

Measuring amber initiator tRNA orthogonality in a genomically recoded organism

Russel M. Vincent[†], Bradley W. Wright[†], Paul R. Jäschke^{†*}

[†]Department of Molecular Sciences, Macquarie University, Sydney 2109, New South Wales, Australia

*Correspondence and requests for materials should be addressed to PRJ (paul.jaschke@mq.edu.au)

ABSTRACT

Using engineered initiator tRNA for precise control of protein translation within cells has great promise within future orthogonal translation systems to decouple housekeeping protein metabolism from that of engineered genetic systems. Previously, *E. coli* strain C321.ΔA.exp lacking all UAG stop codons was created, freeing this ‘amber’ stop codon for other purposes. An engineered ‘amber initiator’ tRNA^{fMet}_{CUA} that activates translation at UAG codons is available, but little is known about this tRNA’s orthogonality. Here, we combine for the first time the amber initiator tRNA^{fMet}_{CUA} in C321.ΔA.exp and measure its cellular effects. We found that the tRNA^{fMet}_{CUA} expression resulted in a nearly 200-fold increase in fluorescent reporter expression with a unimodal population distribution and no apparent cellular fitness defects. Proteomic analysis revealed upregulated ribosome-associated, tRNA degradation, and amino acid biosynthetic proteins, with no evidence for off-target translation initiation. In contrast to previous work, we show that UAG-initiated proteins carry N-terminal methionine exclusively. Together, our results identify beneficial features of using the amber initiator tRNA^{fMet}_{CUA} to control gene expression while also revealing fundamental challenges to using engineered initiator tRNAs as the basis for orthogonal translation initiation systems.

KEYWORDS: amber suppression, start codon, tRNA, orthogonality, translation

The ability to precisely control an orthogonal protein translation system within engineered organisms is one of the grand goals of synthetic biology.¹ A distinct feature of the standard genetic code is the dual function of AUG as both the dominant start

codon as well as a methionine sense codon internal to genes. This dual function of the AUG start codon can create ambiguities in gene prediction and genetic design. For example, in recombinant protein design, in-frame AUG sense codons can behave as start codons if the RNA sequence upstream resembles a Shine-Dalgarno site.² Therefore, engineering an orthogonal system to initiate translation at a unique codon reserved solely for that function would be beneficial. Once established, the unique start codon could be paired with an orthogonal initiator tRNA aminoacylated with natural or unnatural amino acids with useful properties as has been done before in internal codons.³

Specific tRNAs recognise and decipher mRNA codons during translation and sequentially associate a bound amino acid to a growing polypeptide within the ribosome through three anticodon nucleotides (positions 34-36) on the anticodon loop. tRNAs are divided into two functionally and structurally distinct classes: initiator tRNAs and elongator tRNAs. Initiator tRNAs participate in the initiation phase of translation, recognizing start codons, and thus, incorporate the first amino acid. The three consecutive G:C base pairs in the tRNA anticodon stem is a crucial conserved element for initiator tRNA binding to the P-site in the 30S ribosomal subunit via interactions with the 16S ribosomal RNA.⁴⁻⁵ Additionally, the initiator tRNA's conserved C1:A72 base pair mismatch in the acceptor stem specifically interacts with methionyl-tRNA N-formyltransferase enzyme⁶, and initiation factor 2 (IF2)⁷, a protein vital for prokaryotic translation initiation.

Over the past two decades, efforts have focused on engineering stop codon suppression systems that use elongator tRNA aminoacyl-synthetase pairs as an orthogonal system that specifically recognises amber (UAG) sense codons to site-specifically incorporate unnatural amino acids⁸. Unnatural amino acid incorporation enables researchers to build engineered proteins with an extended range of structural and functional properties for applications in protein evolution⁹, therapeutics¹⁰, and elucidating structure and function¹¹. To prevent undesired natural amber codon suppression, competition for UAG codons from release factor 1, and non-specific unnatural amino acid incorporation, researchers have recently built a genomically recoded *Escherichia coli* strain, C321.ΔA with all instances of UAG stop codons and associated release factor (RF1) removed.¹²

The UAG amber stop codon has also been repurposed as a unique initiation codon in *E. coli* by creating an initiator tRNA_{CUA}^{fMet} with anticodon changed from the canonical CAU to CUA. This amber initiator has been shown in chloramphenicol acetyltransferase (CAT) reporter assays on bulk cell cultures to initiate translation *in vivo* from a UAG start codon.¹³ They found that initiation efficiency was only 50-60% relative to the AUG reporter and that it was aminoacylated predominantly by glutamine instead of methionine.¹³⁻¹⁴ Despite excellent prior work characterizing the amber initiator tRNA, a more complete picture of its interactions with the host cell is needed if an orthogonal translation system is to be based on this tRNA.

In this work, we characterized the phenotypic effects of expressing the amber initiator tRNA_{CUA}^{fMet} in the genomically recoded C321.ΔA.exp strain using a modular tRNA/reporter plasmid system. Using this system we found that expressing tRNA_{CUA}^{fMet} in *E. coli* C321.ΔA.exp resulted in initiation from a UAG start codon with similar population-level characteristics to that seen from the canonical AUG initiation codon and wild-type initiator tRNA. We also found previously uncharacterised effects on host proteome, but surprisingly, found no evidence for off-target translation initiation from genomic UAG codons. Additionally, we found that proteins initiated from tRNA_{CUA}^{fMet} carried methionine exclusively at their N-terminus. Together, this work reveals specific opportunities and challenges to using engineered initiator tRNAs in orthogonal translation systems.

RESULTS AND DISCUSSION

To measure orthogonality of the previously described amber initiator tRNA_{CUA}^{fMet}, we built a modular plasmid-based amber initiator tRNA expression and reporter system (Figure 1A). The modular system consists of the amber initiator plasmid (pULTRA::*tac-metY*(CUA)) (Figure 1A, top) harbouring the *metY*(CUA) gene expressing tRNA_{CUA}^{fMet},¹³ and a compatible reporter plasmid pQE-60::T5-*sfGFP* (Figure 1A, bottom) encoding one of three superfolder green fluorescent protein (sfGFP) variants differing in their start codon (either AUG, UAG, or GCC), where *sfGFP*(AUG) denotes the superfolder GFP gene with AUG start codon. To reduce host interference from release factor 1 competing with UAG codons we chose *E. coli* C321.ΔA.exp (Table 1) with a reduced genetic code as the host strain for this work.¹²

We used this plasmid system to measure translation initiation events due to $\text{tRNA}_{\text{CUA}}^{\text{fMet}}$ expression. In the absence of pULTRA::*tac-metY*(CUA) plasmid, the sfGFP(UAG) reporter fluorescence was indistinguishable from the negative control sfGFP(GCC) fluorescence (Figure 1B). However, when the $\text{tRNA}_{\text{CUA}}^{\text{fMet}}$ was present in the cell, sfGFP(UAG) reporter fluorescence increased 196-fold in the induced condition compared to the repressed condition (Figure 1B and Figure S1). Moreover, $\text{tRNA}_{\text{CUA}}^{\text{fMet}}$ expression resulted in sfGFP(UAG) fluorescence at a similar level to that of sfGFP(AUG) fluorescence when $\text{tRNA}_{\text{CUA}}^{\text{fMet}}$ was absent in the system (Figure 1B and Table S1). Fluorescence from reporter sfGFP(GCC) remained nearly unchanged in all conditions, consistent with previous findings.¹⁵ These data confirm the ability of $\text{tRNA}_{\text{CUA}}^{\text{fMet}}$ to efficiently initiate protein synthesis from the UAG stop codon at near-cognate AUG levels.¹³

Surprisingly, $\text{tRNA}_{\text{CUA}}^{\text{fMet}}$ expression caused an increase in the sfGFP(AUG) fluorescence per cell (Figure 1B, right blue bars) suggesting the possibility that $\text{tRNA}_{\text{CUA}}^{\text{fMet}}$ could also recognise and interact with AUG start codons. Previous work concluded that the wild-type host initiator tRNA (with a CAU as opposed to CUA anticodon) does not initiate translation from a UAG start codon.¹⁵ The reciprocal interaction of a CUA anticodon with an AUG start codon should also result in weak translation initiation.

We next used flow cytometry to measure the distribution of fluorescence within the population of *E. coli* C321.Δ*A.exp* cells harbouring the pULTRA::*tac-metY*(CUA) plasmid. All cell populations had unimodally distributed fluorescence in both repressed and induced conditions and had uniform population-wide increase in fluorescence in both the sfGFP(UAG) reporter and the sfGFP(AUG) reporter strains upon $\text{tRNA}_{\text{CUA}}^{\text{fMet}}$ induction (Figure 1C). This is the first measurement, to our knowledge, of population-level translation-initiation from the amber initiator tRNA.

We next performed growth assays to measure the fitness of cells harbouring pULTRA::*tac-metY*(CUA). We compared growth characteristics of cells harbouring plasmids expressing the amber initiator tRNA (pULTRA::*tac-metY*(CUA)) and the wild-type initiator tRNA (pULTRA::*tac-metY*) to the empty vector (pULTRA::*tac-Empty*) under inducing conditions. We found no significant effect (p-value > 0.5) from

expression of either initiator tRNA on maximal growth rate or maximal cell density (Figure 2 and Table S2).

