in vitro* translation

We have used *E. coli* T7 S30 extract for cell-free protein synthesis (CFPS). We used CAT (chloramphenicol acetyltrasferase) containing one in-frame UAG stop codon as a model protein to evaluate the efficiency of $tRNA_{CUD}$ during translation. The mutant CAT plasmid was created from wild-type CAT by mutating position Y109 to stop (TAG). The mRNA of CAT thus produced will have an in frame UAG stop codon at position 109 and terminate with UAA stop codon. The reaction (10µL) was performed with 400ng DNA template of wild type (wt-CAT) or mutated CAT(mut-CAT), 0.1mM amino acid mixture minus leucine, S30 premix without amino acid, T7 S30 extract (Promega L1130), 5 µM ^{14}C leucine, 0.3mg/ml coumaryl tRNA synthetase (CmRS), 0.1mM non-natural amino acid (7-HMC)[30,31], 3µM ligated tRNA (modified or unmodified). The reaction was mixed gently and incubated at 37°C for 20 mins followed by 30°C for 4 hr. Samples were placed in ice bath for 5 mins to stop the reaction. The protein samples were run on 15% SDS PAGE and analyzed by ^{14}C autoradiography using gel imaging scanner (GE *Typhoon* FLA9000). Full length protein was observed at 25kDa due to 7-HMC incorporation at UAG site. Truncated protein product was observed at 15kDa due to termination of translation at UAG site.

In vitro translation of GFP containing multiple UAG sites

Next, we used GFP containing two UAG sites as a model protein to evaluate efficiency of tRNA analog for site-specific protein modification. Three double mutant GFP plasmids were created by site directed mutagenesis at positions K140 and Y200 (GFP 1), positions Y151 and Y200 (GFP 2) and positions Y182 and Y200 (GFP 3). The mRNA of GFP will have two in frame UAG sites and terminate with UAA stop codon. The reaction (10µL) was performed with 200ng DNA template of wild type GFP (wt-GFP) or GFP double mutants, 0.1mM amino acid mixture minus leucine, S30 premix without amino acid, 5µM ^{14}C leucine, 0.3mg/ml coumaryl tRNA synthetase (CmRS), 0.1mM 7-HMC[30,31], T7 S30 extract, 3 µM ligated tRNA (modified or unmodified). The reaction was mixed gently and incubated at 37°C for 20 mins followed by 30°C for 4hr. Samples were placed in ice bath for 5 mins to stop the reaction. The protein samples were run on 15% SDS PAGE and analyzed by ^{14}C autoradiography. Full length protein was observed at 27kDa. Truncated bands were observed at 15.3kDa, 16.5kDa, 19.9kDa, and 21.9kDa due to translation termination at UAG.

