

ENGINEERED INITIATOR tRNAs AND USES THEREOF**CROSS-REFERENCE TO RELATED APPLICATIONS**

This application claims priority to U.S. Provisional Patent Application No. 63/550,806, filed February 7, 2024, which is incorporated herein by reference in its entirety.

BACKGROUND OF THE INVENTION

Transfer RNA ("tRNA") is an RNA typically 76 to 90 nucleotides in length in eukaryotes). It is the physical link between the mRNA coding sequence and the amino acid sequence of the protein encoded by the mRNA. tRNA accomplishes this by physically carrying a specific amino acid charged to the tRNA to the protein synthesizing machinery – ribosome – to facilitate protein translation. There, complementation of a 3-nucleotide codon in the mRNA by a 3-nucleotide anticodon in the tRNA results in protein synthesis based on the mRNA codons. As such, tRNAs are necessary components of translation, the biological synthesis of new proteins in accordance with the genetic code.

Typically, tRNA genes from prokaryotic bacteria are smaller (mean = 77.6 bp) than tRNAs from Archaea (mean = 83.1 bp) and eukaryotes (mean = 84.7 bp). In contrast, mature tRNAs from bacteria are usually longer (median = 77.6 nt) than tRNAs from Archaea (median = 76.8 nt), with eukaryotes (median = 74.5 nt).

The mRNA encodes a protein as a series of contiguous codons, each of which is recognized by a particular tRNA, by way of complementary base pairing between each of the consecutive 3-nt codons on mRNA, with 3-nt anticodons on the tRNA matches.

On the 3' end of the tRNA molecule is a covalent attachment to the amino acid that corresponds to the specific anticodon sequence. Each type of tRNA molecule can be attached to only one type of amino acid, thus each organism has many different types of tRNA. Because the genetic code contains multiple codons that specify the same amino acid, there are several tRNA molecules bearing different anticodons which carry the same amino acid.

The covalent attachment to the tRNA 3' end is catalyzed aminoacyl tRNA synthetases (AATS). During protein synthesis, tRNAs with attached amino acids are delivered to the ribosome by elongation factors (EFs), which aid in association of the tRNA with the ribosome, synthesis of the new polypeptide, and translocation of the ribosome along the mRNA.

If the tRNA's anticodon matches the mRNA, another tRNA already bound to the

ribosome transfers the growing polypeptide chain from its 3' end to the amino acid attached to the 3' end of the newly delivered tRNA, a reaction catalyzed by the ribosome.

The structure of tRNA can be decomposed into its primary structure (*i.e.*, nucleotide sequence of the tRNA), its secondary structure (*e.g.*, tRNA having stems and loops due to base pairing, usually visually depicted as a cloverleaf structure), and its tertiary structure (*e.g.*, the 3-D tRNA structure that usually resembles an "L"). All tRNAs have the similar L-shaped 3D structures that allows them to fit into the P and A sites of the ribosome. The cloverleaf structure becomes the 3D L-shaped structure through coaxial stacking of the helices, which is a common RNA tertiary structure motif. The lengths of each arm, as well as the loop diameter, in a tRNA molecule vary from species to species.

A typical tRNA structure consists of the following structural elements:

Acceptor Arm: the acceptor stem is a 7- to 9-base pair (bp) stem made by the base pairing of the 5'-terminal nucleotide with the 3'-terminal nucleotide (which contains the CCA 3'-terminal group used to attach the amino acid). In general, such 3'-terminal tRNA-like structures are referred to as "genomic tags." The acceptor stem may contain non-Watson-Crick base pairs.

CCA tail: the CCA tail is a cytosine-cytosine-adenine sequence at the 3' end of the tRNA molecule. The amino acid loaded onto the tRNA by aminoacyl tRNA synthetases, to form aminoacyl-tRNA, is covalently bonded to the 3'-hydroxyl group on the CCA tail. This sequence is important for the recognition of tRNA by enzymes and critical in translation. In prokaryotes, the CCA sequence is transcribed in some tRNA sequences. In most prokaryotic tRNAs and eukaryotic tRNAs, the CCA sequence is added during processing and therefore does not appear in the tRNA gene.

D Loop: The D loop (also referred to herein as "**D arm**") is a 4-6-bp stem ending in a loop that often contains dihydrouridine.

Anticodon Loop: the anticodon loop (also referred to herein as "anticodon arm" or "**A arm**") is a 5-bp stem whose loop contains the anticodon. The tRNA 5'-to-3' primary structure contains the anticodon but in reverse order, since 3'-to-5' directionality is required to read the mRNA from 5'-to-3'.

TΨC loop: The TΨC loop (also referred to herein as "**T arm**") is named so because of the characteristic presence of the unusual base Ψ in the loop, where Ψ is pseudouridine, a modified uridine. The modified base is often found within the sequence 5'-TΨCGA-3', with the T (ribothymidine, m⁵U) and A forming a base pair.

Variable Loop: the variable loop sits between the anticodon loop and the ΨU loop

and, as its name implies, varies in size from 3 to 21 bases.

An anticodon is a unit of three nucleotides corresponding to the three bases of an mRNA codon. Each tRNA has a distinct anticodon triplet sequence that can form 3 complementary base pairs to one or more codons for an amino acid. Some anticodons pair with more than one codon due to wobble base pairing. Frequently, the first nucleotide of the anticodon is one not found on mRNA: inosine, which can hydrogen bond to more than one base in the corresponding codon position.

The top half of tRNA consisting of the T arm and the acceptor stem with 5'-terminal phosphate group and 3'-terminal CCA group, and the bottom half consisting of the D arm and the anticodon arm, are independent units in structure as well as in function.

It is important to note that many species have lost specific tRNAs during evolution. For instance, both mammals and birds lack the same 14 out of the possible 64 tRNA genes, but other life forms contain these tRNAs. For translating codons for which an exact pairing tRNA is missing, organisms resort to a strategy called wobbling, in which imperfectly matched tRNA/mRNA pairs still give rise to translation, although this strategy also increases to propensity for translation errors. Because nucleotide triplets can present more combinations than there are amino acids and associated tRNAs, there is redundancy in the genetic code, and several different 3-nucleotide codons can express the same amino acid. This codon bias is what necessitates codon optimization.

In genetic code, it is common for a single amino acid to be specified by all four third-position possibilities, or at least by both pyrimidines and purines; for example, the amino acid glycine is coded for by the codon sequences GGU, GGC, GGA, and GGG. Other modified nucleotides may also appear at the first anticodon position - sometimes known as the "wobble position" - resulting in subtle changes to the genetic code, as for example in mitochondria. Per cell, 61 tRNA types are required to provide one-to-one correspondence between tRNA molecules and codons that specify amino acids, as there are 61 sense codons of the standard genetic code. However, many cells have under 61 types of tRNAs because the wobble base is capable of binding to several, though not necessarily all, of the codons that specify a particular amino acid. At least 31 tRNAs are required to translate, unambiguously, all 61 sense codons.

Aminoacylation (the process of adding an aminoacyl group to a compound) covalently links a specific amino acid to the CCA 3' end of a cognate tRNA molecule. Each tRNA is aminoacylated (or "charged") with a specific amino acid by an aminoacyl tRNA synthetase. There is normally a single aminoacyl tRNA synthetase for each amino acid,

despite the fact that there can be more than one tRNA, and more than one anticodon for an amino acid. Recognition of the appropriate tRNA by the synthetases is not mediated solely by the anticodon, and the acceptor stem often plays a prominent role.

Protein synthesis is generally initiated with methionine or formylmethionine in all organisms, by using AUG as the initiation / start codon, and a special methionine tRNA - the initiator tRNA (tRNAⁱⁿⁱ). Eubacteria, mitochondria, and chloroplasts initiate protein synthesis with formylmethionine. Meanwhile, archaeobacteria and eukaryotic cytoplasm initiate with methionine. In eubacteria, codons such as GUG and UUG are also occasionally used for initiation. However, these codons are read by the same initiator tRNA and are translated as formylmethionine. In eukaryotic systems, AUG is almost exclusively the codon used for initiation. In rare instances where ACG, CUG, AUU, and AGG are used as start codon, protein synthesis is still thought to be initiated with methionine.

However, it has been found in *E. coli*, and more recently in mammalian cells, that protein synthesis could be initiated with codons other than AUG and amino acids other than methionine, using anticodon sequence mutants of the initiator tRNA. Such anticodon mutation allows the mutant initiator tRNA to initiate protein synthesis by using a non-AUG codon complementary to the new / mutated anticodon. Because in many cases, the anticodon is an important determinant for recognition of the tRNA by aminoacyl-tRNA synthetases, the mutant tRNAs are often aminoacylated with different amino acids.

In any event, translation of an mRNA requires an initiator tRNA for initiation of protein synthesis. The provision of an engineered initiator tRNA provides a means to produce proteins or peptides, at least *in vitro*, with the first amino acid of the protein / peptide being a non-Met residue. This can be tremendously valuable for, *e.g.*, generating a peptide library with enhanced diversity and novel functional characteristics, without the limitation of must having Met as the first residue.