To better understand the global proteomic effect of expressing the amber initiator tRNA in C321. Δ A.*exp*, as well as to investigate possible causes for the observed increase in AUG reporter fluorescence (Figure 1), we next measured global protein production in these cells. We employed the data-independent acquisition method SWATH-MS (Sequential Window Acquisition of all Theoretical fragment ion Mass Spectra) which enables the collection of large, accurate, and reproducible quantitative proteomic datasets.¹⁶⁻¹⁷ We collected data from *E. coli* C321. Δ A.*exp* samples with and without tRNA^{fMet}_{CUA} expression three and five hours post-induction.

We first interrogated the SWATH-MS data set with the aim of detecting canonical *E. coli* proteins with altered abundance under tRNA^{fMet}_{CUA} expression. Summed fragment ion peak areas (area under the curve) of measured precursor ions (peptides) allowed us to compare protein abundance between *E. coli* C321. Δ A.*exp* expressing tRNA^{fMet}_{CUA} versus the control. Three hours post-induction we identified only two differentially expressed proteins that reached our significance threshold (fold-change >1.5, p-value <0.01). The lactose operon repressor protein (LacI) was up-regulated 4.9-fold because it was constitutively expressed from the medium-copy pULTRA::*tac-metY*(CUA) plasmid (Figure 3A, brown point and Table S3). The FhuE receptor protein was also up-regulated 3.7-fold (Figure 3A, purple point). Together, these data show that at three hours post-induction, the amber initiator tRNA^{fMet}_{CUA} does not appear to have a large-scale effect on the C321. Δ A.*exp* proteome.

In contrast to three hours post-induction, at five hours we identified 262 differentially expressed proteins (Figure 3A and Table S4). To assess the biological relevance of the differentially expressed proteins, we performed enrichment analysis employing the Protein Analysis THrough Evolutionary Relationships (PANTHER) functional annotation tool.¹⁸ This analysis revealed that tRNA^{fMet}_{CUA} expression had a significant impact on levels of proteins involved in the ribosome and translation process, leading to enriched biological process categories such as ribosomal biogenesis and assembly, and peptide metabolic pathways (Figure S2 and Table S5). Our data also showed that ribosomal proteins were abundant in both tRNA^{fMet}_{CUA}-expressing and

control cells three hours post-induction, but that they declined five hours post-induction in control cells. In contrast, when tRNA^{fMet}_{CUA} was expressed in the cell, ribosomal protein levels did not decline (Figure 3B and Figure S3, cluster 3).

In addition to the sustained production of the majority of ribosomal structural proteins five hours post-induction, eight ribosomal RNA (rRNA) modification enzymes, two initiation factors (IF2 and IF3), and elongation factor 4 (LepA) were all significantly up-regulated 2.0 - 9.1-fold compared to control cells (Table S4). This expression profile was similar to ribosomal proteins that continued to be expressed at levels similar to three hours post-induction while levels in control cells declined (Figure S3, cluster 3). Notably, LepA plays a direct role in ribosomal biogenesis¹⁹ and a simultaneous increase in observed rRNA modification proteins could suggest a rise in mature ribosome concentration. A recent study has linked ribosomal rRNA biogenesis to the conserved anticodon stem sequence of the initiator tRNA.²⁰ Our results complement the findings of this previous study by showing that ribosomal protein is also highly upregulated in response to increased cellular levels of initiator tRNA. We speculate that an increased number of mature ribosomes, in addition to up-regulated initiation factors (IF2 and IF3), could explain the increased translation from the AUG start codon reporter we observed (Figure 1B, right blue bars), rather than tRNA^{fMet}_{CUA} directly interacting with AUG start codons.

Previous evidence showed that tRNA^{fMet}_{CUA} is a poor substrate for aminoacyl-tRNA synthetases^{13-14, 21} and increases the level of uncharged tRNA in the cell²¹ triggering the tRNA catabolic pathways (Figure 4). Additionally, the levels of uncharged tRNA^{fMet}_{CUA} may arise due to peptidyl tRNA hydrolase activity on amber initiator tRNA charged with glutamine.²² This potential imbalance in charged to uncharged tRNA ratio can be seen in our experiment through the increased abundance of poly(A) polymerase (PAP I) and polynucleotide phosphorylase (PNPase), both involved in tRNA degradation.²³

We also observe an upregulation of guanosine pyrophosphatase (RelA)²⁴ that is a key player in the cellular response to amino acid starvation through the stringent response.²⁵ Despite RelA upregulation, it is unlikely that cells expressing the amber initiator tRNA are activating the stringent response, which is normally characterized by a reduction in the number of ribosomal proteins and expression of starvation sigma

factors.²⁶ Neither of these responses are observed in our experiment, and instead we see sigma factor 70 (σ^{70}) increased 2.9-fold and increased ribosomal protein expression (Table S4). We do observe the up-regulation of leucine-responsive regulatory protein (Lrp) and transcriptional regulator TyrR, normally induced during the stringent response.²⁵ Both Lrp and TyrR, in the absence of leucine and aromatic amino acids respectively, induce amino acid biosynthesis.²⁷⁻²⁸ Providing further evidence that the stringent response is not being activated in response to the amber initiator tRNA expression, the increased levels of transcriptional regulators Lrp and TyrR in our work did not result in up-regulation of associated amino acid biosynthesis pathways normally observed in the stringent response.

We next searched our SWATH dataset for off-target peptides produced from open reading frames (ORFs) beginning with UAG start codons in the C321. Δ A.*exp* genome. The C321. Δ A.*exp* strain has 321 instances of functional UAG stop codons in the genome replaced with UAA stop codons¹² but we hypothesized that of the 27,108 UAG codons remaining in the genome, some may initiate translation in the presence of the amber initiator tRNA_{CUA}^{fMet}. To identify peptides initiating from these UAG codons we generated an off-target proteome database using the genome sequence of C321. Δ A.*exp* (Figure 5A). We considered all genomic UAG codons as potential start codons and considered UGA and UAA as stop codons. We did not treat UAG as a stop codon because C321. Δ A.*exp* lacks release factor 1 which is required for termination at UAG stop codons. We generated a list of 27,108 ORFs and filtered out all ORFs shorter than 150 nucleotides (50 amino acids) in length since small proteins (less than 50 amino acids) are generally ignored in whole-proteome studies to reduce background noise.²⁹ This filtering process resulted in an off-target database of 9,857 ORFs initiating from UAG start codons. To further filter the ORF list we identified those ORFs with a Shine-Dalgarno (SD) sequence upstream of the UAG start codon, resulting in 990 ORFs in the final off-target database (Figure 5A and Table S6). Additionally, we considered that tRNA_{CUA}^{fMet} could be charged with glutamine in addition to methionine^{13-14, 21} and thus created two versions of each protein in the database, one with an N-terminal methionine and one with a glutamine.

We then combined the off-target database with the conventional *E. coli* MG1655 database (4,443 proteins) and searched our generated spectral library against it. Using

the SWATH-MS data, we mined the combined spectral library and quantified 1,803 proteins across all three biological replicates that satisfied our significance criteria (protein and peptide level FDR <1% with at least 1 unique peptide observed). We observed 76 peptides that matched proteins from the off-target database, however, the identified peptides were shared with existing canonical *E. coli* proteins due to the presence of a UAG start codon upstream and in-frame with a canonical start codon (Figure 5B, case 2). We considered these peptides to originate from the conventional *E. coli* proteome since none were unique to the off-target database (Table S7). Despite not detecting any peptides originating from off-target ORFs we were able to detect peptides from four known proteins that have been previously shown to be present at 1-5 copies per cell in *E. coli* grown in LB: P02942 (Tsr protein), P04949 (FliC protein), P07017 (Tar protein), and P39332 (YjgH protein).³⁰ Together, these results show that despite detecting and quantifying over 1,800 proteins across a large dynamic range, we saw no evidence for tRNA^{fMet}_{CUA} expression leading to off-target translation.

To further probe the possibility of translation initiation from genomically encoded off-target ORFs with UAG start codons, we built a single-copy reporter plasmid to determine whether we could detect protein production at the low levels expected for these ORFs. We designed and built the bacterial artificial chromosome (BAC) plasmids pSMART-BAC::*pRha-NanoLuc*(UAG) and pSMART-BAC::*pRha-NanoLuc*(AUG) (Figure 6A) and transformed them into cells already containing the pULTRA::*tac-metY*(CUA) plasmid or with the empty vector pULTRA::*tac-Empty*. We performed a luciferase assay on cells induced with varying concentrations of rhamnose designed to induce varying amounts of NanoLuc luciferase production with constant induction of the amber initiator tRNA^{fMet}_{CUA} or empty vector. As expected, we found that NanoLuc(UAG) luminescence was absolutely dependent on the amber initiator tRNA^{fMet}_{CUA} expression (Figure 6B).

At all rhamnose concentrations, when the amber initiator tRNA^{fMet}_{CUA} was expressed, the luminescence resulting from NanoLuc(AUG) and NanoLuc(UAG) expression was equal (Figure 6B). We also found that the NanoLuc(AUG) reporter luminescence was independent of the presence of the amber initiator, having the same signal in the pULTRA::*tac-Empty* cells as those containing pULTRA::*tac-metY*(CUA).