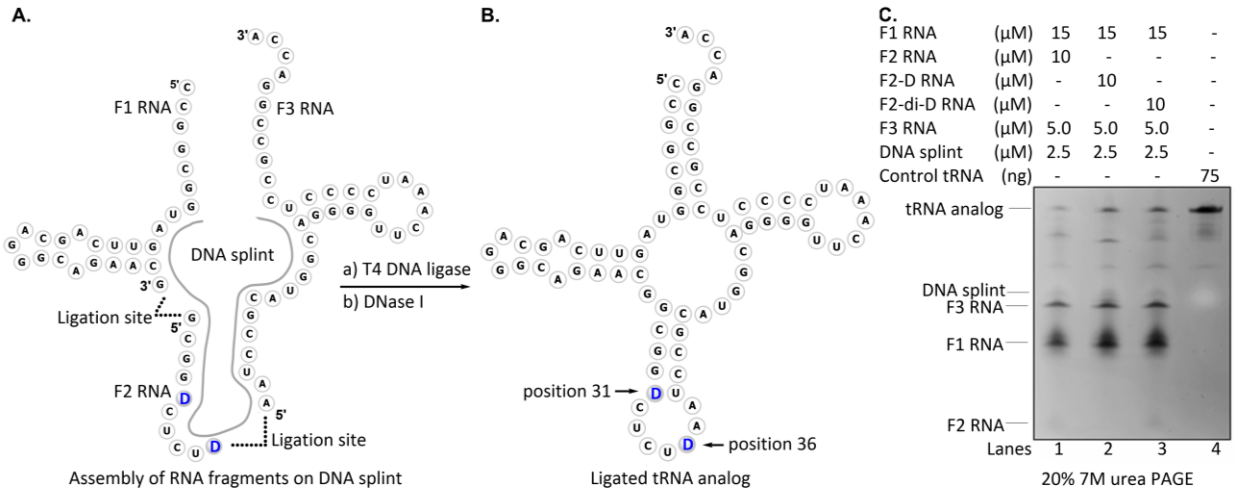

Fig. 3 A. Scheme for the synthesis of the tRNA analog. The tRNA analog ($tRNA_{CUD}$) is derived from *Mj*$tRNA^{Tyr}_{CUA}$ in which A is substituted with D at positions 31 and 36 in the sequence of F2 RNA fragment. **B**. Sequence of ligated tRNA analog. **C.** Analysis of ligated tRNA on 20% denaturing PAGE. **Lane 1**, tRNA analog synthesized from ligation of F1, F3 and F2 modified with two D. **Lane 2**, tRNA analog synthesized from ligation of F1, F3 and F2 modified with one D(F2 with one D is solid-phase synthesized), **Lane 3**, Unmodified tRNA synthesized from ligation of F1, F3 and unmodified F2. **Lane 4**, *in vitro* transcribed $tRNA^{Tyr}_{CUA}$ as positive control.

RESULTS AND DISCUSSION

Generation of tRNA analog for enhanced UAG stop codon suppression

We hypothesized that strengthening the codon-anticodon pairing using tRNA analog will facilitate UAG stop codon readthrough resulting in improved site–specific protein modification. We constructed a tRNA analog, $tRNA_{CUD}$ for stronger codon-anticodon pairing during translation. The tRNA analog was created with modification at positions 31 and 36 of ASL with 2, 6-diaminopurine (D) (Fig 3)[32]. Compared to the A-U base pair, the D-U base pair has the potential for one extra intermolecular hydrogen bond due to the presence of an additional -NH_2 group at position 2 of the purine ring in the minor groove of RNA (Fig. 1). The nucleobase D can be incorporated into RNA enzymatically using 2, 6-diaminopurine triphosphate (DTP) and T7 RNA polymerase[33]. However, *in vitro* transcription of the complete tRNA analog using T7 RNA polymerase would lead to D incorporation at all the 17 adenine bases throughout the tRNA, which is not desirable.

To overcome this problem, we designed an approach to selectively introduce D at only two adenine bases in the ASL of tRNA. The D-modified tRNA analog, $tRNA_{CUD}$ was generated by DNA splint mediated ligation of three *in vitro* transcribed RNA fragments using T4 DNA ligase (Fig. 3A, B). The three RNA fragments F1 (26nt), F2 RNA (11nt, containing two D substitutions) and F3 RNA (40nt) were synthesized using T7 RNA polymerase-based transcription (Fig. 2). To avoid problems like non-templated nucleotide addition at 5' and 3' end and requirement of GGG sequence at the transcription initiation site[42,43], we fused the DNA templates with hammerhead ribozyme sequence[36] downstream of the T7 promoter. We also inserted C2'-methoxy modification at 5' termini of the last two nucleotides in F1 and F3 RNA encoding DNA templates (Fig. 2A, E).

The transcripts generated after ribozyme self-cleavage contained free 5'and 3' hydroxyl groups. Hence, the RNA fragments were monophosphorylated at the 5'end by T4 PNK so that they could act as substrates for T4 DNA ligase (Fig. 2D, F). A DNA splint (41nt) was designed to simultaneously anneal to all the three RNA pieces, and act as substrate for T4 DNA ligase[37] (Fig. 3A). The bridging DNA splint helps in sequential alignment of RNA pieces to maintain correct tRNA sequence for T4 DNA ligase mediated ligation[38]. We used T4 DNA ligase instead of T4 RNA ligase, as the T4 RNA ligase might result in undesirable side products due to circularization of individual RNA fragments and ligated products.