SUMMARY OF THE INVENTION

One aspect of the invention provides an engineered initiator tRNA (tRNAⁱⁿⁱ), comprising: (1) a mutated anti-codon comprising one or more base changes at the 1st (*i.e.*, the C position), 2nd (*i.e.*, the A position), and/or 3rd (*i.e.*, the U position) positions of the wild-type CAU anti-codon of the wild-type tRNAⁱⁿⁱ, wherein the mutated anti-codon does not comprise base change to A or U at the 1st position; and, (2) optionally, a further mutation that promotes folding of the A-stem of the tRNAⁱⁿⁱ.

In certain embodiments, the further mutation comprises: (i) a first double mutation in

the stem region of the D-arm of the wild-type tRNAⁱⁿⁱ, wherein the first double mutation preserves an original G:C base pair in the stem region of the D-arm; optionally, the first double mutation is associated with a C-to-G change in the 1st position of the anti-codon; and/or, (ii) a second double mutation comprising a C-to-G mutation in the loop region of the T-arm of the wild-type tRNAⁱⁿⁱ, and a G-to-C mutation in the loop region of the D-arm of the wild-type tRNAⁱⁿⁱ; optionally, the second double mutation is associated with a U-to-G change in the 3rd position of the anti-codon.

In certain embodiments, the engineered tRNAⁱⁿⁱ does not comprise the further mutation, and which initiates translation from / incorporates the first amino acid in a nascent polypeptide chain encoded by a polynucleotide comprising a cognate start codon for the mutated anti-codon, at a relative efficiency of at least about 40%, 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140% or more, compared to the efficiency of the wild-type tRNAⁱⁿⁱ using the wild-type AUG start codon.

In certain embodiments, the mutated anti-codon is CUA, CCA, CGU, GGA, CAA, CGA, CAG, or CAC.

In certain embodiments, the engineered tRNAⁱⁿⁱ comprises the further mutation, and which has enhanced ability to initiate translation from / incorporate the first amino acid in a nascent polypeptide chain encoded by a polynucleotide comprising a cognate start codon for the mutated anti-codon, compared to a control engineered tRNAⁱⁿⁱ having the mutated anti-codon but lacking the further mutation.

In certain embodiments, the first double mutation is associated with a C-to-G change in the 1st position of the anti-codon, and comprises a C13G mutation and a G23C mutation.

In certain embodiments, the second double mutation is associated with a U-to-G change in the 3rd position of the anti-codon, and comprises a C57G mutation and a G20C mutation.

In certain embodiments, in the engineered tRNAⁱⁿⁱ: (1) the mutated anti-codon is GAA, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 2); (2) the mutated anti-codon is GGA, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 3); (3) the mutated anti-codon is GUA, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 4); (4) the mutated anti-codon is GCA, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 5); (5) the mutated anti-codon is CAA, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 6); (6) the mutated anti-codon is CGA, and optionally there is no said further mutation

(*e.g.*, SEQ ID NO: 7); (7) the mutated anti-codon is CUA, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 8); (8) the mutated anti-codon is CCA, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 9); (9) the mutated anti-codon is GAG, and wherein the further mutation comprises a C13G and a G23C double mutation, and a C57G and a G20C double mutation (*e.g.*, SEQ ID NO: 10); (10-12) the mutated anti-codon is GGG, GUG, or GCG, and wherein the further mutation comprises a C13G and a G23C double mutation, and a C57G and a G20C double mutation (*e.g.*, SEQ ID NOs: 11-13, respectively); (13) the mutated anti-codon is CAG, and wherein the further mutation comprises a C57G and a G20C double mutation (*e.g.*, SEQ ID NO: 14); (14) the mutated anti-codon is CGG, and wherein the further mutation comprises a C57G and a G20C double mutation (*e.g.*, SEQ ID NO: 15); (15-16) the mutated anti-codon is CUG, or CCG, and wherein the further mutation comprises a C57G and a G20C double mutation (*e.g.*, SEQ ID NOs: 16 & 17, respectively); (17) the mutated anti-codon is GAU, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 18); (18-20) the mutated anti-codon is GGU, GUU, or GCU, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NOs: 19-21, respectively); (21) the mutated anti-codon is CGU, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 22); (22-23) the mutated anti-codon is CUU, or CCU, and optionally there is no said further mutation (*e.g.*, SEQ ID NOs: 23 & 24, respectively); (24-27) the mutated anti-codon is GAC, GGC, GUC, or GCC, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NOs: 25-28, respectively); (28) the mutated anti-codon is CAC, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 29); or, (29-31) the mutated anti-codon is CGC, CUC, or CCC, and optionally there is no said further mutation (*e.g.*, SEQ ID NOs: 30-32, respectively).

In certain embodiments, the wild-type tRNAⁱⁿⁱ is a prokaryotic tRNAⁱⁿⁱ.

In certain embodiments, the wild-type tRNAⁱⁿⁱ is *E. coli*. fMet-tRNAⁱⁿⁱ (SEQ ID NO: 1).

Another aspect of the invention provides a translation system for the synthesis of a polypeptide, comprising the engineered initiator tRNA (tRNAⁱⁿⁱ) described herein.

Another aspect of the invention provides a method of constructing a polypeptide library, the method comprising translating a coding sequence encoding a member polypeptide of the polypeptide library, in a cell-free translation system comprising the engineered initiator tRNA (tRNAⁱⁿⁱ) described herein.

Another aspect of the invention provides a method of constructing a library of a

complex between a polypeptide and an mRNA encoding the polypeptide, comprising translating in a cell-free translation system comprising the engineered initiator tRNA (tRNAⁱⁿⁱ) described herein.

Another aspect of the invention provides a polypeptide library constructed by the method of the invention described herein, or a polypeptide-mRNA complex library constructed by the method of the invention.

It should be understood that any one embodiment of the invention described herein, including those embodiments described only under one aspect of the invention or one section below, or only in the examples or claims, can be combined with any one or more other embodiments of the invention unless such combination is expressly disclaimed or is improper.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGs. 1A-1C show initiator tRNA anticodon mutations and translation efficiency. FIG. 1A shows sequence of tRNAⁱⁿⁱ_{CAU} (SEQ ID NO: 1) highlighting initiator tRNA identity elements in red. Mutations to the 31 orthogonal anticodons are illustrated in the adjacent table. FIG. 1B shows HiBiT translation assays of 32 tRNAⁱⁿⁱ with orthogonal anticodons initiating translation from their respective cognate start codons and translating the coding sequence for peptide ^ΔFGGHGGHHGGGGSSKKSGWRLF (SEQ ID NO: 33). Color codes are matched to anticodons in FIG. 1A and the native tRNAⁱⁿⁱ_{CAU} is plotted in all graphs as reference in green. FIG. 1C shows relative translation efficiencies of all mutated tRNAⁱⁿⁱ normalized to tRNAⁱⁿⁱ_{CAU}.

FIGs. 2A & 2B show tRNA fold analysis. FIG. 2A shows most frequent energy structure and their respective frequency in the structural ensemble as calculated by the RNAfold webserver. Bases are color coded according to their probability to be folded as predicted with dark red representing a probability of 1 to dark blue representing a probability of 0. FIG. 2B shows the results of 20% native TBE gels run in 50 mM HEPES pH 7.2 and 3 mM MgCl₂.

FIGs. 3A-3C show structural engineering of misfolded tRNAs. FIG. 3A shows common misfolding patterns observed for tRNAs with anticodon mutations to a G in the 1st and 3rd position. Examples shown are tRNAⁱⁿⁱ_{GCA} and tRNAⁱⁿⁱ_{GAG}. Mis-paired base-pairs and anticodons are highlighted in red. Native fold base-pairing in misfolded structures is indicated with a dashed grey line. The sequences of the shown initiator tRNAs are SEQ ID NOs: 36, 5, 37, and 10, respectively. FIG. 3B shows HiBiT translation assays of mutant

tRNAs relative to tRNAⁱⁿⁱ_{CAU}. FIG. 3C is a summary of mutant initiator tRNAs synthesized and their respective translation efficiency compared to the wildtype initiator tRNA. All translation efficiencies are normalized to tRNAⁱⁿⁱ_{CAU}.

FIGs. 4A-4C show the results of orthogonal initiator codon use assessment. FIG. 4A shows exemplary MALDI experiment design for orthogonal use of codons UUC, UUG, CUC, CUG and AUC. Each initiator tRNA (tRNAⁱⁿⁱ_{GAA} is displayed as example) is charged with acetylated phenylalanine (^{Ac}F) and translated individually using a library that encodes peptides (SEQ ID NOs: 38-42, respectively, from top to bottom) with a different molecular weight depending on the start codon. Each peptide contains a FLAG tag highlighted in yellow. The displayed anticodon box highlights all tRNAs tested with a white background and those used in the shown experiment in color. tRNAs denoted with * correspond to double mutations and ** to quadruple mutations listed in FIG. 3C. FIG. 4B shows an overlay of five MALDI spectra recorded for tRNAⁱⁿⁱ_{GAA} (red), tRNAⁱⁿⁱ_{CAA} (blue), tRNAⁱⁿⁱ_{GAG} (yellow), tRNAⁱⁿⁱ_{CAG} (cyan) and tRNAⁱⁿⁱ_{GAU} (purple) translated from the library displayed in FIG. 3A. FIG. 4C shows percentage of non-cognate start codon translation initiation relative to the respective cognate codon initiation for each tRNA.