This result is in contrast to what we found with the sfGFP reporter on a medium-copy vector (Figure 1) where the presence of the amber initiator seemed to boost production of sfGFP(AUG). One explanation for this different behaviour could be that it is a result of the vector copy number: at low-copy number the mRNA encoding the NanoLuc protein is likely limiting production of the protein, whereas at much higher copy number the ribosomal abundance may be limiting sfGFP production. If this is true, and the expression of the amber initiator tRNA increases ribosomal abundance (Figure 3), then it follows that increased ribosome abundance in the medium copy sfGFP(AUG) reporter strain would cause an increase in fluorescence due to increased translation initiation from the reporter mRNA. In contrast, at low-copy increased ribosomal abundance has no effect on the NanoLuc(AUG) reporter because mRNA levels are limiting, not ribosomal abundance.

To identify the N-terminal amino acid incorporated into the NanoLuc reporter protein we purified the protein from C321.ΔA.*exp* cells harbouring the pSMART-BAC::*pRha-NanoLuc*(UAG) and pULTRA::*tac-metY*(CUA) plasmids. We designed the pSMART-BAC::*pRha-NanoLuc*(UAG/AUG) plasmids to contain a C-terminal 6xHis-tag to enable purification of the full-length protein (Figure 6A). The internal and N-terminal peptides of the purified NanoLuc proteins were subjected to a targeted mass spectrometry technique called parallel reaction monitoring (PRM). The strength of the PRM method is that it provides high specificity and sensitivity and would allow us to specify peptides to target for accurate identification.³¹

The PRM assay showed that we could detect three peptides produced from tryptic digests of NanoLuc(UAG) (Figure 6C). The NanoLuc(UAG) N-terminal peptides contained methionine exclusively at the N-terminus (Figure 6D). The PRM assay was setup to agnostically target N-terminal peptides from NanoLuc(UAG) with any amino acid in the N-terminal position.²² This result is in contrast to previous work with the amber initiator tRNA, which showed both methionine and glutamine as the N-terminal amino acid from a reporter containing UAG as the start codon.³² However, these N-terminal amino acid differences could be a result of the different reporter gene used (NanoLuc versus CAT) or by the different bacterial strains used in the experiments (*E. coli* versus *Mycobacterium smegmatis*).

Using the PRM method and the single-copy NanoLuc(UAG) reporter plasmid, we searched for off-target peptides in C321.ΔA.*exp* cells harbouring the pSMART-BAC::*pRha-NanoLuc*(UAG) and pULTRA::*tac-metY*(CUA) plasmids. We also measured peptides from cells containing the pSMART-BAC::*pRha-NanoLuc*(UAG) and pULTRA::*tac-Empty* plasmids as a control. In addition to inducing the pULTRA plasmid, we also weakly induced the pSMART-BAC::*pRha-NanoLuc*(UAG) reporter with 100 μM rhamnose (Figure 6B) as an internal control to determine our ability to detect UAG-initiating proteins from single-copy genes. Since the PRM method derives its sensitivity from limiting the peptides the mass spectrometer measures to a pre-specified list, it was not possible to target all 990 UAG-initiating ORFs with a Shine-Dalgarno sequence in C321.ΔA.*exp*. Instead, we limited our search space to peptides from 194 off-target ORFs with the strongest upstream Shine-Dalgarno sequences (Table S6), in addition to the peptides of NanoLuc(UAG). Simultaneously, we targeted four weakly expressed *E. coli* host proteins previously observed in our SWATH-MS dataset. As before, we found no evidence for off-target protein expression from any of the off-target ORFs. We did detect one peptide that could have originated from a UAG-initiating ORF, but this peptide was observed in both the empty vector control and amber initiator-containing cells (Figure S5) suggesting that expression of this peptide is not dependent on tRNA_{CUA}^{fMet}. At the same time, we were able to detect peptides from the four weakly expressed proteins from the *E. coli* host proteome (Table S11) along with NanoLuc(UAG) reporter protein from the single-copy pSMART plasmid, but only when cells contained pULTRA::*tac-metY*(CUA), not pULTRA::*tac-Empty* (Figure 6E).

In conclusion, we found that the amber initiator tRNA_{CUA}^{fMet} efficiently initiates at the UAG amber stop codon from two different reporters, but does not initiate at detectable levels from genomic UAG codons. The amber initiator is thus shown to be orthogonal from a translation initiation point of view. Additionally, we showed that expressed NanoLuc(UAG) reporter proteins contain exclusively methionine at the N-termini, ensuring proteins created from this alternative initiation system have the same properties as wild-type proteins. We also show for the first time that amber initiator expression causes increased ribosome levels, tRNA degradation, and upregulation of amino acid biosynthetic proteins, although these effects do not seem to affect strain fitness. Together, these results point towards a potential role for the amber initiator

tRNA, after additional engineering to reduce its reliance on host aminoacyl-synthetases, in a fully orthogonal translation initiation system.³³

METHODS

Bacterial strains. All designed and constructed plasmids were expressed in the genetically recoded organism *E. coli* strain C321.ΔA.exp, which was a gift from George Church (Addgene #49018). All strain variants were grown in lysogeny broth Lennox (LB^L) or on LB^L agar, supplemented with zeocin for all *E. coli* strain C321.ΔA.exp cultures and carbenicillin for cultures with reporter plasmid and spectinomycin for cultures with amber initiator plasmid.

Plasmid construction. pQE-60::T5-sfGFP reporter plasmids were designed to contain superfolder GFP (sfGFP) with AUG/UAG/GCC start codons. The pULTRA initiator tRNA-expression plasmids were created by using Gibson assembly³⁴ to insert a 1971bp gBlock (IDT) contained *metY* genes from MG1655 (GenBank: U00096.3) with anticodon sequences of CAU (wild-type initiator) and CUA (amber initiator) (Tables S8 and S9) into the pULTRA-CNF³⁵ vector (Addgene #48215) (Figure 1A). To create an empty vector, we constructed the pULTRA::*tac*-Empty plasmid via PCR backbone amplification of pULTRA::*tac-metY*(CUA) and ligation, such that all elements of the pULTRA::*tac-metY*(CUA) plasmid except the *metY*(CUA) gene were retained. The pSMART-BAC::*pRha*-NanoLuc(UAG) and pSMART-BAC::*pRha*-NanoLuc(AUG) plasmids were constructed by cutting pSMART-BAC with NotI followed by Gibson assembly with gblocks encoding a rhamnose-inducible promoter (*pRha*) followed by the NanoLuc genes³⁶ with a C-terminal 6x-His tag and either UAG or AUG start codon. All plasmids were purified and stored in 10 mM Tris-Cl, pH 8.5. These plasmids were transformed into C321.ΔA.exp strains containing either pULTRA::*tac-metY*(CUA) or pULTRA::*tac*-Empty.

Culture growth conditions for assay measurements. Bacterial cultures were grown overnight at 37° C in LB^L with appropriate antibiotics. After overnight growth, each culture was diluted 1:100 into 400 μL LB^L in a 96-well deep well plate containing either 1 mM IPTG to induce, or 2% glucose (w/v) to repress, *metY*(CUA) expression. To control for the effects of the reagents used for repression and induction of the *tacI* promoter, we used the same repressed (2% glucose) and induced (1 mM IPTG) culture conditions for

all plasmid combinations, regardless of whether they contained a plasmid with a *tacI* promoter or not.

Fluorescence measurements. Measurements of fluorescence intensity from the amber initiator plasmid system was performed as before¹⁵ with the following modifications: OD₆₀₀ was used to estimate culture density, followed by fluorescence measurement (ex=485 nm, em=520 nm, bandwidth = 9 nm, PHERAstar FSX). Cultures from the above experiment were also measured on Cytoflex S flow cytometer (FITC, 488 nm excitation laser, 525/40 nm emission band-pass filter). 10,000 measured events were triggered on a side scattering threshold. Data was processed using FlowJo v10.

Fitness analysis. Strains of C321.ΔA.*exp* containing the pULTRA::*tac-metY*(CAU), pULTRA::*tac-metY*(CUA), or pULTRA::*tac*-Empty plasmid were grown overnight as described above. Cell densities were measured, and cells were passaged to a starting OD₆₀₀ of 0.1 into fresh 200 μL LB^L with appropriate antibiotics and 1 mM IPTG. Cultures were grown in a flat-bottom 96-well plate sealed with gas-permeable seal at 37°C (300 rpm). Absorbance at OD₆₀₀ over time was measured on a SPECTROstar NANO plate reader at 5-minute intervals. Growth rate (μ_r) and maximal cell density (max OD₆₀₀) was determined for each culture using the R package GrowthCurver³⁷ represented in Table S2.

Nanoluciferase Assay. Measurements of luminescence intensity from the amber initiator plasmid system was performed as before¹⁵ using the Nano-Glo Luciferase Assay System kit (Promega, #N1110) with the following modifications: expression of nanoluciferase was induced with the following concentrations of rhamnose: 0, 5, 50, 100, 500 μM, while 100 μL of culture were mixed with 100 μL of assay buffer and incubated for 15 min at room temperature, followed by luminescence measurement (em=230-750 nm, gain = 900, PHERAstar FSX).

NanoLuc expression and purification. Single colonies from C321.ΔA.*exp* strain harbouring pULTRA::*tac-metY*(CUA) and pSMART-BAC::*pRha-NanoLuc*(TAG) or pSMART-BAC::*pRha-NanoLuc*(ATG) as a control was used to inoculate 3 mL of LB^L supplemented with zeocin, spectinomycin, and chloramphenicol and shaken at 37°C overnight. We used the overnight cultures to purify C-terminal 6X histidine-tagged nanoluciferase protein following Ni-NTA purification as before¹⁵ with the following

modifications: expression of tRNA^{Met}_{CUA} and nanoluciferase was induced with 1 mM IPTG and 1 mM rhamnose.