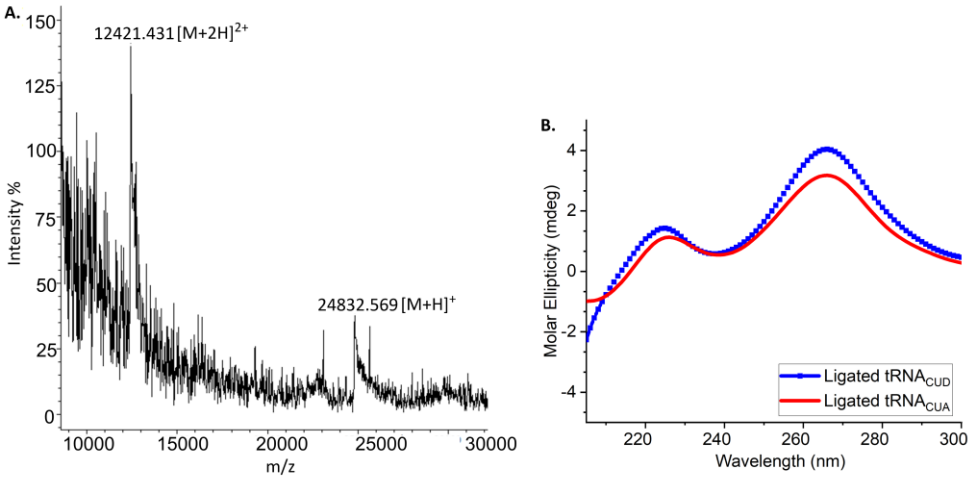

Fig. 4. Characterization of the ligated tRNA analog **A**. MALDI-TOF-MS spectra of purified tRNA analog. Calc. m/z – 24835.410 Da and obs. m/z - 24832.569 Da $[M+H]^+$ and 12421.431 Da $[M+2H]^{2+}$. **B**. CD spectra measured at 25°C for 3 µM refolded ligated $tRNA_{CUA}$ and $tRNA_{CUD}$.

The ligated tRNA species (77nt) contained two D bases in the ASL of tRNA. As a positive control, we also synthesized the corresponding unmodified $tRNA_{CUA}$. Interestingly, our results demonstrate more ligated product yield in modified RNA as compared to unmodified RNA (Fig. 3C lanes 2, 3). For comparison, we also used F2 RNA containing one D which was custom synthesized on solid-phase (Fig. 3C lane C). It is important to note that the correct sized ligated tRNA product was obtained only when optimum concentration of each RNA fragment was provided during the three-way ligation. Also, the ratio of RNA terminal used for ligation determined the homogeneity of ligation products. We optimized the ratio of F1 (26nt), F2 (11nt), F3 (40nt) and DNA splint (41nt) and found that a 6: 4: 2: 1 ratio gave maximum yield of the desired product. The full-length tRNA analog was only obtained in the presence of all three RNA pieces. The presence of D at the ligation site resulted in stable interaction at the RNA-DNA duplex junction, and no detrimental effect on T4 DNA ligase function at modified ligation junction was seen. The tRNA product (77 nt) was gel purified and characterized by MALDI-TOF (Fig. 4A) to give the expected peaks.

Secondary structural characterization of the tRNA analog

The correct folding of the tRNA analog is essential to allow tRNA flexibility for efficient translation. Hence, the secondary structure of the tRNA analog was evaluated using CD spectroscopy. We refolded the tRNA species in Tris-HCl (pH 7.4) supplemented with 20mM $MgCl_2$. The CD spectra indicated that the canonical three dimensional structure of the tRNA was not disrupted in the presence of non-natural base D (Fig. 4B) and was comparable to the CD spectra of $tRNA_{CUA}$ or other purified tRNAs[39]. For ligated $tRNA_{CUA}$ and $tRNA_{CUD}$, four major peaks were observed at 210 nm (negative), 223 nm (positive), 240 nm (negative) and ~265 nm (positive) corresponding to a double strand tRNA in A-conformation[40].