FIGs. 5A-5B show mutant folding predictions and gel shift. FIG. 5A shows double mutant design for CAG and CGG. Structures shown represent the most frequent energy (mfe) structure within the ensemble. Percentages represent the predicted percentage of the RNA to fold as illustrated within the structural ensemble. Images and color coding are extracted from the RNAfold webserver. Mutated nucleotides are indicated by red arrows. FIG. 5B shows the result of 20% native TBE gels run in 50 mM HEPES pH 7.2 and 3 mM MgCl₂ of all mutated tRNAs compared to their wildtype variants.

FIGs. 6A-6D show MALDI spectra overlay to assess non-cognate start codon translation initiation. Initiator tRNAs were acylated individually with acetylated phenylalanine as demonstrated in FIGs. 4A-4C, and translated using the illustrated pool of libraries. MALDI spectra are showing translation products initiating with (FIG. 6A) ^{Ac}F_{GUA} (blue) and ^{Ac}F_{CUA} (orange) (peptide sequences SEQ ID NOs: 40-41, respectively, from top to bottom); (FIG. 6B) ^{Ac}F_{GCA} (blue) and ^{Ac}F_{CCA} (orange) (peptide sequences SEQ ID NOs: 40-41, respectively, from top to bottom); (FIG. 6C) ^{Ac}F_{GGA} (blue), ^{Ac}F_{CGA} (orange), and ^{Ac}F_{CGU} (green) (peptide sequences SEQ ID NOs: 40-42, respectively, from top to bottom); and (FIG. 6D) ^{Ac}F_{GAG} (cyan), ^{Ac}F_{CAG} (light blue), ^{Ac}F_{GAU} (orange) and ^{Ac}F_{CAU} (green) (peptide sequences SEQ ID NOs: 40-43, respectively, from top to bottom).

DETAILED DESCRIPTION OF THE INVENTION

The invention described herein is partly based on the discovery that a wild-type initiator tRNA, when mutated to have certain anticodon sequence changes (such as changing the first anticodon nucleotide from C to G, or the last / 3rd anticodon nucleotide from U to G), tends to have drastically reduced ability to initiate protein synthesis from a matching non-AUG “start” codon. The invention is also based on the further discovery that, when certain additional mutations are introduced in such mutant initiator tRNA’s, their ability to initiate protein synthesis from the same matching non-AUG codon is substantially restored.

Thus, one aspect of the invention provides an engineered initiator tRNA (tRNAⁱⁿⁱ), comprising: (1) a mutated anti-codon comprising one or more base changes at the 1st (*i.e.*, the C position), 2nd (*i.e.*, the A position), and/or 3rd (*i.e.*, the U position) positions of the wild-type CAU anti-codon of the wild-type tRNAⁱⁿⁱ, wherein the mutated anti-codon does not comprise base change to A or U at the 1st position; and, (2) optionally, a further mutation that promotes folding of the A-stem of the tRNAⁱⁿⁱ.

In this regard, since the wild-type anticodon sequence is CAU, the first position of the mutated anticodon in the subject initiator tRNA is either the wild-type C, or a mutated G. The second position of the mutated anticodon can be any one of A, U, C or G. The third position of the mutated anticodon can also be any one of A, U, C or G. In other words, there are altogether 16 “GXX-type” mutant anticodon options (with each X being one of A, U, C, or G), and 15 “CXX-type” mutant anticodon options (with each X being one of A, U, C, or G, except for the wild-type anticodon CAU). See FIG. 1A.

As the data in FIG. 1B shows (also numerically summarized in the table of FIG. 1C), altering the anti-codon sequences, sometimes by just a single nucleotide, can have profound impact on the ability of such mutant initiator tRNA’s to initiate protein synthesis from a matching non-AUG “start codon” that fully complements the mutated anticodon sequence. For example, as shown in FIG. 1C, changing the wild-type CAU anticodon sequence to GAU (a mere one-nucleotide change) caused the mutant initiator tRNA bearing the GAU anticodon sequence to lose 96% of the ability to initiate protein synthesis at the matching AUC “start codon” (see the “4%” data value associated with the “GAU” anticodon sequence in FIG. 1C).

On the other hand, mutating the anticodon from CAU to CUA actually enhanced the ability of the mutant initiator tRNA with the mutated anticodon to initiate protein synthesis from the matching UAG codon (normally a stop codon) to 140% of the wild-type level (e.g., 100% for CAU as the anticodon).

Nevertheless, the CUA anticodon mutation is an exception, since all other anticodon mutations resulted in reduced ability to initiate from matching non-AUG start codon compared to the wild-type sequence (see FIG. 1C), though some such mutant initiator tRNA's still provide adequate level of protein synthesis initiation.

Thus, in certain embodiments, the engineered initiator tRNA retains at least about 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95% or more of the ability to initiate protein synthesis from a matching non-AUG codon, compared to that of the wild-type initiator tRNA to initiate protein synthesis from AUG start codon.

As used herein, any art-recognized method may be used to measure the ability of a given initiator tRNA to initiate protein synthesis from a codon matching / fully complementary to the mutated anticodon.

For example, the ability to initiate protein synthesis may be measured using any of the methods as described in the examples herein below, such as the HiBiT translation assays.

In one embodiment, the ability is measured by an *in vitro* translation system that utilizes a coding sequence having a start codon fully complementary to the anti-codon sequence of the initiator tRNA (*e.g.*, CAU in the wild-type anticodon as 100% initiation, and mutant "start codon" fully complement with the mutated anticodon sequence in the mutant initiator tRNA). In certain embodiments, the translation product of the *in vitro* translation system is a detectable protein or polypeptide, such as a fluorescent protein, or a protein with a sequence tag that can be readily detected and/or quantitated, such as a FLAG tag, 6His tag, polyglutamate tag, chitin binding protein (CBP) tag, maltose binding protein (MBP) tag, Strep-tag, glutathione-S-transferase (GST) tag, ALFA-tag, V5-tag, Myc-tag, HA-tag, Spot-tag, T7-tag or NE-tag.

In certain embodiments, the protein sequence tag is HiBiT-tag. HiBiT (Promega) is a small, 11-amino-acid, epitope tag capable of producing a bioluminescent signal when bound to its complementation partner, LgBiT. When the coding sequence comprises a "start codon" fully complementary to the mutated anticodon of a mutated initiator tRNA, and also comprises a coding sequence for HiBiT-tag (such as HiBiT-tag located at the C-terminal end of the encoded peptide), the ability of the mutated initiator tRNA to initiate translation from the fully complementary codon (which replaces a normal AUG start codon) and synthesize the polypeptide comprising the HiBiT-tag can be readily measured by using the HiBiT detection system comprising LbBiT (that binds HiBiT-tag).

In certain embodiments, the engineered initiator tRNA does not comprise the further mutation, and initiates translation from / incorporates the first amino acid in a nascent

polypeptide chain encoded by a polynucleotide comprising a cognate start codon for (*i.e.*, one that is fully reverse complementary to) the mutated anti-codon, at a relative efficiency of at least about 40%, 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140% or more, compared to the efficiency of the wild-type tRNAⁱⁿⁱ using the wild-type AUG start codon.

In certain embodiments, the engineered initiator tRNA does not comprise a mutated anticodon having any one of the following sequences: CUA, CCA, CGU, GGA, CAA, CGA, CAG, and CAC.

In certain embodiments, the engineered initiator tRNA comprises a mutated anticodon having any one of the following sequences: CUA, CCA, CGU, GGA, CAA, CGA, CAG, and CAC.

In certain embodiments, the engineered initiator tRNA retains less than 40%, 35%, 30%, 25%, 20%, 15%, 10%, 5% or has no detectable level of ability to initiate protein translation from a non-AUG start codon that fully complements the mutated anticodon, in the absence of a further mutation, and wherein the presence of the further mutation restores the ability to initiate protein translation above 40%, *e.g.*, to 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100% or more compared to the ability of wild-type initiator tRNA to initiate translation with a AUG start codon.

In certain embodiments, the initiator tRNA comprises the further mutation, which has enhanced ability to initiate translation from / incorporate the first amino acid in a nascent polypeptide chain encoded by a polynucleotide comprising a cognate start codon for the mutated anti-codon, compared to a control engineered tRNAⁱⁿⁱ having the mutated anti-codon but lacking the further mutation.