Sample preparation for mass spectrometry proteome analysis. Strains were grown overnight and back diluted 1:100 into fresh media in 250 mL flasks with appropriate antibiotics and 1 mM IPTG. Cells were harvested three hours and five hours post-induction (3,500 x g/4°C/10 min). Cell pellets were lysed with 0.5 mL CellLytic™ B Cell Lysis buffer (Sigma-Aldrich #B7310) adding 0.1 mg lysozyme, 1X protease inhibitor (Roche #04693132001), and 25 units of benzonase (Sigma-Aldrich #E1014). The lysate was incubated (15 min/room temperature) then centrifuged (16,000 x g/room temperature/10 min). 1 mL of soluble protein supernatant was collected per culture and analysed with minor changes from previous work.³⁸ Proteins from whole cell lysate and Ni-NTA purified fractions were precipitated with 1 mL ice-cold acetone (-20°C/overnight) then centrifuged (15,000 x g/4°C/10 min), followed by protein pellet resuspension in 100 mM Tris-HCl, 8 M urea (pH 8.0). Protein samples were digested with 1:100 (w/w) protein/trypsin (37°C/overnight), then digested again (37°C/4 hours) with fresh trypsin as before.

Peptide library generation. To produce a peptide ion library, samples from biological triplicates were pooled and separated by high-pH RP-HPLC as described previously.³⁹ Mass spectrometry analysis was performed on a TripleTOF 6600 in data-dependent mode as previously described³⁹ except selection of ion precursors were performed over 140-minute runs. Spectra were interrogated using ProteinPilot 5.0 using the Paragon algorithm (SCIEX)¹⁷ and searched against the complete *E. coli* K-12 MG1655 protein sequence database from Uniprot (proteome ID-UP000000625 (last modified 2nd April, 2018)) and in-house generated C321.ΔA.exp UAG off-target database (File S5). The default paragon algorithm includes a fixed cysteine carbamidomethylation modification, caused by dithiothriitol and iodoacetamide treatment, and variable modifications sourced from <http://unimod.org>.¹⁶ Proteins and peptides were accepted with a protein false discovery rate of <1%.

SWATH-MS analysis. Spectra collection was performed in data independent mode as described previously³⁹ except, MS/MS spectra were collected across the *m/z* range of 350-1500 with a 35 ms accumulation time. Ion libraries of individual strains were imported into PeakView software 2.1 using the SWATH MicroApp 2.0 (SCIEX) and

matched against SWATH-MS data of individual replicates. Retention time calibration was performed using endogenous peptides and data were processed as previously described³⁹ to generate cumulative protein areas from extracted ion chromatograms which were exported for further analysis (Figure S3 and S4). Proteins with two or more peptides, p -value < 0.01 , and fold-change > 2 and < 0.5 were considered as significantly differentially expressed (Tables S3 and S4).

Gene ontology analysis. The differentially expressed proteins were analysed against the total set of 1,803 quantified proteins using the overrepresentation test in PANTHER (<http://pantherdb.org>) with default parameters.¹⁸ Gene ontology biological processes with $p < 0.05$ (after Binomial test with Bonferroni multiple hypothesis test correction) were considered significant.

Generation of in-house off-target database. To identify peptides initiating from these genomic UAG codons we generated an off-target proteome database using the genome sequence of C321. Δ A.*exp*. We considered all genomic UAG codons as potential start codons and considered UGA and UAA as stop codons. We did not treat UAG as a stop codon because C321. Δ A.*exp* lacks release factor 1 which is required for termination at UAG stop codons. We generated a list of 27,108 ORFs and filtered out all ORFs shorter than 150 nucleotides (50 amino acids) in length. This process resulted in an off-target database of 9,857 protein sequences that could potentially form functional ORFs initiating from UAG start codons (Figure 5A). To determine the SD-like sequences upstream of potential ORFs, we used a previously described method⁴⁰ to predict thermodynamic interactions between the anti-Shine Dalgarno and the upstream sequences of the potential ORFs, using the RNA coFold method of the ViennaRNA Package 2.0⁴¹. We used a similar binding energy threshold of -4.5 kcal/mol to filter down the off-target database, resulting in 990 potential off-target proteins with upstream SD-like sequences (Table S6).

Targeted protein mass spectrometry of N-terminal peptide and off-target proteins. The targeted mass spectrometry method of parallel reaction monitoring (PRM)³¹ was used to identify the N-terminal peptide species of the nanoluciferase reporter protein with a UAG start codon. Additionally, we analysed the whole cell proteome for off-target protein expression. Samples were analysed on a high resolution Q-Exactive mass spectrometer (ThermoFisher Scientific) coupled to an EASY-nLC1000

liquid chromatography system. Peptides were eluted over a 120-minute linear gradient with increasing concentration of elution buffer (99.9% (v/v) acetonitrile, 0.1% formic acid). The analysis consisted of one survey (full) scan at 70,000 resolution (400 m/z) for targeted precursors (Table S10 and Table S11) (AGC target $2e^5$, maximum injection time 100 ms, 2 m/z isolation width). Measured precursor ions were selected for HCD fragmentation with normalized collision energy of 30 followed by full ms/ms of product ions at 17,500 resolution (m/z 200, AGC target of $2e^5$ and 60 ms maximum injection time). Precursor and fragment ion spectra were analysed using Skyline version 4.2.⁴² PRM assay result for identification of N-terminal amino acid can be downloaded from the Panorama web repository server (<https://panoramaweb.org/H5K69A.url>).

SUPPORTING INFORMATION

Figure S1: Fold change in fluorescence due to tRNA^{Met}_{CUA} induction; Figure S2: Proportion of differentially expressed proteins represented in each gene ontology category; Figure S3: SWATH protein expression profile clusters; Figure S4: Evaluation of SWATH-MS quantification technique; Figure S5: Off-target peptide detected in C321.ΔA.exp cells regardless of amber initiator expression; Table S1: Translation initiation efficiency; Table S8: Reporter and tRNA gene sequences; Table S9: Synthesized oligonucleotides (PDF).

Table S2: A) Growth curve data represented as OD600 measured every 5 minutes B) Average specific growth rate (min⁻¹) and maximal OD600 (A.U.) (XLSX); Table S3: List of all quantified proteins using SWATH-MS technique three hours post-induction; Table S4: List of all quantified proteins using SWATH-MS technique five hours post-induction; Table S5: Gene ontology categorization of differentially expressed proteins five hours post-induction; Table S6: List of off-target ORFs harbouring SD-like sequences upstream; Table S7: List of all off-target proteins harmonized with Uniprot IDs from conventional *E. coli* database; Table S10: Parallel reaction monitoring targeted precursor list for N-terminal amino acid identification; Table S11: Parallel reaction monitoring targeted precursor list for off-target peptide identification (XLSX).

Supporting File S1: Amber initiator expression plasmid pULTRA_tac-metY(CUA) (GB); Supporting File S2: Empty vector control plasmid pULTRA_tac-Empty (GB); Supporting File S3: GFP reporter plasmid pQE-60_T5-sfGFP(UAG) (GB); Supporting File S4: Nanoluciferase reporter plasmid pSMART-BAC_pRha-Nanoluc(UAG) (GB); Supporting File S5: FASTA file of combined proteomic database including potential off-target proteins for mass spectrometry identification (FASTA).

ABBREVIATIONS

tRNA^{Met}_{CUA}, amber initiator tRNA with CUA anticodon; SWATH-MS, Sequential Window Acquisition of all Theoretical fragment ion Mass Spectra; IF1, Initiation Factor 1; IF2, Initiation Factor 2; IF3, Initiation Factor 3; sfGFP(AUG), superfolder green fluorescence

protein gene with wild-type AUG start codon; sfGFP(UAG), superfolder green fluorescence protein gene with amber stop codon UAG as start codon; NanoLuc(AUG), the NanoLuc gene with wild-type AUG start codon; NanoLuc(UAG), the NanoLuc gene with amber stop codon UAG as start codon.

AUTHOR INFORMATION

Corresponding Author

Paul R Jaschke

paul.jaschke@mq.edu.au

Department of Molecular Sciences, Macquarie University, NSW, 2109, Australia

ORCID

Russel M Vincent: 0000-0002-6976-2295

Bradley Wright: 0000-0002-6233-831X

Paul R Jaschke: 0000-0003-2359-1301

Author Contributions

RMV and PRJ conceived the study. RMV and BWW performed experiments, analyzed data, and created figures and tables. RMV and PRJ drafted the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

The authors declare no competing conflicts of interest.

ACKNOWLEDGEMENTS

We thank Matthew McKay, Ardeshir Amirkhani, David Cantor, Karthik Kamath and Dana Pascovici of the Australian Proteome Analysis Facility for assistance in protein and data analysis. We also thank Michael A. Sorensen, Richard Fahlman, Olivier Borkowski, Heinrich Kroukamp, and Jeff Glasgow for helpful discussions. PRJ was supported by the Molecular Sciences Department, Faculty of Science & Engineering, and the Deputy Vice-Chancellor (Research) of Macquarie University. BWW and RMV are recipients of the Macquarie University Research Excellence PhD scholarship (MQRES). Aspects of this research was conducted at the Australian Proteome Analysis Facility, facilitated by the Australian Government's National Collaborative Research Infrastructure Scheme.