Since Mg^{2+} ion plays a crucial role in tRNA folding and function, we sought to determine the optimal concentration of $MgCl_2$ required to achieve maximal tRNA refolding *in vitro*. Towards this end, CD spectra were measured with $MgCl_2$ concentrations of 1mM, 5mM, 10mM and 20mM at 25 °C in 10mM Tris-HCl (pH 7.4). We observed optimum tRNA folding at 20mM $MgCl_2$. Interestingly, although $tRNA_{CUA}$ had considerable secondary and possibly tertiary structure, it showed no activity in translation in the absence of Mg^{2+}. It is important to note that the ligation methodology did not affect the native folding of tRNA after ligation. The refolded tRNA species were used for *in vitro* translation in *E. coli* T7 based cell-free translation to decode in frame UAG stop site in our protein of interest.

Evaluating the efficiency of tRNA analog during *E. coli* cell-free translation

1. Effect of tRNA analog on decoding one UAG site

We examined the ability of $tRNA_{CUD}$ to act as an efficient adaptor as well as competitor of RF1 during site-specific protein modification. We asked whether the stability of anticodon-codon complex enhances UAG stop site recognition by the tRNA analog during *in vitro* translation. We used an *E. coli* T7 S30 extract system as a simplified platform for coupled transcription/translation of an appropriate DNA sequence. We added $tRNA_{CUA}$ or $tRNA_{CUD}$ along with cognate aaRS and non-natural amino acid, 7-HMC to allow expression of the model protein CAT (Fig. 5). The CAT template was designed such that RNA bears an in frame UAG codon at position 101 and terminates with UAA stop codon. If the aminoacylated $tRNA_{CUD}$ was successful in decoding UAG, full length 25 kDa CAT protein was expected, otherwise a truncated 15kDa protein would be formed. Full-length wild-type CAT was used as a positive control (Fig. 5, lane 2). After *in vitro* translation, we observed 50% more full-length modified CAT protein with respect to $tRNA_{CUD}$ when compared to $tRNA_{CUA}$. The yield of full-length modified CAT in the presence of $tRNA_{CUA}$ and $tRNA_{CUD}$ are 24% and 56% respectively with respect to full length CAT protein which was used as positive control (Fig. 5, lanes 4, 5). Further, the amount of truncated protein was significantly less when $tRNA_{CUD}$ was present, indicating that the tRNA analog was able to effectively outcompete RF-1 during translation. We conclude that $tRNA_{CUD}$ more effectively decodes the UAG stop site during ribosomal translation. Our results represent the importance of tRNA anticodon modification and codon-anticodon interaction in protein translation.

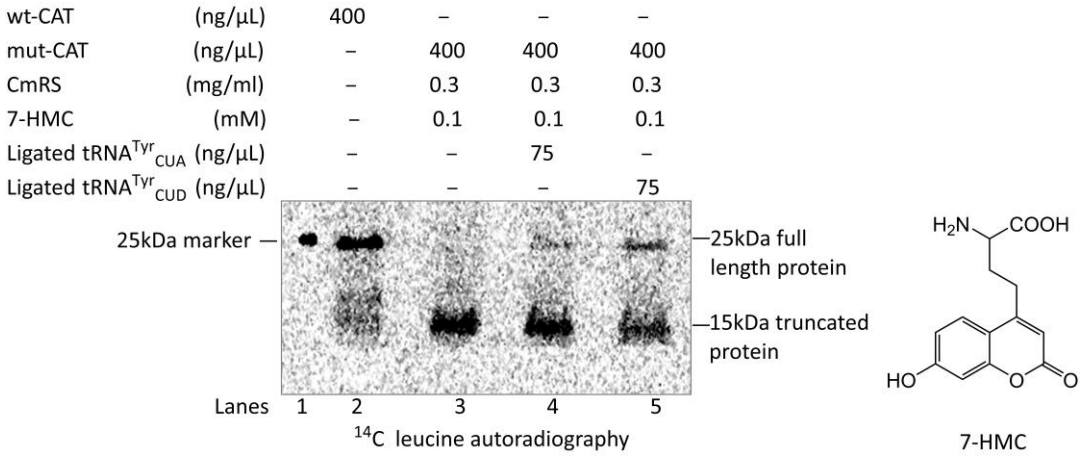

Fig. 5. Analysis of *in vitro* translated products by ^{14}C autoradiography. **Lane 2** is positive control showing expression of 25kDa wt-CAT protein. **Lane 3** shows 15 kDa truncated CAT protein without any suppressor tRNA. **Lane 4** shows expression of full-length modified CAT protein in presence of 3µM $tRNA_{CUA}$ and 7-HMC. **Lane 5** shows expression of full-length modified CAT protein in presence of 3 µM $tRNA_{CUD}$ and 7-HMC.