In certain embodiments, the further mutation promotes folding of the A-stem of the initiator tRNA (tRNAⁱⁿⁱ). The folding of the initiator tRNA can be predicted by any art-recognized tRNA secondary structure determination tool, such as the RNAfold webserver at <http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>. See representative results in FIGs. 2A, 2B, 3A, 5A and 5B). Alternatively or in addition, native gel electrophoresis can serve as a source to obtain information on folding, such as to verify that the mutant tRNA is folded in accordance with the native CAU initiator tRNA. Finally, NMR-spectroscopy can be used to fairly accurately determine folding.

In certain embodiments, the further mutation comprises a first double mutation in the stem region of the D-arm of the wild-type tRNAⁱⁿⁱ, wherein the first double mutation preserves an original G:C base pair in the stem region of the D-arm. In certain embodiments, the first double mutation is associated with a C-to-G change in the 1st position of the anti-

codon. In certain embodiments, the first double mutation is associated with a C-to-G change in the 1st position of the anti-codon, and comprises a C13G mutation and a G23C mutation.

In certain embodiments, the further mutation comprises a second double mutation comprising a C-to-G mutation in the loop region of the T-arm of the wild-type tRNAⁱⁿⁱ, and a G-to-C mutation in the loop region of the D-arm of the wild-type tRNAⁱⁿⁱ. While not wishing to be bound by any particular theory, it is believed that these mutations preserve an original tertiary G:C base pair between D- and T-arm. In certain embodiments, the second double mutation is associated with a U-to-G change in the 3rd position of the anti-codon. In certain embodiments, the second double mutation is associated with a U-to-G change in the 3rd position of the anti-codon, and comprises a C57G mutation and a G20C mutation.

In certain embodiments, the further mutation comprises both the first double mutation and the second double mutation as described herein.

While not wishing to be bound by any particular theory, it is believed that a mutated anticodon having G at the first position, when causing or associated with detrimental effect on translation initiation from a codon fully reverse complementary to the mutated anticodon, the first double mutation (*e.g.*, the C13G/G23C double mutation) would generally operate to alleviate or even eliminate the detrimental effect, to the extent that the ability to initiate translation from said codon fully reverse complementary to the mutated anticodon is substantially restored.

Similarly, a mutated anticodon having G at the third position, when causing or associated with detrimental effect on translation initiation from a codon fully reverse complementary to the mutated anticodon, the second double mutation (*e.g.*, the C57G/G20C double mutation) would generally operate to alleviate or even eliminate the detrimental effect, to the extent that the ability to initiate translation from said codon fully reverse complementary to the mutated anticodon is substantially restored.

Further, a mutated anticodon having G at the first and the third positions, when causing or associated with detrimental effect on translation initiation from a codon fully reverse complementary to the mutated anticodon, the first and the second double mutations (*e.g.*, the C13G/G23C/C57G/G20C quadruple mutation) would generally operate to alleviate or even eliminate the detrimental effect, to the extent that the ability to initiate translation from said codon fully reverse complementary to the mutated anticodon is substantially restored.

In certain embodiments, in the engineered initiator tRNA, the mutated anti-codon is GAA, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*,

SEQ ID NO: 2).

GGCGGGGUGG AGGAGCCUGG UACCUCGUCG GGCUGAAAAC CCGAAGGUCC
UCGGUUCAAA UCCGGCCCCC GCAACCA (SEQ ID NO: 2)

In certain embodiments, in the engineered initiator tRNA, the mutated anti-codon is GGA, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 3).

GGCGGGGUGG AGGAGCCUGG UACCUCGUCG GGCUGGAAAC CCGAAGGUCC
UCGGUUCAAA UCCGGCCCCC GCAACCA (SEQ ID NO: 3)

In certain embodiments, in the engineered initiator tRNA, the mutated anti-codon is GUA, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 4).

GGCGGGGUGG AGGAGCCUGG UACCUCGUCG GGCUGUAAAC CCGAAGGUCC
UCGGUUCAAA UCCGGCCCCC GCAACCA (SEQ ID NO: 4)

In certain embodiments, in the engineered initiator tRNA, the mutated anti-codon is GCA, wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 5).

GGCGGGGUGG AGGAGCCUGG UACCUCGUCG GGCUGCAAAC CCGAAGGUCC
UCGGUUCAAA UCCGGCCCCC GCAACCA (SEQ ID NO: 5)

In certain embodiments, in the engineered initiator tRNA, the mutated anti-codon is CAA, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 6).

GGCGGGGUGG AGCAGCCUGG UAGCUCGUCG GGCUCAAAAC CCGAAGGUCC
UCGGUUCAAA UCCGGCCCCC GCAACCA (SEQ ID NO: 6)

In certain embodiments, in the engineered initiator tRNA, the mutated anti-codon is CGA, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 7).

GGCGGGGUGG AGCAGCCUGG UAGCUCGUCG GGCUCGAAAC CCGAAGGUCC
UCGGUUCAAA UCCGGCCCCC GCAACCA (SEQ ID NO: 7)

In certain embodiments, in the engineered initiator tRNA, the mutated anti-codon is CUA, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 8).

GGCGGGGUGG AGCAGCCUGG UAGCUCGUCG GGCUCUAAAC CCGAAGGUCC
UCGGUUCAAA UCCGGCCCCC GCAACCA (SEQ ID NO: 8)

In certain embodiments, in the engineered initiator tRNA, the mutated anti-codon is CCA, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 9).

GGCGGGGUGG AGCAGCCUGG UAGCUCGUCG GGCUCCAAAC CCGAAGGUCC
UCGGUUCAAA UCCGGCCCCC GCAACCA (SEQ ID NO: 9)

In certain embodiments, in the engineered initiator tRNA, the mutated anti-codon is GAG, and wherein the further mutation comprises a C13G and a G23C double mutation, and a C57G and a G20C double mutation (*e.g.*, SEQ ID NO: 10).

GGCGGGGUGG AGGAGCCUGC UACCUCGUCG GGCUGAGAAC CCGAAGGUCC
UCGGUUGAAA UCCGGCCCCC GCAACCA (SEQ ID NO: 10)

In certain embodiments, in the engineered initiator tRNA, the mutated anti-codon is GGG, GUG, or GCG, and wherein the further mutation comprises a C13G and a G23C double mutation, and a C57G and a G20C double mutation.

GGCGGGGUGG AGGAGCCUGC UACCUCGUCG GGCUGGGAAC CCGAAGGUCC
UCGGUUGAAA UCCGGCCCCC GCAACCA (SEQ ID NO: 11)

GGCGGGGUGG AGGAGCCUGC UACCUCGUCG GGCUGUGAAC CCGAAGGUCC
UCGGUUGAAA UCCGGCCCCC GCAACCA (SEQ ID NO: 12)

GGCGGGGUGG AGGAGCCUGC UACCUCGUCG GGCUGCGAAC CCGAAGGUCC
UCGGUUGAAA UCCGGCCCCC GCAACCA (SEQ ID NO: 13)

In certain embodiments, in the engineered initiator tRNA, the mutated anti-codon is CAG, and wherein the further mutation comprises a C57G and a G20C double mutation (*e.g.*, SEQ ID NO: 14).

GGCGGGGUGG AGCAGCCUGC UAGCUCGUCG GGCUCAGAAC CCGAAGGUCC
UCGGUUGAAA UCCGGCCCCC GCAACCA (SEQ ID NO: 14)

In certain embodiments, in the engineered initiator tRNA, the mutated anti-codon is CGG, and wherein the further mutation comprises a C57G and a G20C double mutation (*e.g.*, SEQ ID NO: 15).

GGCGGGGUGG AGCAGCCUGC UAGCUCGUCG GGCUCGGAAC CCGAAGGUCC
UCGGUUGAAA UCCGGCCCCC GCAACCA (SEQ ID NO: 15)

In certain embodiments, in the engineered initiator tRNA, the mutated anti-codon is CUG, or CCG, and wherein the further mutation comprises a C57G and a G20C double mutation.

GGCGGGGUGG AGCAGCCUGC UAGCUCGUCG GGCUCUGAAC CCGAAGGUCC
UCGGUUGAAA UCCGGCCCCC GCAACCA (SEQ ID NO: 16)

GGCGGGGUGG AGCAGCCUGC UAGCUCGUCG GGCUCCGAAC CCGAAGGUCC
UCGGUUGAAA UCCGGCCCCC GCAACCA (SEQ ID NO: 17)

In certain embodiments, in the engineered initiator tRNA, the mutated anti-codon is GAU, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 18).

GGCGGGGUGG AGGAGCCUGG UACCUCGUCG GGCUGAUAAC CCGAAGGUCG
UCGGUUCAA UCCGGCCCC GCAACCA (SEQ ID NO: 18)

In certain embodiments, in the engineered initiator tRNA, the mutated anti-codon is GGU, GUU, or GCU, and wherein the further mutation comprises C13G and a G23C double mutation.