REFERENCES

1. Liu, C. C.; Jewett, M. C.; Chin, J. W.; Voigt, C. A., Toward an orthogonal central dogma. *Nat. Chem. Bio.* **2018**, *14*, 103-106.
2. Whitaker, W. R.; Lee, H.; Arkin, A. P.; Dueber, J. E., Avoidance of Truncated Proteins from Unintended Ribosome Binding Sites within Heterologous Protein Coding Sequences. *ACS Synthetic Biology* **2015**, *4* (3), 249-257.
3. Liu, C. C.; Schultz, P. G., Adding new chemistries to the genetic code. *Annu Rev Biochem* **2010**, *79*, 413-44.
4. RajBhandary, U. L., Initiator transfer RNAs. *Journal of Bacteriology* **1994**, *176* (3), 547-552.

5. Lancaster, L.; Noller, H. F., Involvement of 16S rRNA Nucleotides G1338 and A1339 in Discrimination of Initiator tRNA. *Molecular Cell* **2005**, *20*, 623-632.
6. Varshney, U.; RajBhandary, U. L., Role of methionine and formylation of initiator tRNA in initiation of protein synthesis in *Escherichia coli*. *Journal of Bacteriology* **1992**, *174* (23), 7819-7826.
7. Guenneugues, M.; Caserta, E.; Brandi, L.; Spurio, R.; Meunier, S.; Pon, C. L.; Boelens, R.; Gualerzi, C. O., Mapping the fMet-tRNA(f)(Met) binding site of initiation factor IF2. *EMBO Journal* **2000**, *19* (19), 5233-5240.
8. Wang, L.; Schultz, P. G., A general approach for the generation of orthogonal tRNAs. *Chemistry and Biology* **2001**, *8* (9), 883-890.
9. Liu, C. C.; Mack, A. V.; Tsao, M. L.; Mills, J. H.; Lee, H. S.; Choe, H.; Farzan, M.; Schultz, P. G.; Smider, V. V., Protein evolution with an expanded genetic code. *Proceedings of the National Academy of Sciences* **2008**, *105* (46), 17688-17693.
10. Dozier, J. K.; Distefano, M. D., Site-specific pegylation of therapeutic proteins. *Int J Mol Sci* **2015**, *16* (10), 25831-25864.
11. Wang, J.; Xie, J.; Schultz, P. G., A genetically encoded fluorescent amino acid. *Journal of the American Chemical Society* **2006**, *128* (27), 8738-8739.
12. Lajoie, M. J.; Rovner, A. J.; Goodman, D. B.; Aerni, H. R.; Haimovich, A. D.; Kuznetsov, G.; Mercer, J. A.; Wang, H. H.; Carr, P. A.; Mosberg, J. A.; Rohland, N.; Schultz, P. G.; Jacobson, J. M.; Rinehart, J.; Church, G. M.; Isaacs, F. J., Genomically recoded organisms expand biological functions. *Science* **2013**, *342* (6156), 357-360.
13. Varshney, U.; RajBhandary, U. L., Initiation of protein synthesis from a termination codon. *Proceedings of the National Academy of Sciences of the United States of America* **1990**, *87* (4), 1586-1590.
14. Varshney, U.; Lee, C. P.; RajBhandary, U. L., From elongator tRNA to initiator tRNA. *Proceedings of the National Academy of Sciences of the United States of America* **1993**, *90* (6), 2305-9.
15. Hecht, A.; Glasgow, J.; Jaschke, P. R.; Bawazer, L. A.; Munson, M. S.; Cochran, J. R.; Endy, D.; Salit, M., Measurements of translation initiation from all 64 codons in *E. coli*. *Nucleic Acids Research* **2017**, *45* (7), 3615-3626.
16. Collins, B. C.; Hunter, C. L.; Liu, Y.; Schilling, B.; Rosenberger, G.; Bader, S. L.; Chan, D. W.; Gibson, B. W.; Gingras, A. C.; Held, J. M.; Hirayama-Kurogi, M.; Hou, G.; Krisp, C.; Larsen, B.; Lin, L.; Liu, S.; Molloy, M. P.; Moritz, R. L.; Ohtsuki, S.; Schlapbach, R.; Selevsek, N.; Thomas, S. N.; Tzeng, S. C.; Zhang, H.; Aebersold, R., Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *Nature communications* **2017**, *8* (1), 291.
17. Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R., Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Molecular & Cellular Proteomics* **2012**, *11* (6), O1111.016717-O1111.016717.
18. Mi, H.; Huang, X.; Muruganujan, A.; Tang, H.; Mills, C.; Kang, D.; Thomas, P. D., PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Research* **2017**, *45* (D1), D183-D189.
19. Gibbs, M. R.; Moon, K.-M.; Chen, M.; Balakrishnan, R.; Foster, L. J.; Fredrick, K., Conserved GTPase LepA (Elongation Factor 4) functions in biogenesis of the 30S subunit of the 70S ribosome. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114* (5), 980-985.

20. Shetty, S.; Varshney, U., An evolutionarily conserved element in initiator tRNAs prompts ultimate steps in ribosome maturation. *Proceedings of the National Academy of Sciences* **2016**, *113* (41), E6126-E6134.
21. Varshney, U.; Lee, C. P.; Seong, B. L.; RajBhandary, U. L., Mutants of initiator tRNA that function both as initiators and elongators. *Journal of Biological Chemistry* **1991**, *266* (27), 18018-18024.
22. Thanedar, S.; Kumar, N. V.; Varshney, U., The fate of the initiator tRNAs is sensitive to the critical balance between interacting proteins. *The Journal of biological chemistry* **2000**, *275* (27), 20361-7.
23. Svenningsen, S. L.; Kongstad, M.; Stenum, T. S.; Muñoz-Gómez, A. J.; Sørensen, M. A., Transfer RNA is highly unstable during early amino acid starvation in *Escherichia coli*. *Nucleic Acids Research* **2017**, *45* (2), 793-804.
24. Raina, M.; Ibba, M., tRNAs as regulators of biological processes. *Frontiers in genetics* **2014**, *5*, 171-171.
25. Traxler, M. F.; Summers, S. M.; Nguyen, H.-T.; Zacharia, V. M.; Hightower, G. A.; Smith, J. T.; Conway, T., The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Molecular microbiology* **2008**, *68* (5), 1128-48.
26. Sharma, U. K.; Chatterji, D., Transcriptional switching in *Escherichia coli* during stress and starvation by modulation of sigma 70 activity. *FEMS Microbiology Reviews* **2010**, *34* (5), 646-657.
27. Shimada, T.; Saito, N.; Maeda, M.; Tanaka, K.; Ishihama, A., Expanded roles of leucine-responsive regulatory protein in transcription regulation of the *Escherichia coli* genome: Genomic SELEX screening of the regulation targets. *Microbial genomics* **2015**, *1* (1), e000001-e000001.
28. Pittard, J.; Camakaris, H.; Yang, J., The TyrR regulon. *Molecular Microbiology* **2004**, *55* (1), 16-26.
29. Ramamurthi, K. S.; Storz, G., The small protein floodgates are opening; now the functional analysis begins. *BMC Biol* **2014**, *12*, 96.
30. Schmidt, A.; Kochanowski, K.; Vedelaar, S.; Ahrne, E.; Volkmer, B.; Callipo, L.; Knoops, K.; Bauer, M.; Aebersold, R.; Heinemann, M., The quantitative and condition-dependent *Escherichia coli* proteome. *Nature biotechnology* **2016**, *34* (1), 104-10.
31. Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J., Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Molecular & cellular proteomics : MCP* **2012**, *11* (11), 1475-88.
32. Govindan, A.; Miryala, S.; Mondal, S.; Varshney, U., Development of assay systems for amber codon decoding at the steps of initiation and elongation in mycobacteria. *J Bacteriol* **2018**, *200*, e00372-18.
33. Liu, C. C.; Jewett, M. C.; Chin, J. W.; Voigt, C. A., Toward an orthogonal central dogma. 2018; Vol. 14, pp 103-106.
34. Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. a.; Smith, H. O.; Iii, C. A. H.; America, N., Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **2009**, *6* (5), 343-5.
35. Schultz, K. C.; Supekova, L.; Ryu, Y.; Xie, J.; Perera, R.; Schultz, P. G., A genetically encoded infrared probe. *J Am Chem Soc* **2006**, *128* (43), 13984-5.
36. Hall, M. P.; Unch, J.; Binkowski, B. F.; Valley, M. P.; Butler, B. L.; Wood, M. G.; Otto, P.; Zimmerman, K.; Vidugiris, G.; Machleidt, T.; Robers, M. B.; Benink, H. A.; Eggers, C. T.; Slater, M. R.; Meisenheimer, P. L.; Klaubert, D. H.; Fan, F.; Encell, L. P.; Wood, K. V., Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chem Biol* **2012**, *7* (11), 1848-57.