2. Effect of tRNA analog on decoding two UAG sites

As the tRNA analog could more efficiently decode the UAG stop codon, we asked whether $tRNA_{CUD}$ will be useful for amber suppression at multiple positions in a protein. To test this hypothesis, we constructed three GFP double mutants carrying two UAG stop codons at different positions in the protein sequence. The mRNAs bears in frame UAG stop sites at residues 140 and 200 in GFP 1 (Fig. 6a), residues 151 and 200 in GFP 2, residues 182 and 200 in GFP 3, and terminates with UAA stop codon. If the aminoacylated $tRNA_{CUD}$ successfully decodes both UAG sites, full length GFP (27kDa) containing two 7-HMC units will be formed. Otherwise, only truncated GFP variants are expected.

$tRNA_{CUA}$ was unable to decode two UAG sites simultaneously resulting in generation of truncated products only (Fig. 6, lanes 6, 7, 8) whereas $tRNA_{CUD}$ was able to successfully decode both UAG sites resulting in full length GFP (Fig. 6, lanes 3, 4, 5). Truncated proteins were observed at 15.3kDa, 16.5kDa and 19.9 kDa (Fig. 6, lanes 9, 10, 11) due to translation termination at the first UAG stop site in GFP 1, GFP 2 and GFP 3 respectively. Another truncated band was observed at 21.9kDa (Fig. 6, lanes 9, 10, 11) due to the inability of tRNA to encode 7-HMC at the second UAG site (position 200). Full-length wild-type GFP was used as a positive control (Fig. 6. lane 2). For estimating suppression efficiency, we compared the ratio of full-length protein resulting from the two tRNA species after UAG codon readthrough. The suppression efficiency of full length GFP 1, GFP 2 and GFP 3 in the presence of $tRNA_{CUD}$ was 32.8%, 12.5%, 14.6% respectively compared to wt-GFP (Fig. 6. lane 2) whereas the suppression efficiency was 0% for $tRNA_{CUA}$. The reason for increased protein yield is likely the effective out-competition of RF1 by $tRNA_{CUD}$ during the UAG readthrough step in translation.