GGCGGGGUGG AGGAGCCUGG UACCUCGUCG GGCUGGUAAC CCGAAGGUCG
UCGGUUCAA UCCGGCCCC GCAACCA (SEQ ID NO: 19)

GGCGGGGUGG AGGAGCCUGG UACCUCGUCG GGCUGUUAAC CCGAAGGUCG
UCGGUUCAA UCCGGCCCC GCAACCA (SEQ ID NO: 20)

GGCGGGGUGG AGGAGCCUGG UACCUCGUCG GGCUGCUAAC CCGAAGGUCG
UCGGUUCAA UCCGGCCCC GCAACCA (SEQ ID NO: 21)

In certain embodiments, in the engineered initiator tRNA, the mutated anti-codon is CGU, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 22).

GGCGGGGUGG AGCAGCCUGG UAGCUCGUCG GGCUCGUAAC CCGAAGGUCG
UCGGUUCAA UCCGGCCCC GCAACCA (SEQ ID NO: 22)

In certain embodiments, in the engineered initiator tRNA, the mutated anti-codon is CUU, or CCU, and optionally there is no said further mutation.

GGCGGGGUGG AGCAGCCUGG UAGCUCGUCG GGCUCUUAAC CCGAAGGUCG
UCGGUUCAA UCCGGCCCC GCAACCA (SEQ ID NO: 23)

GGCGGGGUGG AGCAGCCUGG UAGCUCGUCG GGCUCCUAAC CCGAAGGUCG
UCGGUUCAA UCCGGCCCC GCAACCA (SEQ ID NO: 24)

In certain embodiments, in the engineered initiator tRNA, the mutated anti-codon is GAC, GGC, GUC, or GCC, and wherein the further mutation comprises C13G and a G23C double mutation.

GGCGGGGUGG AGGAGCCUGC UAGCUCGUCG GGCUGACAAC CCGAAGGUCG
UCGGUUCAA UCCGGCCCC GCAACCA (SEQ ID NO: 25)

GGCGGGGUGG AGGAGCCUGC UAGCUCGUCG GGCUGGCAAC CCGAAGGUCG
UCGGUUCAA UCCGGCCCC GCAACCA (SEQ ID NO: 26)

GGCGGGGUGG AGGAGCCUGC UAGCUCGUCG GGCUGUCAAC CCGAAGGUCG
UCGGUUCAA UCCGGCCCC GCAACCA (SEQ ID NO: 27)

GGCGGGGUGG AGGAGCCUGC UAGCUCGUCG GGCUGCCAAC CCGAAGGUCG
UCGGUUCAA UCCGGCCCC GCAACCA (SEQ ID NO: 28)

In certain embodiments, in the engineered initiator tRNA, the mutated anti-codon is CAC, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 29).

GGCGGGGUGG AGCAGCCUGG UAGCUCGUCG GGCUCACAAC CCGAAGGUCC
 UCGGUUCAA UCCGGCCCC GCAACCA (SEQ ID NO: 29)

In certain embodiments, in the engineered initiator tRNA, the mutated anti-codon is CGC, CUC, or CCC, and optionally there is no said further mutation.

GGCGGGGUGG AGCAGCCUGG UAGCUCGUCG GGCUCGCAAC CCGAAGGUCC
 UCGGUUCAA UCCGGCCCC GCAACCA (SEQ ID NO: 30)

GGCGGGGUGG AGCAGCCUGG UAGCUCGUCG GGCUCUCAAC CCGAAGGUCC
 UCGGUUCAA UCCGGCCCC GCAACCA (SEQ ID NO: 31)

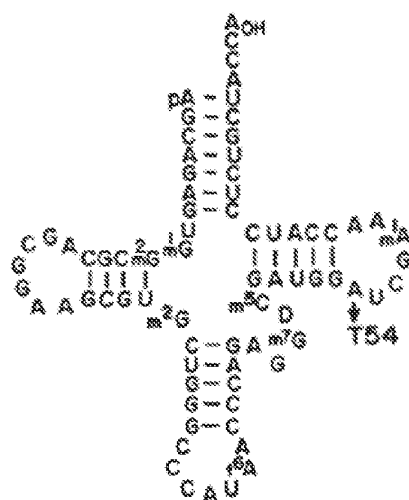
GGCGGGGUGG AGCAGCCUGG UAGCUCGUCG GGCUCCCAAC CCGAAGGUCC
 UCGGUUCAA UCCGGCCCC GCAACCA (SEQ ID NO: 32)

In certain embodiments, the engineered tRNAⁱⁿⁱ comprises, consists essentially of, or consists of the polynucleotide sequence of any one of SEQ ID NOs: 2-32.

In certain embodiments, the wild-type tRNAⁱⁿⁱ is a prokaryotic tRNAⁱⁿⁱ. For example, in certain embodiments, the wild-type tRNAⁱⁿⁱ is that of the *E. coli* fMet-tRNAⁱⁿⁱ (SEQ ID NO: 1), or any one of the bacterial initiator tRNA described in Marck and Grosjean (RNA 8:1189-232, 2002, incorporated herein by reference).

In certain embodiments, the wild-type tRNAⁱⁿⁱ is an archaea bacterial tRNAⁱⁿⁱ. For example, in certain embodiments, the wild-type tRNAⁱⁿⁱ is any one of the archaea initiator tRNA described in Marck and Grosjean (RNA 8:1189-232, 2002, incorporated herein by reference).

In certain embodiments, the wild-type tRNAⁱⁿⁱ is a eukaryotic tRNAⁱⁿⁱ. For example, in certain embodiments, the wild-type tRNAⁱⁿⁱ is that of a vertebrate (*e.g.*, human), an insect (*e.g.*, *Drosophila melanogaster*), a worm (*e.g.*, *C. elegans*), a yeast (*e.g.*, *S. pombe*, or *S. cerevisiae*), a plant (*e.g.*, *Arabidopsis thaliana*), or a fungi (*e.g.*, *Encephalitozoon cuniculi*) Met-tRNAⁱⁿⁱ. See Marck and Grosjean (RNA 8:1189-232, 2002, incorporated herein by reference). The human initiator tRNA sequence and secondary structure is shown below as SEQ ID NO: 35.



(human initiator tRNA)

Since the elucidation of the first nucleotide sequence of a mature tRNA molecule and its putative two-dimensional cloverleaf structure by Holley in 1965, more than 4,000 different mature tRNA and tRNA genes originating from a variety of organisms have been sequenced and subsequently compiled in the ever expanding tRNA data bank (550 mature tRNA sequences, 500 tRNA genes from various organisms, and 3,700 tRNA genes from fully sequenced genomes of 63 organisms as of 1998. See www.uni-bayreuth.de/departments/biochemie/trna/). This exponentially increasing information reinforces the evidence for a universally adopted cloverleaf secondary structure as initially proposed by Holley.

The sequence of the initiator tRNA gene (tDNA) is the most highly conserved among those of all tRNA species, across all three domains of life (Marck and Grosjean, RNA 8:1189-232, 2002, the entire content, including all disclosed initiator tRNA sequences therein, are incorporated herein by reference). This observation was made after Marck and Grosjean analyzed 50 genomes of the three domains of life (7 eukarya, 13 archaea, and 30 bacteria), and over 4,000 sequences corresponding to cytoplasmic, nonorganellar tRNAs. For each genome, the complete set of tRNAs required to read the 61 sense codons were identified. The initiator tDNA-iMet is the most conserved across the three domains, yet domain-specific signatures exist. Also, according to which tRNA feature is considered (5'-extra G encoded in tDNAs-His, AUA codon read by tRNA-Ile with anticodon CAU, presence of intron, absence of "two-out-of-three" reading mode and short V-arm in tDNA-Tyr) Archaea sequester either with Bacteria or Eukarya. No common features between Eukarya and Bacteria not shared with Archaea could be unveiled. Thus, from the tRNomic point of view, Archaea appears as an "intermediate domain" between Eukarya and Bacteria.

Further, all known vertebrate initiator tRNAs have identical sequences (Sprinzl *et al.*, Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* 26:148-153, 1998, the entire content, including all disclosed initiator tRNA sequences therein, are incorporated herein by reference). Specifically, Sprinzl *et al.* analyzed 3,279 sequences of tRNA genes and tRNAs published up to December 1996. The sequences, references and footnotes of tRNAs and tRNA genes listed in Table 1 of Sprinzl are deposited in the European Bioinformatics Institute (EBI) Data Library, and are available at www.uni-bayreuth.de/departments/biochemie/trna/ (incorporated herein by reference).

As used herein, the nucleotide positions of the initiator tRNA, such as C13, G23, C57, and G20, are based on the numbering scheme indicated in FIG. 1A or SEQ ID NO: 1, which depicts the initiator tRNA sequence of wild-type *E. coli* tMet-tRNAⁱⁿⁱ. Though all the initiator tRNA's do not have identical sequences or structure, in properly aligned tRNA sequences, nucleotides occupying the same position in different tRNA sequences (*e.g.*, those nucleotides referred to herein as "corresponding" nucleotides in different tRNA's) should play a comparable structural or functional role.

Therefore, as used herein, "a C13G mutation" as used herein not only includes a C-to-G substitution at position 13 of the initiator tRNA in FIG. 1A, but also any C-to-G substitution at an initiator tRNA position corresponding to position 13 of the initiator tRNA in FIG. 1A.