37. Sprouffske, K.; Wagner, A., Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC bioinformatics* **2016**, *17*, 172.
38. Kamath, K. S.; Krisp, C.; Chick, J.; Pascovici, D.; Gygi, S. P.; Molloy, M. P., *Pseudomonas aeruginosa* Proteome under Hypoxic Stress Conditions Mimicking the Cystic Fibrosis Lung. *Journal of proteome research* **2017**, *16* (10), 3917-3928.
39. Bucio-Noble, D.; Kautto, L.; Krisp, C.; Ball, M. S.; Molloy, M. P., Polyphenol extracts from dried sugarcane inhibit inflammatory mediators in an *in vitro* colon cancer model. *Journal of proteomics* **2018**, *177*, 1-10.
40. Hockenberry, A. J.; Jewett, M. C.; Amaral, L. A. N.; Wilke, C. O., Within-Gene Shine-Dalgarno Sequences Are Not Selected for Function. *Molecular Biology and Evolution* **2018**, *35* (10), 2487-2498.
41. Lorenz, R.; Bernhart, S. H.; Höner zu Siederdissen, C.; Tafer, H.; Flamm, C.; Stadler, P. F.; Hofacker, I. L., ViennaRNA Package 2.0. *Algorithms for Molecular Biology* **2011**, *6* (1), 26.
42. MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J., Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics (Oxford, England)* **2010**, *26* (7), 966-8.

Table 1. Description of strains and plasmids used in this study.

Strain or Plasmid	Description	Reference or Source
Strain		
C321.ΔA.exp	<i>E. coli</i> MG1655 Δ(<i>ybhB-bioAB</i>):: <i>zeoR</i> Δ <i>prfA</i> ; all 321 UAG stop codons changed to UAA.	Addgene #49018
Plasmids		
pULTRA:: <i>tac-metY</i> (CUA)	pULTRA plasmid with <i>metY</i> gene containing amber initiator anticodon (CUA) driven by <i>tacI</i> promoter.	35 and this study
pULTRA:: <i>tac-metY</i> (CAU)	pULTRA plasmid with <i>metY</i> gene containing wild-type initiator anticodon (CAU) driven by <i>tacI</i> promoter.	35 and this study
pULTRA:: <i>tac</i> -Empty	pULTRA plasmid with <i>tacI</i> promoter, lacking any insert.	35 and this study
pQE-60:: <i>T5-sfGFP</i> (UAG)	pQE-60 plasmid with <i>sfGFP</i> gene initiated by UAG start codon and C-terminal 6xHis-tag inserted into the plasmid's multiple cloning site.	Qiagen and this study
pQE-60:: <i>T5-sfGFP</i> (AUG)	pQE-60 plasmid with <i>sfGFP</i> gene initiated by AUG start codon and C-terminal 6xHis-tag inserted into the plasmid's multiple cloning site.	Qiagen and this study
pQE-60:: <i>T5-sfGFP</i> (GCC)	pQE-60 plasmid with <i>sfGFP</i> gene initiated by GCC start codon and C-terminal 6xHis-tag inserted into the plasmid's multiple cloning site.	Qiagen and this study
pSMART-BAC:: <i>pRha-NanoLuc</i> (UAG)	pSMART-BAC plasmid with single-copy origin of replication (<i>oriS</i> , <i>repE</i> , <i>parABC</i>); TrfA-inducible high-copy origin of replication (<i>oriV</i>); Cm ^r -chloramphenicol resistance gene; NanoLuc gene with UAG start codon under control of rhamnose promoter (<i>pRha</i>) inserted into the vector's multiple-cloning site.	Lucigen Corp. and this study
pSMART-BAC:: <i>pRha-NanoLuc</i> (AUG)	pSMART-BAC plasmid with single-copy origin of replication (<i>oriS</i> , <i>repE</i> , <i>parABC</i>); TrfA-inducible high-copy origin of replication (<i>oriV</i>); Cm ^r -chloramphenicol resistance gene; NanoLuc gene with AUG start codon under control of rhamnose promoter (<i>pRha</i>) inserted into the vector's multiple-cloning site.	Lucigen Corp. and this study

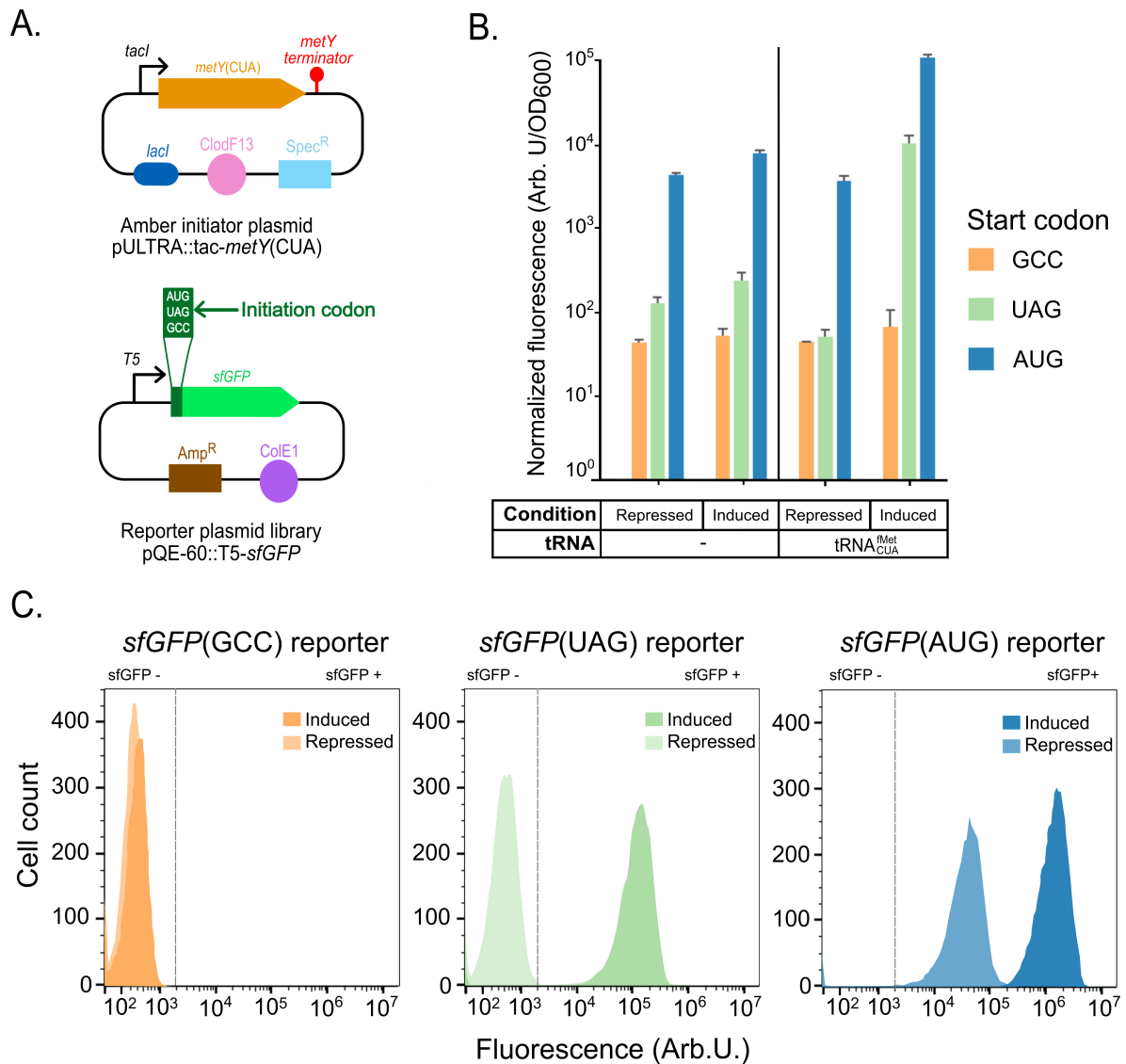


Figure 1. Expression of amber initiator tRNA causes increase in reporter fluorescence from both amber UAG and canonical AUG start codons. (A) Amber initiator and reporter plasmid system design. Amber initiator plasmid pULTRA::tac-metY(CUA) transcribing tRNA^{fMet}_{CUA}, and reporter plasmid pQE-60 library expressing sfGFP containing three different initiation codons (AUG, UAG, and GCC). (B) Normalized expression levels from sfGFP reporters under different tRNA^{fMet}_{CUA} transcription conditions. Repressed, 2% glucose; Induced, 1 mM IPTG; tRNA, presence or absence of tRNA^{fMet}_{CUA} expressing amber initiator plasmid. Average of three biological replicates with error bars showing one standard deviation. (C) Representative fluorescence (Arb. U.) histogram of gated populations containing pULTRA::tac-metY(CUA) expressing tRNA^{fMet}_{CUA} with pQE-60 reporter library. Induced cell population (darker shades) and repressed cell population (lighter shades) shown for each reporter. Vertical dashed lines demarcate fluorescent and non-fluorescent populations based on the 99th percentile of the control population (fluorescence of *E. coli* C321.ΔA.exp population expressing tRNA^{fMet}_{CUA} with sfGFP(GCC) reporter).