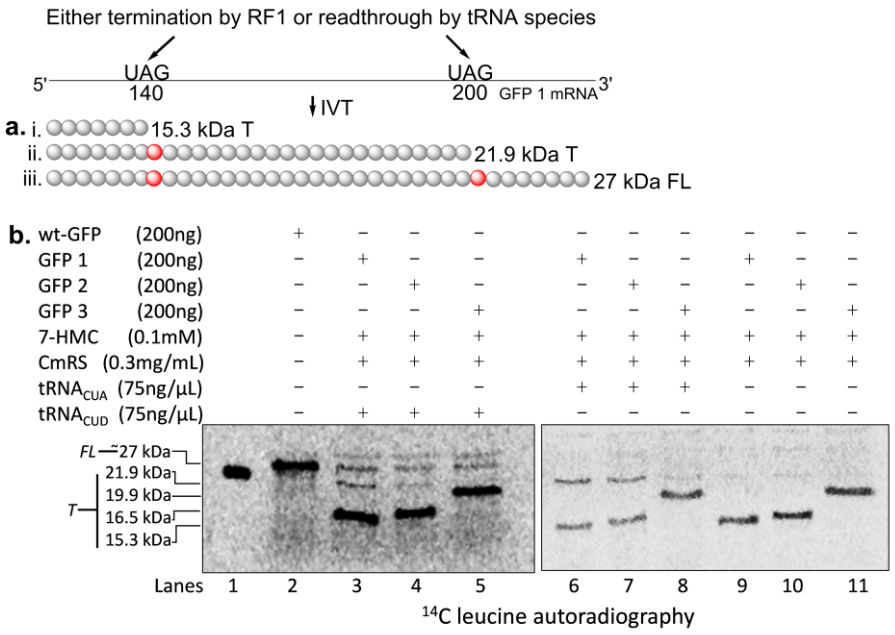

Fig. 6. Scheme of UAG stop site decoding by tRNA species **a.**140 and 200 in GFP 1. **b.** Analysis of *in vitro* translated proteins by ^{14}C Leucine autoradiography. **Lane 2** is positive control showing expression of FL 27kDa wt-GFP. **Lanes 3, 4, 5** shows expression of FL 27kDa modified GFP 1, GFP 2 and GFP 3 in presence of 3 µM $tRNA_{CUD}$ and 7-HMC. **Lanes 6, 7, 8** shows expression of only truncated GFP products in presence of 3 µM $tRNA_{CUA}$ and 7-HMC. **Lanes 9, 10, 11** is negative control, shows expression of truncated protein in the absence of any suppressor tRNA. Lane **1**: 25kDa marker. T- truncated. FL – full length. IVT – *in vitro* translation.

Our study highlights the importance of anticodon base modification in the context of hydrogen bonding. It is well-known that the yield of modified protein depends on the UAG site selection, called mRNA context effect[41]. One report suggested that RF1 interactions during translation depend on mRNA sequence context, and contributed approximately half of the observed difference in reassignment efficiency[42]. As $tRNA_{CUD}$ was able to outcompete RF-1 to a significant extent, we speculate that using $tRNA_{CUD}$ may help alleviate mRNA context effect. The use of $tRNA_{CUD}$ did not seem to adversely affect the functioning of the various components of the translational machinery during translation.

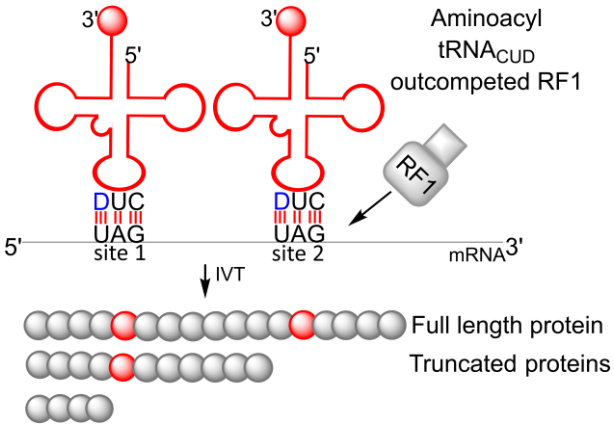

Fig. 7. Model summarizing *in vitro* translation of protein using $tRNA_{CUD}$. IVT – *in vitro* translation.

CONCLUSION

The main goal of this study was to construct a modified suppressor tRNA that will form stronger codon-anticodon complex during translation and result in higher yield of NAA containing protein products. Our results demonstrate the importance of tRNA modification on UAG suppression for site-specific protein modification. We synthesized an anticodon modified tRNA, which was structurally and conformationally similar to $tRNA_{CUA}$. We have shown that the codon-anticodon duplex stability may be increased by using D, that provide three H-bonds in the D-U pair instead of two in the A-U pair at the interface of mRNA codon (5'-UAG-3') and tRNA anticodon (5'-CUA-3'). $tRNA_{CUD}$ was used in a cell-free *E. coli* translation system, where it resulted in significant increase in the yield of protein product containing NAA at two sites. Our $tRNA_{CUD}$ may be useful to study suppression efficiency *in vivo* by injecting into *Xenopus* oocytes[43].

The applications of NAA-incorporated proteins is limited by the selectivity and overall efficiency of the stop codon reassignment methodology. Our results add a new dimension to be considered for improved stop codon reassignment. The tRNA modification studies may address issues related to translation recoding that stem from poor understanding of codon-anticodon interaction and repurposing translation recoding for genetic code expansion. In future it will be interesting to investigate the dependence of mRNA base or sequence context on codon-anticodon stability. We approached the relation between tRNA modification and mRNA codon reading efficiency to indicate importance of anticodon loop modification during translation. Engineering of translational components will rewire translation to foster genetic code expansion so that artificial biopolymers can be synthesized, tested and evolved.

ACKNOWLEDGMENTS

We thank IISER Bhopal and SERB for funding of this work. We also thank Prof. Peter Schultz for the *Mj*tRNA$^{Tyr}_{CUA}$/aaRS plasmid and Prof. Vikas Jain for the GFP plasmid. PM thanks CSIR for a doctoral fellowship.