Likewise, "a G23C mutation" as used herein not only includes a G-to-C substitution at position 23 of the initiator tRNA in FIG. 1A, but also any G-to-C substitution at an initiator tRNA position corresponding to position 23 of the initiator tRNA in FIG. 1A.

"A C20G mutation" as used herein not only includes a C-to-G substitution at position 20 of the initiator tRNA in FIG. 1A, but also any C-to-G substitution at an initiator tRNA position corresponding to position 20 of the initiator tRNA in FIG. 1A.

"A G57C mutation" as used herein not only includes a G-to-C substitution at position 57 of the initiator tRNA in FIG. 1A, but also any G-to-C substitution at an initiator tRNA position corresponding to position 57 of the initiator tRNA in FIG. 1A.

In certain embodiments, the wild-type tRNAⁱⁿⁱ is not a mitochondrial tRNAⁱⁿⁱ.

Another aspect of the invention provides a translation system for the synthesis of a polypeptide, comprising the engineered initiator tRNA (tRNAⁱⁿⁱ) of the invention described herein above.

In certain embodiments, the translation system contains the initiator tRNA of the present invention, and other components sufficient to enable the synthesis of a peptide or

protein. Such other components may include an mRNA or a polynucleotide coding sequence for the peptide or protein, (large and small) ribosomal subunits, other non-initiator tRNAs charged with amino acid residues to be incorporated into the peptide or protein or aminoacyl tRNA synthetases (AARS's) required to charge tRNA with the cognate amino acids and the cognate amino acids, energy source required for peptide bond formation (*e.g.*, ATP, GTP), EF-Tu, translation initiation factors, translation elongation factors, translation termination factors, release factors (RF), ribosome regeneration factor (RRF), energy regenerating systems (creatine phosphate and creatine phosphokinase for eukaryotic systems, and phosphoenol pyruvate and pyruvate kinase for the *E. coli* lysate), and other co-factors (Mg^{2+} , K^+ , *etc.*).

In certain embodiments, the translation system comprises one or more (*e.g.*, all) of the 20-22 standard or proteinogenic amino acids.

In certain embodiments, the translation system comprises a non-standard amino acid, *e.g.*, an amino acid that is not the standard 20-22 proteinogenic amino acids. Such non-standard amino acids may comprise side chain groups not found in the 20-22 proteinogenic amino acids, or comprises side chain groups modified based on the 20-22 proteinogenic amino acids. In certain embodiments, the non-standard amino acids may comprise or may further comprise modification at the amino group of the amino acid (such as N-alkyl modified amino acids, such as N-methylated amino acids).

In certain embodiments, the translation system comprises D-amino acids, or L-amino acids, or both.

In certain embodiments, the translation system comprises β -amino acids.

In certain embodiments, the translation system comprises peptoid amino acids.

In certain embodiments, the translation system may comprise enzymes that charge non-standard amino acids to tRNA, such as initiator tRNA's of the invention. In certain embodiments, the translation system may comprise a flexizyme, which is a *de novo* ribozyme capable of charging a wide variety of non-natural amino acids on tRNAs (see Ohuchi *et al.*, The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus. *Curr Opin Chem Biol.* 11(5):537-42, 2007).

In certain embodiments, the translation system is an *in vitro* translation system. In certain embodiments, the translation system is a commercially available *in vitro* translation system, such as a rabbit reticulocyte *in vitro* translation system, a wheat germ extract *in vitro* translation system, an insect cell extract *in vitro* translation system, or an *Escherichia coli in vitro* translation system.

The rabbit reticulocyte lysate is a highly efficient *in vitro* eukaryotic protein synthesis system used for translation of exogenous RNAs (either natural or generated *in vitro*). Reticulocytes are highly specialized immature red cells that are primarily responsible for the synthesis of hemoglobin. More than 90% of the proteins made in the reticulocyte are hemoglobin. Reticulocytes have already lost their nuclei, but contain adequate mRNA, as well as complete translation machinery, for extensive globin synthesis. The endogenous globin mRNA is typically eliminated by incubation with Ca^{2+} -dependent micrococcal nuclease, which is later inactivated by chelation of the Ca^{2+} by EGTA. Ambion offers a nuclease-treated reticulocyte lysate. This type of lysate is the most widely used RNA-dependent cell-free system because of its low background and its efficient utilization of exogenous RNAs even at low concentrations.

Wheat germ extract is a convenient alternative to the rabbit reticulocyte lysate cell-free system. This extract has low background incorporation due to its low level of endogenous mRNA. Wheat germ lysate efficiently translates exogenous RNA from a variety of different organisms, from viruses and yeast to higher plants and mammals. Typically, translation by wheat germ extracts is more cap-dependent than translation by retic extracts. If capping of the RNA is impossible and the protein yield from an uncapped mRNA is low, the coding sequence can be subcloned into a prokaryotic vector and expressed directly from a DNA template in an *E. coli* cell-free system.

The *E. coli* cell-free systems consist of a crude extract that is rich in endogenous mRNA. The extract is incubated during preparation so that this endogenous mRNA is translated and subsequently degraded. Thus the levels of endogenous mRNA in the prepared lysate is low. In comparison to eukaryotic systems, the *E. coli* extract has a relatively simple translational apparatus with less complicated control at the initiation level, allowing this system to be very efficient in protein synthesis. Bacterial extracts are often unsuitable for translation of RNA, because exogenous RNA is rapidly degraded by endogenous nucleases. There are some viral mRNAs (TMV, STNV, and MS2) that translate efficiently, because they are somewhat resistant to nuclease activity and contain stable secondary structure.

Other related *in vitro* translation system using a ribosome of *Escherichia coli* is described in: Kung *et al.* (*JBC* 252:19, 6889-6894, 1977); Gonza *et al.* (*PNAS USA* 82:1648-1652, 1985); Pavlov and M. Ehrenberg (*Archives of Biochemistry and Biophysics* 328(1): 9-16, 1996); Shimizu *et al.* (*Nature Biotechnology* 19(8): 751-755, 2001); and Ohashi *et al.* (*Biochemical and Biophysical Research Communications* 352(1): 270-276, 2007).

Non-limiting examples of a commercially available cell-free translation system

include *Escherichia coli* derived systems such as RTS-100™ of Roche Diagnostics, reconstituted translation systems such as PURESYS™ of PGI and PUREx-press® In Vitro Protein Synthesis Kit of New England BioLabs and systems using a wheat germ extract such as those of ZOEGENE Corporation or CellFree Sciences.

Ultimately, the choice of *in vitro* cell-free translation system partly depends on whether the initiator tRNA is based on eukaryotic initiator tRNA or prokaryotic initiator tRNA.

In certain embodiments, the translation system of the invention may contain RNA polymerase for simultaneously performing transcription from DNA.

Another aspect of the invention provides a method of constructing a polypeptide library, the method comprising translating a coding sequence (or multiple coding sequences, each) encoding a member polypeptide of the polypeptide library, in a cell-free translation system comprising the engineered initiator tRNA (tRNAⁱⁿⁱ) of the invention described herein.

In a related aspect, the invention described herein provides a method of constructing a library of a complex between a polypeptide and an mRNA encoding the polypeptide, comprising translating in a cell-free translation system comprising the engineered initiator tRNA (tRNAⁱⁿⁱ) of the invention described herein.

In yet another related embodiment, the invention provides a polypeptide library constructed by the methods of the invention described herein, or a polypeptide-mRNA complex library constructed by the method of the invention described herein.

In certain embodiments, the polypeptide library comprises circular peptides, such as peptides having a disulfide bond covalently linking two residues of the peptide, one or both of which may be terminal residue(s) of the polypeptide.

Examples of a cyclization method are described in Kawakami *et al.*, *Nature Chemical Biology* 5, 888-890 (2009); Yamagishi *et al.*, *ChemBioChem* 10, 1469-1472 (2009); Sako *et al.*, *Journal of American Chemical Society* 130, 7932-7934 (2008); Goto *et al.*, *ACS Chemical Biology* 3, 120-129 (2008); Kawakami *et al.*, *Chemistry & Biology* 15, 32-42 (2008); WO2008/117833, WO2012/074129.

EXAMPLES

Example 1 Translation Initiation Efficiency Varies Using Orthogonal Anticodon Mutant Initiator tRNA

This example shows that translation initiation efficiency varies widely when the anticodon sequence of the *E. coli* initiator tRNA was mutated, and when such mutant initiator tRNAs were used to translate mRNA having a cognate non-AUG start codon.