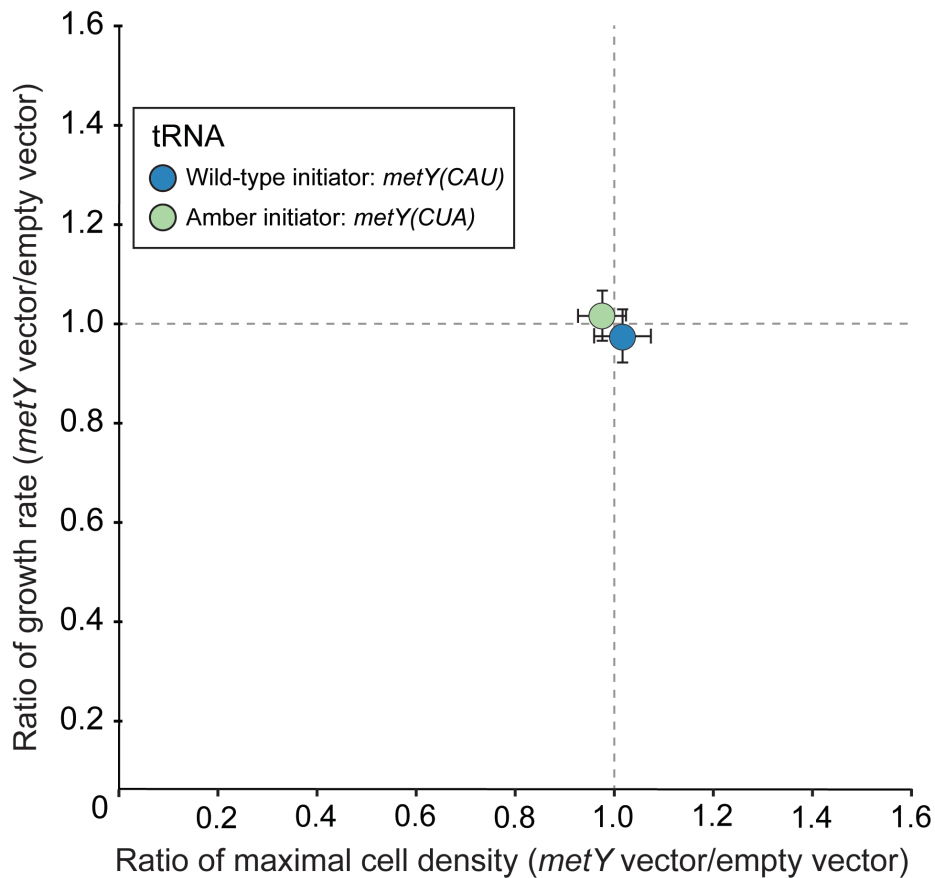


Figure 2. Amber initiator and wild-type initiator tRNA expression results in minimal fitness impact on C321. $\Delta A.exp$ cells. Ratios of maximal cell densities (horizontal axis) and growth rate (vertical axis) were determined for strains containing either pULTRA::*tac-metY*(CUA) or pULTRA::*tac-metY*(CAU) plasmid versus their corresponding strains with the empty vector pULTRA::*tac*-Empty. Strains unaffected by initiator tRNA expression are expected to have ratios of 1. Strains exhibiting slower growth are below the horizontal gray line, and strains exhibiting lower maximum cell density are to the left of the vertical gray line. All strains were grown under the same induced (1 mM IPTG) condition. Each color represents the average of biological replicates ($n = 3$), and error bars show one standard deviation.

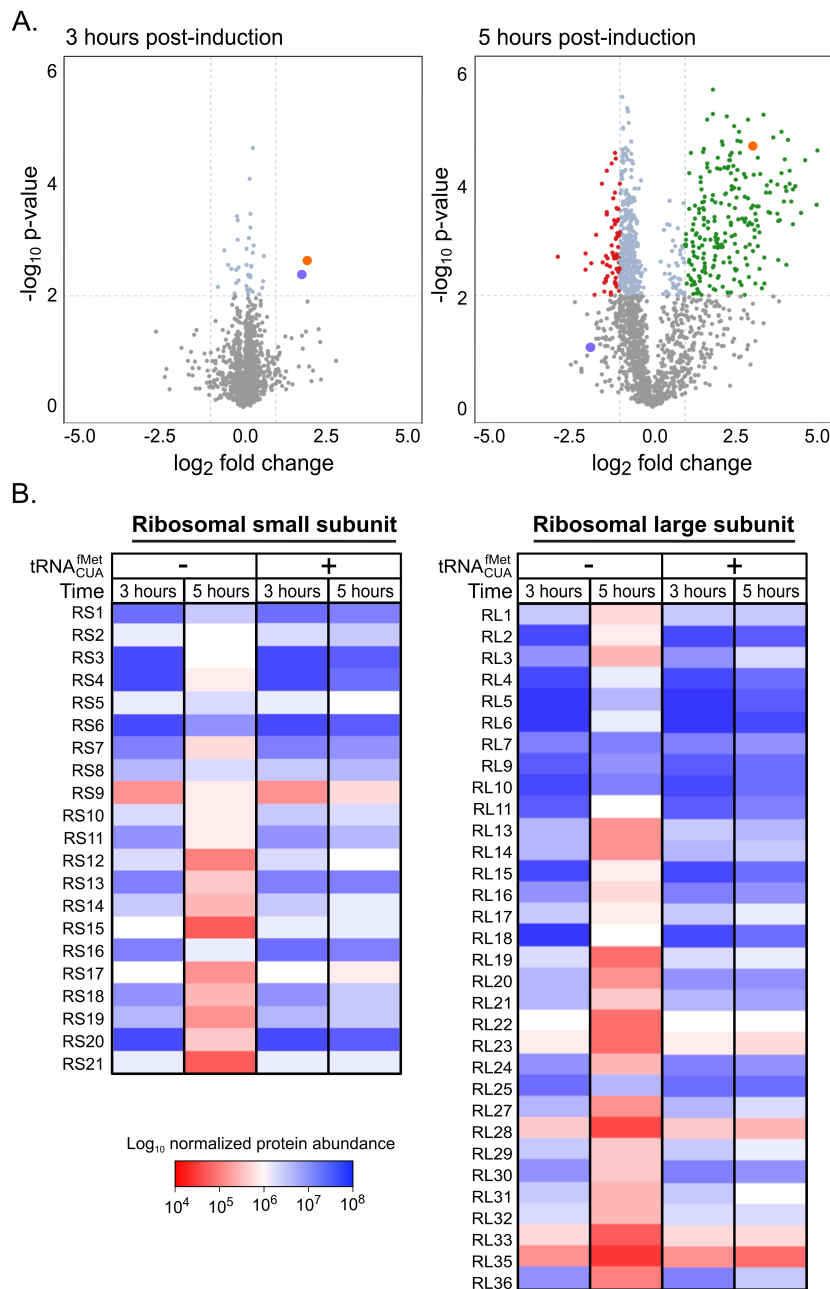


Figure 3. tRNA^{fMet}_{CUA} expression causes delayed downregulation of ribosomal proteins in C321.Δ*exp* cells. (A) Volcano plot of 1,803 quantified proteins from *E. coli* C321.Δ*exp* cells at 3-hours and 5-hours after amber initiator tRNA^{fMet}_{CUA} induction. The orange points represent LacI and purple points represent FhuE. Vertical grey dashed lines represent fold change threshold: $\log_2 \geq 1$ or ≤ -1 while the horizontal grey dashed line represents significance threshold: $-\log_{10} p\text{-value} \geq 2$ equivalent to $p\text{-value} \leq 0.01$. Green points represent significantly up-regulated proteins, red point represent significantly down-regulated, blue points represent unaltered expression, and grey points represent proteins that did not reach the significance threshold ($p\text{-value} \geq 0.01$). (B) Heat-map of ribosomal proteins quantified in C321.Δ*exp* control cells versus C321.Δ*exp*(pULTRA::*tac-metY*(CUA)) cells expressing tRNA^{fMet}_{CUA}. The C321.Δ*exp* control and C321.Δ*exp*(pULTRA::*tac-metY*(CUA)) strains were grown in LB in the presence of 1 mM IPTG and harvested at three and five hours post-induction.

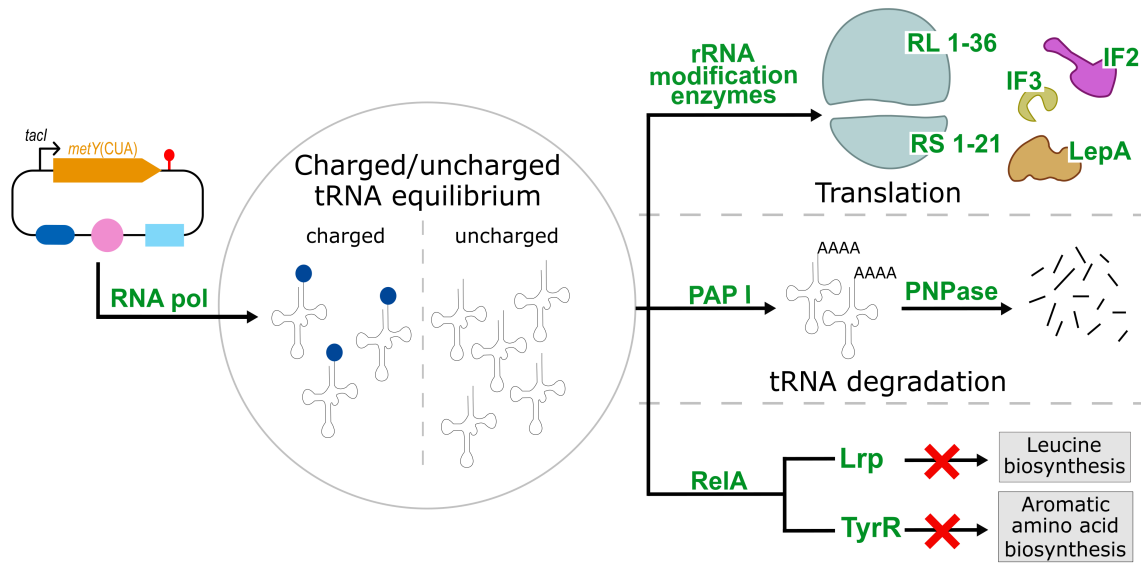


Figure 4. Proposed model of cellular effects of amber initiator tRNA^{fMet}_{CUA} expression. Increased RNA polymerase subunits transcribe tRNA^{fMet}_{CUA} from pULTRA::*tac-metY(CUA)* resulting in reduction of charged tRNA^{fMet}_{CUA}. Black arrows represent differentially regulated pathways, green text represent up-regulated proteins, and gray boxes depict unaltered pathways.