REFERENCES

1. Rodnina, M. V. Translation in prokaryotes. *Cold Spring Harb. Perspect. Biol.***10**, 1–22 (2018).
2. Lim, V. I. & Venclovas, Č. Codon-anticodon pairing A model for interacting codon-anticodon duplexes located at the ribosomal A- and P-sites. *FEBS Lett.***313**, 133–137 (1992).
3. Agris, P. F. *et al.* Celebrating wobble decoding: Half a century and still much is new. *RNA Biol.***15**, 537–553 (2018).
4. Satoh, A., Takai, K., Ouchi, R., Yokoyamaand, S. & Takaku, H. Effects of anticodon 2'-O-methylations on tRNA codon recognition in an Escherichia coli cell-free translation. *RNA***6**, 680–686 (2000).
5. Yarian, C. *et al.* Accurate translation of the genetic code depends on tRNA modified nucleosides. *J. Biol. Chem.***277**, 16391–16395 (2002).
6. Lim, V. I. & Curran, J. F. Analysis of codon:anticodon interactions within the ribosome provides new insights into codon reading and the genetic code structure. *RNA***7**, 942–957 (2001).
7. Crick, F. H. C. Codon—anticodon pairing: The wobble hypothesis. *J. Mol. Biol.***19**, 548–555 (1966).
8. Crick, F. & Watson, J. Molecular structure of nucleic acids. *Nature***171**, 737–738 (1953).
9. Hoernes, T. P. *et al.* Translation of non-standard codon nucleotides reveals minimal requirements for codon-anticodon interactions. *Nat. Commun.***9**, 7321–7328 (2018).
10. Jason, W. Review Expanding and reprogramming the genetic code. *Nature***550**, 53–60 (2017).
11. Smolskaya, S. Site-Specific Incorporation of Unnatural Amino Acids into Escherichia coli Recombinant Protein : Methodology Development and Recent Achievement. *Biomolecules***9**, 255–272 (2019).
12. Wang, L., Brock, a, Herberich, B. & Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science***292**, 498–500 (2001).
13. Wang, L., Xie, J. & Schultz, P. G. Expanding the Genetic Code. *Annu. Rev. Biophys. Biomol. Struct.***35**, 225–249 (2006).
14. Wang, K., Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat. Biotechnol.***25**, 770–777 (2007).
15. Rauch, B. J., Porter, J. J., Mehl, R. A. & Perona, J. J. Improved Incorporation of Noncanonical Amino Acids by an Engineered tRNA. *Biochemistry***55**, 618–628 (2016).
16. Johnson, D. B. F. *et al.* RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat. Chem. Biol.***7**, 779–86 (2011).
17. Chemla, Y., Ozer, E., Algov, I. & Alfonta, L. Context effects of genetic code expansion by stop codon suppression. *Curr. Opin. Chem. Biol.***46**, 146–155 (2018).
18. Albayrak, C. & Swartz, J. R. Cell-free co-production of an orthogonal transfer RNA activates efficient site-specific non-natural amino acid incorporation. *Nucleic Acids Res.***41**, 5949–5963 (2013).
19. Auffinger, P. & Westhof, E. An extended structural signature for the tRNA anticodon loop. *RNA***7**, 334–341 (2001).
20. Grosjean, H., Cedergren, R. J. & McKay, W. Structure in tRNA data. *Biochimie***64**, 387–397 (1982).
21. Hoernes, T. P. *et al.* Atomic mutagenesis of stop codon nucleotides reveals the chemical prerequisites for release factor-mediated peptide release. *Proc. Natl. Acad. Sci. U. S. A.***115**, 382–389 (2018).
22. Bailly, C. & Waring, M. J. The use of diaminopurine to investigate structural properties of nucleic acids and molecular recognition between ligands and DNA. *Nucleic Acids Res.***26**, 4309–4314 (1998).
23. Noren, C. J., Anthony-cahill, S. J., Griffith, M. C. & Schultz, P. G. A General Method for Site-Specific Incorporation of Unnatural Amino Acids into Proteins. *Science***244**, 182–188 (1989).
24. Dondapati, S. K., Stech, M., Zemella, A. & Kubick, S. Cell-Free Protein Synthesis: A Promising Option for Future Drug Development. *BioDrugs***34**, 327–348 (2020).
25. Gao, W., Cho, E., Liu, Y. & Lu, Y. Advances and challenges in cell-free incorporation of unnatural amino acids into proteins. *Front. Pharmacol.***10**, 611–618 (2019).
26. Gao, W., Bu, N. & Lu, Y. Efficient Incorporation of Unnatural Amino Acids into Proteins with a Robust Cell-Free System. *Methods Protoc.***2**, 16–28 (2019).
27. Kelwick, R. J. R., Webb, A. J., Freemont, P. S. & Myers, C. J. Biological Materials : The Next Frontier for Cell-Free Synthetic Biology. *Front. Bioeng. Biotechnol.***8**, 399–414 (2020).
28. Venkat, S., Chen, H., Gan, Q. & Fan, C. The application of cell-free protein synthesis in genetic code expansion for post-translational modifications. *Front. Pharmacol.***10**, 248–254 (2019).
29. Young, T. S., Ahmad, I., Yin, J. A. & Schultz, P. G. An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli. *J. Mol. Biol.***395**, 361–374 (2010).
30. Saraogi, I., Zhang, D., Chandrasekaran, S. & Shan, S. O. Site-specific fluorescent labeling of nascent proteins on the translating ribosome. *J. Am. Chem. Soc.***133**, 14936–14939 (2011).
31. Wang, J., Xie, J. & Schultz, P. G. A genetically encoded fluorescent amino acid. *J. Am. Chem. Soc.***128**, 8738–8739 (2006).