FIG. 1A shows the secondary sequence of *E. coli* tRNAⁱⁿⁱ_{CAU} (SEQ ID NO: 1) as predicted by RNAfold. The various identity elements of the initiator tRNA were highlighted / color coded. The sequence in FIG 1A, with the anticodon CAU and the C13, G20, G23, and C57 nucleotides are double underlined):

GGCGGGGUGG AGCAGCCUGG UAGCUCGUCG GGCUCCAUAAC CCGAAGGUCG
UCGGUUCAAA UCCGGCCCCC GCAACCA (SEQ ID NO: 1)

Mutations to the 31 orthogonal anticodons are illustrated in the adjacent table. FIG. 1B shows HiBiT translation assays of 32 tRNAⁱⁿⁱ with orthogonal anticodons initiating translation from their respective cognate start codons and translating the coding sequence for peptide ^{Ac}FGGHGGHHGGGGSSKKSGWRLF (SEQ ID NO: 33). For example, when the anticodon was mutated to GAA, and the cognate start codon in the mRNA is UUC (as opposed to AUG). The result of translation efficiency using such initiator tRNA and mRNA pair was determined to be 8% of that of the wild-type initiator tRNA paired with AUG start codon.

Similar results were obtained for all 31 orthogonal mutations of the CAU anticodon (with the first position being either C or G but not A or U). The results were shown in FIG. 1B and summarized in FIG. 1C.

The data shows that certain anticodon mutations (such as CUA) are very efficiency at initiating translation at their cognate non-AUG start codons (such as UAG), while others essentially lost most or all initiation ability (such as GCC and CCC). Many mutant anticodon with G at the first position tend to fall into the second category of mutants.

To further investigate this phenomenon, tRNA fold analysis was conducted for each of these anticodon mutants, using the RNA secondary structure prediction software RNAfold. The results are shown in FIG. 2A, including the most frequent energy structure and their respective frequency in the structural ensemble as calculated by the RNAfold webserver. Bases are color coded according to their probability to be folded as predicted with dark red representing a probability of 1 to dark blue representing a probability of 0.

FIG. 2B shows the results of 20% native TBE gels run in 50 mM HEPES pH 7.2 and 3 mM MgCl₂. Secondary and tertiary structures of tRNA anticodon mutants are preserved in the native TBE gel.

The results suggests that many of those initiator tRNA with anticodon mutations may

be structurally heterogeneous and inactive, while alternative structures may be stable but they don't adopt the active tRNA fold required to initiate translation, thus negatively impacting their ability to initiate translation, even though the cognate non-AUG start codon perfectly matches the mutated anticodon in the initiator tRNA.

Example 2 *Structural Engineering of Orthogonal Anticodon Mutant Initiator tRNA to Improve Translation Initiation Efficiency*

FIG. 3A shows common misfolding patterns observed for tRNAs with anticodon mutations to a G in the 1st and/or the 3rd position. Examples shown are tRNAⁱⁿⁱ_{GCA} and tRNAⁱⁿⁱ_{GAG}. Mis-paired base-pairs and anticodons are highlighted in red. Native fold base-pairing in misfolded structures is indicated with a dashed grey line. As shown in FIG. 3A, the mutated anticodon GCA has a G in the first position of the anticodon, which G tends to base-pair with C13 which normally is involved in forming a GC base pair between C13 and G23 in the D arm of the initiator tRNA (see SEQ ID NO: 36).

Surprisingly, flipping this GC base pair to G13 (the "C13G" mutation) and C23 (the "G23C" mutation) prevents the 1st anticodon G to form a base pair with C13 in a misfolded (TA) stem and substantially restores the secondary structure of the initiator tRNA with the GCA anticodon sequence (see FIG. 3A and SEQ ID NO: 5).

Similarly, the mutated anticodon GAG has a G in the first position and the third position of the anticodon. The first position G also base-pairs with C13 normally in the D arm of the initiator tRNA, and the third position G base-pairs with C57 normally in the T loop. Flipping this GC base pair to G13 (the "C13G" mutation) and C23 (the "G23C" mutation), and flipping the tertiary GC base pair between C57 and G20 substantially restores the predicted secondary structure of the initiator tRNA with the GAG anticodon sequence (see FIG. 3A and SEQ ID NO: 10).

FIG. 3B shows HiBiT translation assays of mutant tRNAs relative to tRNAⁱⁿⁱ_{CAU}. FIG. 3C is a summary of mutant initiator tRNAs synthesized and their respective translation efficiency compared to the wildtype initiator tRNA. All translation efficiencies are normalized to tRNAⁱⁿⁱ_{CAU}.

Surprisingly, translation initiation efficiency is generally improved for anticodon mutants with G in the first position of the anticodon, when the C13G/G23C double mutations were introduced into such mutant initiator tRNAs (see FIG. 3C). For example, the GAA initiator tRNA, without the C13G/G23C double mutation, has a translation efficiency of

merely 8%. With the C13G/G23C double mutation, the translation efficiency is drastically improved to 118%.

Meanwhile, translation initiation efficiency is generally improved for anticodon mutants with G in the third position of the anticodon, when the C57G/G20C double mutations were introduced into such mutant initiator tRNAs (see FIG. 3C). For example, the CAG initiator tRNA, without the C57G/G20C double mutation, has a translation efficiency of about 41%. With the C57G/G20C double mutation, the translation efficiency is more than doubled to 83%.

Example 3 Orthogonal Initiator Codon Usage Assessment

IF3, responsible for accepting or rejecting particular codon-anticodon combinations, can allow base pair mismatches (particularly U/U and G/U) between the 3rd anticodon base/first codon base. This example demonstrates that some of the orthogonal initiator tRNAs having anticodon mutations can be used in the presence of a particular selection of alternative start codons with high specificity for their cognate codon for orthogonal use of tRNAs in translation.

FIG. 4A shows exemplary MALDI experiment design for orthogonal use of codons UUC, UUG, CUC, CUG and AUC. Each initiator tRNA (tRNAⁱⁿⁱ_{GAA}, for the UUC “start” codon, is displayed as example) is charged with acetylated phenylalanine (^{Ac}F) and translated individually using a library that encodes peptides with a different molecular weight depending on the start codon. Each peptide contains a FLAG tag at the C-terminal end of the peptides (highlighted in yellow). The displayed anticodon box highlights all initiator tRNAs tested with a white background and those used in the shown experiment in color. tRNAs denoted with * correspond to double mutations and ** to quadruple mutations listed in FIG. 3C. FIG. 4B shows an overlay of five MALDI spectra recorded for tRNAⁱⁿⁱ_{GAA} (red), tRNAⁱⁿⁱ_{CAA} (blue), tRNAⁱⁿⁱ_{GAG} (yellow), tRNAⁱⁿⁱ_{CAG} (cyan) and tRNAⁱⁿⁱ_{GAU} (purple) translated from the library displayed in FIG. 3A.

FIG. 4C shows percentage of non-cognate start codon translation initiation relative to the respective cognate codon initiation for each tRNA. The results show that the selected initiator tRNA's with mutated anticodons are generally quite specific for their cognate non-AUG start codons. For example, the initiator tRNA with GAG as the anticodon mostly only initiates translation from the cognate CUC start codon. Minor initiation at the UUC codon was also observed but only at about 9% compared to 100% at CUC. Initiation was not

detectable in all other non-matching start codons.

Similar results were also observed in FIGs. 6A-6D when assessing non-cognate start codon translation initiation.

FIGs. 5A-5B show mutant folding predictions and gel shift. FIG. 5A shows double mutant design for CAG and CGG. Structures shown represent the most frequent energy (mfe) structure within the ensemble. Percentages represent the predicted percentage of the RNA to fold as illustrated within the structural ensemble. Images and color coding are extracted from the RNAfold webserver. Mutated nucleotides are indicated by red arrows. FIG. 5B shows the result of 20% native TBE gels run in 50 mM HEPES pH 7.2 and 3 mM MgCl₂ of all mutated tRNAs compared to their wildtype variants.

The methods described herein below are suitable for use in the Examples herein. However, the methods are by no means limiting, and are provided solely for the purpose of illustration. On the other hand, the methods and conditions described herein are parts of the general description of the invention, and are thus specific embodiments that can be combined with one or more other embodiments described herein.

Methods

Transcription of tRNAs

DNA templates for transcription were first extended and then amplified in two steps by PCR according to standard protocols from NEB using PHUSION® High-Fidelity DNA Polymerase (0.1 μM of each Primer and template, 0.2 ng/μL DNA template, 200 μM dNTPs). PCR products were precipitated by addition of 0.1 M NaOAc and 2-fold volume of EtOH. The DNA pellet was dissolved in half the volume of the original PCR reaction in water.

In vitro transcription of tRNAs was performed in 40 mM Tris pH 8.0, 1 mM spermidine, 0.01 % (v,v) Triton X-100, 10 mM DTT, 22.5 mM MgCl₂, 0.0225 N KOH, 3.75 mM rNTPs, 5 mM GMP, 0.18 μM T7 RNA polymerase and 10% (v/v) amplified and concentrated PCR template and 1 U/mL of thermostable inorganic pyrophosphatase (NEB) at 37 °C overnight. Template DNA was removed by addition of 0.025 U/μL DNaseI (NEB) in 1× DNase buffer and incubation for 1 hour at 37°C. Following DNase treatment, the RNA was precipitated by addition of 30 mM EDTA, 0.3 M NaCl and finally 1× volume of isopropanol.