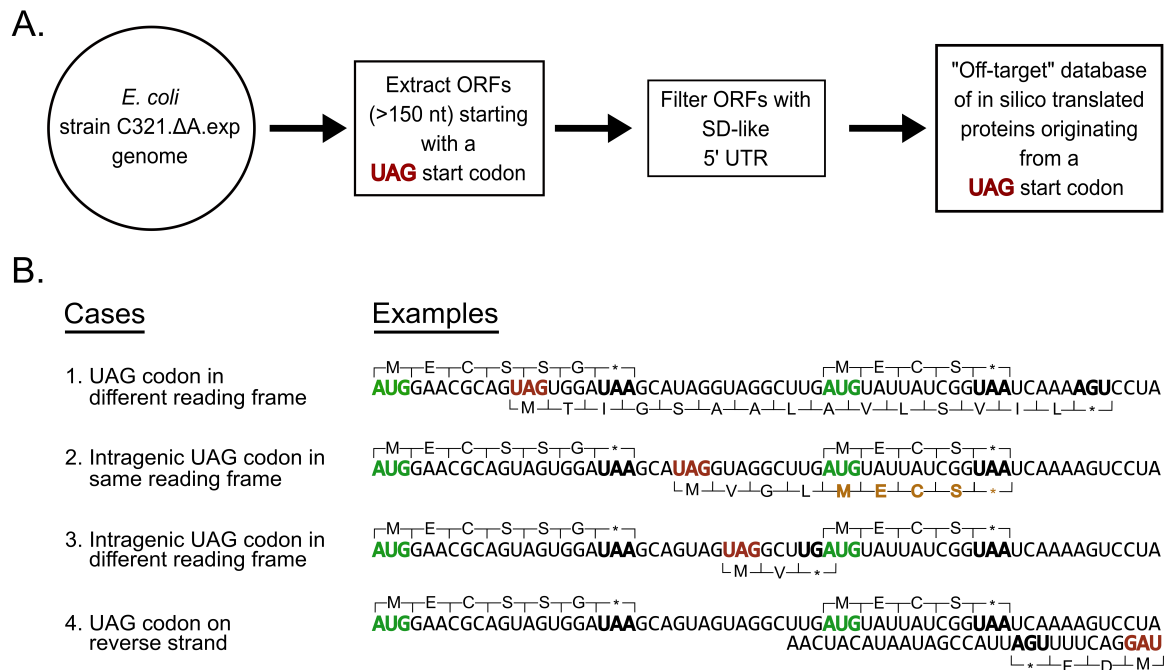


Figure 5. Method used to create UAG open reading frame off-target protein database. (A) Algorithmic workflow to create an off-target UAG-initiating open reading frame (ORF) database from the C321.ΔA.exp genome sequence. ORFs less than 150 nucleotides long were filtered out resulting in a reduction from 27,108 ORFs to 9,857. These ORFs were further filtered down to include only ORFs with Shine-Dalgarno-like 5' upstream sequences. The resulting 990 ORFs were then translated in two ways with UAG coding for methionine and glutamine to generate an off-target protein database. (B) UAG codon contexts that could result in ORFs with UAG start codon. The top proteins represent annotated sequences translated from canonical AUG start codon and bottom proteins represent hypothetical off-target protein sequences translated from a UAG start codon. ORFs displayed are for illustrative purposes only and are intentionally shorter than the minimum 150 nucleotide filter used to generate the off-target protein database. Green codons represent a canonical start codon; bold black codons represent a canonical stop codon in *E. coli* C321.ΔA.exp; red codons represent a UAG start codon; brown amino acids represent shared peptides with sequences translated from a canonical AUG start codon and an amber UAG start codon.

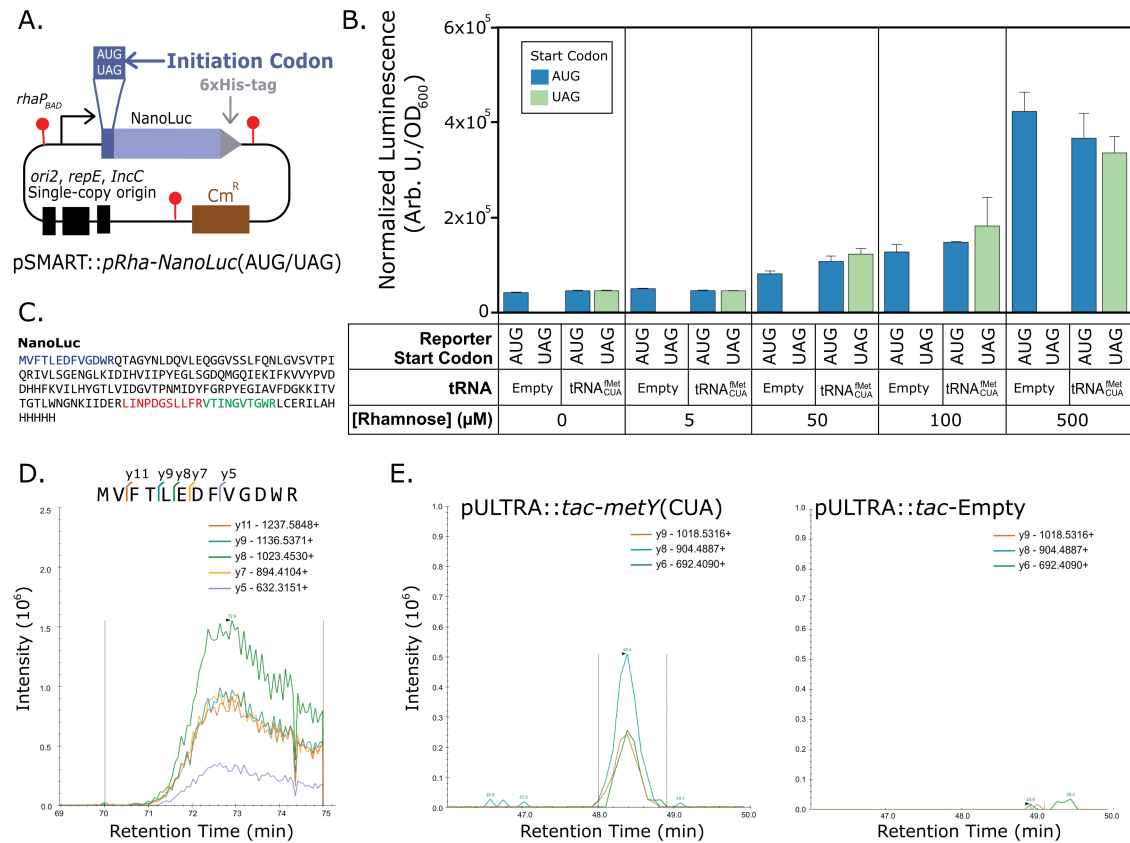


Figure 6. Single-copy bacterial artificial chromosome (BAC) reporter plasmid expressing NanoLuc shows N-terminal amino acid is methionine and that peptides are only observed when the amber initiator tRNA is co-expressed with the reporter. (A) Single-copy BAC NanoLuc plasmid design. Red ball and sticks represent transcriptional terminators; *ori2*, *repE*, and *IncC* encode single-copy origin. (B) NanoLuc luciferase reporter expression from C321.Δ*A.exp* cells harbouring either pULTRA::*tac-Empty* or pULTRA::*tac-metY*(CUA) tRNA expression plasmids along with either pSMART::*pRha-Nanoluc*(UAG) or pSMART::*pRha-Nanoluc*(AUG) reporter plasmids. (C) Amino acid sequence of NanoLuc protein with peptides observed by PRM mass spectrometry highlighted in color. (D) Mass spectrometric identification of N-terminal amino acid of purified NanoLuc protein expressed with UAG start codon. Product ions specific to a methionyl N-terminal peptide from NanoLuc(UAG) shown with identical retention times and expected masses. Peptide sequence and pre-cursor peptide fragmentation patterns shown above chromatogram. NanoLuc-6xHis was Ni-NTA purified from C321.Δ*A.exp*(pULTRA::*tac-metY*(CUA); pSMART::*pRha-Nanoluc*(UAG)) cells induced with 1 mM IPTG and 1 mM rhamnose. (E) NanoLuc peptides expressed from single-copy BAC detected in C321.Δ*A.exp* cells only when expressing the amber initiator. NanoLuc peptides observed in PRM mass spectrometry assay. Left, C321.Δ*A.exp*(pULTRA::*tac-metY*(CUA); pSMART::*pRha-Nanoluc*(UAG)) cells; Right, C321.Δ*A.exp*(pULTRA::*tac-Empty*; pSMART::*pRha-Nanoluc*(UAG)) cells.