32. Wang, L. & Schultz, P. G. A general approach for the generation of orthogonal tRNAs. *Chem. Biol.***8**, 883–890 (2001).
33. Jr, H. B. G., Arar, K., Gewirtz, A. & Hou, Y. Unrestricted accessibility of short oligonucleotides to RNA. *RNA***11**, 1441–1447 (2005).
34. Tor, Y. & Dervan, P. B. Site-Specific Enzymatic Incorporation of an Unnatural Base, N6-(6-Aminohexyl)isoguanosine, into RNA. *J. Am. Chem. Soc.***115**, 4461–4467 (1993).
35. Doudna, J. A. Hammerhead ribozyme structure: U-turn for RNA structural biology. *Structure***3**, 747–750 (1995).
36. Kharma, N. *et al.* Automated design of hammerhead ribozymes and validation by targeting the PABPN1 gene transcript. *Nucleic Acids Res.***44**, 39–51 (2015).
37. Li, N. S. *et al.* A general and efficient approach for the construction of RNA oligonucleotides containing a 5′-phosphorothiolate linkage. *Nucleic Acids Res.***39**, 31–47 (2011).
38. Booy, E. P., Meng, H. & Mckenna, S. A. Splint ligation of RNA with T4 DNA ligase. *Metod. Mol. Biol.***941**, 257–269 (2012).
39. Blum, A. D., Uhlenbeck, O. C. & Tinoco, I. Circular Dichroism Study of Nine Species of Transfer Ribonucleic Acid. *Biochemistry***11**, 3248–3256 (1972).
40. Agudelo, D., Bourassa, P., Beauregard, M., Bérubé, G. & Tajmir-Riahi, H. A. tRNA Binding to Antitumor Drug Doxorubicin and Its Analogue. *PLoS One***8**, 69248–69256 (2013).
41. Bartoschek, M. D. *et al.* Identification of permissive amber suppression sites for efficient non-canonical amino acid incorporation in mammalian cells. *Nucleic Acids Res.***49**, E62 (2021).
42. Schwark, D. G., Schmitt, M. A. & Fisk, J. D. Dissecting the contribution of release factor interactions to amber stop codon reassignment efficiencies of the methanocaldococcus jannaschii orthogonal pair. *Genes (Basel).***9**, 546–562 (2018).
43. Beene, D L., Dougherty, D. A. & Lester, H. A. Unnatural amino acid mutagenesis in mapping ion channel function. *Curr. Opin. Neurobiol.***13**, 264–270 (2003).