The RNA pellet was redissolved in water and purified by PAGE using 10% CRITERION™ TBE-Urea gels. The RNA was visualized by UV shadowing (254 nm),

excised from the gel and eluted in 0.3 M NaOAc. Eluted RNA was precipitated with two volumes of EtOH.

The purity of RNA transcripts was analyzed by polyacrylamide gel electrophoresis using 10% NOVEX™ TBE-Urea Gels. For analysis of RNA folding, samples were prepared in 50 mM HEPES pH 7.2, 3 mM MgCl₂ and 10% glycerol and analyzed using 20% NOVEX™ TBE Gels and imaged by UV shadowing at 254 nm.

HiBiT translation assay

HiBiT translation assays were performed as described previously. Briefly, HiBiT tagged (KKSGWRLE, SEQ ID NO: 34) peptides were translated in a 20 µL volume in the reprogrammed *in vitro* translation system. Each reaction contained 0.04 µM DNA template and 25 µM initiator tRNA (tRNA^{fMet} mutants aminoacylated with Ac-L-Phe), 0.2 mM of each required amino acid supplemented with LgBiT protein and the furimazine substrate (Promega). Luminescence was monitored by a SpectraMax M2 plate reader for two replicates over 120 minutes. Luminescence was normalized and calibrated to tRNA^{fMet}_{CAU} measured in parallel for all mutant tRNAs.

MALDI

20 µL *in vitro* translation mixture was diluted to 100 µL with PBS buffer and incubated for 1 h with anti-FLAG M2 magnetic beads (Sigma). The beads were washed three times with 400 µL of PBS buffer followed by elution in 20 µL Pierce IgG Elution Buffer. The solution was acidified to 0.1% trifluoroacetic acid and the purified peptides were desalted on a ZipTipC18 (Millipore) into 70% acetonitrile/water saturated with αcyano-4-hydroxycinnamic acid. Peptides were analyzed by MALDI-TOF MS using a RAPID-FLEX (Bruker) in reflector/positive mode.

CLAIMS

1. An engineered initiator tRNA (tRNAⁱⁿⁱ), comprising:
 - (1) a mutated anti-codon comprising one or more base changes at the 1st (*i.e.*, the C position), 2nd (*i.e.*, the A position), and/or 3rd (*i.e.*, the U position) positions of the wild-type CAU anti-codon of the wild-type tRNAⁱⁿⁱ, wherein the mutated anti-codon does not comprise base change to A or U at the 1st position; and,
 - (2) optionally, a further mutation that promotes folding of the A-stem of the tRNAⁱⁿⁱ.
2. The engineered tRNAⁱⁿⁱ of claim 1, wherein the further mutation comprises:
 - (i) a first double mutation in the stem region of the D-arm of the wild-type tRNAⁱⁿⁱ, wherein the first double mutation preserves an original G:C base pair in the stem region of the D-arm; optionally, the first double mutation is associated with a C-to-G change in the 1st position of the anti-codon; and/or,
 - (ii) a second double mutation comprising a C-to-G mutation in the loop region of the T-arm of the wild-type tRNAⁱⁿⁱ, and a G-to-C mutation in the loop region of the D-arm of the wild-type tRNAⁱⁿⁱ; optionally, the second double mutation is associated with a U-to-G change in the 3rd position of the anti-codon.
3. The engineered tRNAⁱⁿⁱ of claim 1 or 2, which does not comprise the further mutation, and which initiates translation from / incorporates the first amino acid in a nascent polypeptide chain encoded by a polynucleotide comprising a cognate start codon for the mutated anti-codon, at a relative efficiency of at least about 40%, 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140% or more, compared to the efficiency of the wild-type tRNAⁱⁿⁱ using the wild-type AUG start codon.
4. The engineered tRNAⁱⁿⁱ of any one of claims 1-3, wherein the mutated anti-codon is CUA, CCA, CGU, GGA, CAA, CGA, CAG, or CAC.
5. The engineered tRNAⁱⁿⁱ of claim 1 or 2, which comprises the further mutation, and which has enhanced ability to initiate translation from / incorporate the first amino acid in a nascent polypeptide chain encoded by a polynucleotide comprising a cognate start codon for the mutated anti-codon, compared to a control engineered tRNAⁱⁿⁱ having the mutated anti-codon but lacking the further mutation.
6. The engineered tRNAⁱⁿⁱ of claim 5, wherein the first double mutation is associated

with a C-to-G change in the 1st position of the anti-codon, and comprises a C13G mutation and a G23C mutation.

7. The engineered tRNAⁱⁿⁱ of claim 5 or 6, wherein the second double mutation is associated with a U-to-G change in the 3rd position of the anti-codon, and comprises a C57G mutation and a G20C mutation.
8. The engineered tRNAⁱⁿⁱ of claim 1, wherein:
 - (1) the mutated anti-codon is GAA, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 2);
 - (2) the mutated anti-codon is GGA, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 3);
 - (3) the mutated anti-codon is GUA, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 4);
 - (4) the mutated anti-codon is GCA, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 5);
 - (5) the mutated anti-codon is CAA, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 6);
 - (6) the mutated anti-codon is CGA, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 7);
 - (7) the mutated anti-codon is CUA, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 8);
 - (8) the mutated anti-codon is CCA, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 9);
 - (9) the mutated anti-codon is GAG, and wherein the further mutation comprises a C13G and a G23C double mutation, and a C57G and a G20C double mutation (*e.g.*, SEQ ID NO: 10);
 - (10) the mutated anti-codon is GGG, and wherein the further mutation comprises a C13G and a G23C double mutation, and a C57G and a G20C double mutation (*e.g.*, SEQ ID NO: 11);
 - (11) the mutated anti-codon is GUG, and wherein the further mutation comprises a C13G and a G23C double mutation, and a C57G and a G20C double mutation (*e.g.*, SEQ ID NO: 12);
 - (12) the mutated anti-codon is GCG, and wherein the further mutation comprises a C13G and a G23C double mutation, and a C57G and a G20C double mutation

(*e.g.*, SEQ ID NO: 13);

- (13) the mutated anti-codon is CAG, and wherein the further mutation comprises a C57G and a G20C double mutation (*e.g.*, SEQ ID NO: 14);
- (14) the mutated anti-codon is CGG, and wherein the further mutation comprises a C57G and a G20C double mutation (*e.g.*, SEQ ID NO: 15);
- (15) the mutated anti-codon is CUG, and wherein the further mutation comprises a C57G and a G20C double mutation (*e.g.*, SEQ ID NO: 16);
- (16) the mutated anti-codon is CCG, and wherein the further mutation comprises a C57G and a G20C double mutation (*e.g.*, SEQ ID NO: 17);
- (17) the mutated anti-codon is GAU, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 18);
- (18) the mutated anti-codon is GGU, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 19);
- (19) the mutated anti-codon is GUU, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 20);
- (20) the mutated anti-codon is GCU, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 21);
- (21) the mutated anti-codon is CGU, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 22);
- (22) the mutated anti-codon is CUU, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 23);
- (23) the mutated anti-codon is CCU, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 24);
- (24) the mutated anti-codon is GAC, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 25);
- (25) the mutated anti-codon is GGC, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 26);
- (26) the mutated anti-codon is GUC, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 27);
- (27) the mutated anti-codon is GCC, and wherein the further mutation comprises C13G and a G23C double mutation (*e.g.*, SEQ ID NO: 28);
- (28) the mutated anti-codon is CAC, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 29); or,
- (29) the mutated anti-codon is CGC, and optionally there is no said further

- mutation (*e.g.*, SEQ ID NO: 30).
- (30) the mutated anti-codon is CUC, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 31).
- (31) the mutated anti-codon is CCC, and optionally there is no said further mutation (*e.g.*, SEQ ID NO: 32).
9. The engineered tRNAⁱⁿⁱ of any one of claims 1-8, wherein the wild-type tRNAⁱⁿⁱ is a prokaryotic tRNAⁱⁿⁱ.
10. The engineered tRNAⁱⁿⁱ of claim 9, wherein the wild-type tRNAⁱⁿⁱ is the *E. coli*. fMet-tRNAⁱⁿⁱ (SEQ ID NO: 1).
11. A translation system for the synthesis of a polypeptide, comprising the engineered initiator tRNA (tRNAⁱⁿⁱ) of any one of claims 1-10.
12. A method of constructing a polypeptide library, the method comprising translating a coding sequence encoding a member polypeptide of the polypeptide library, in a cell-free translation system comprising the engineered initiator tRNA (tRNAⁱⁿⁱ) of any one of claims 1-10.
13. A method of constructing a library of a complex between a polypeptide and an mRNA encoding the polypeptide, comprising translating in a cell-free translation system comprising the engineered initiator tRNA (tRNAⁱⁿⁱ) of any one of claims 1-10.
14. A polypeptide library constructed by the method of claim 12, or a polypeptide-mRNA complex library constructed by the method of Claim 13.