Towards Reassignment of the Methionine Codon AUG to Two Different Noncanonical Amino Acids in Bacterial Translation

Alessandro De Simone,[1] Carlos G. Acevedo-Rocha,[2] Michael Georg Hoesl,[3] Nediljko Budisa[1,*]

[1] Department of Chemistry, Biocatalysis Group, Technical University Berlin/Berlin Institute of Technology, Müller-Breslau-Str. 10, Berlin DE-10623, Germany

[2] Biosyntia ApS, Fruebjergvej 3, boks 54, DK-2100 Copenhagen Ø, Denmark

[3] Clariant Produkte (Deutschland) GmbH, Semmelweisstraße 1, DE-82152 Planegg

* Corresponding author's e-mail address: budisa@chem.tu-berlin.de

───── THIS PAPER IS DEDICATED TO THE LOVING MEMORY OF IVANA WEYGAND-ĐURAŠEVIĆ (1952 – 2014) ─────

Abstract: Genetic encoding of noncanonical amino acids (ncAAs) through sense codon reassignment is an efficient tool for expanding the chemical functionality of proteins. Incorporation of multiple ncAAs, however, is particularly challenging. This work describes the first attempts to reassign the sense methionine (Met) codon AUG to two different ncAAs in bacterial protein translation. *Escherichia coli* methionyl-tRNA synthetase (MetRS) charges two tRNAs with Met: tRNA[fMet] initiates protein synthesis (starting AUG codon), whereas elongator tRNA[Met] participates in protein elongation (internal AUG codon(s)). Preliminary *in vitro* experiments show that these tRNAs can be charged with the Met analogues azidohomoalanine (Aha) and ethionine (Eth) by exploiting the different substrate specificities of *Ec*MetRS and the heterologous MetRS/tRNA[Met] pair from the archaeon *Sulfolobus acidocaldarius,* respectively. Here, we explored whether this configuration would allow a differential decoding during *in vivo* protein initiation and elongation. First, we eliminated the elongator tRNA[Met] from a methionine auxotrophic *E. coli* strain, which was then equipped with a rescue plasmid harboring the heterologous pair. Although the imported pair was not fully orthogonal, it was possible to incorporate preferentially Eth at internal AUG codons in a model protein, suggesting that *in vivo* AUG codon reassignment is possible. To achieve full orthogonality during elongation, we imported the known orthogonal pair of *Methanosarcina mazei* pyrrolysyl-tRNA synthetase (PylRS) / tRNA[Pyl] and devised a genetic selection system based on the suppression of an amber stop codon in an important glycolytic gene, *pfkA*, which restores enzyme functionality and normal cellular growth. Using an evolved PylRS able to accept Met analogues, it should be possible to reassign the AUG codon to two different ncAAs by using directed evolution.

Keywords: genetic code, sense codon reassignment, methionyl-tRNA synthetase, genetic selection system, synthetic biology, unnatural amino acid.

INTRODUCTION

Classical Methods for *in vivo* Multiple Incorporation of ncAAs

Over the last years, the genetic encoding of non-canonical amino acids (ncAAs) has allowed the introduction of many new chemical functionalities into target proteins. Generally, this is achieved by two main methods. On one hand, ncAAs that are structurally analogous (isostructural) to the canonical ones, and that can be recognized by the endogenous host cell machinery, can be encoded residue-specifically (*i.e.*, global reassignment) using the supplementation-based incorporation method (SPI).[1,2] On the other hand, ncAAs which are orthogonal to the host cell translation system are usually incorporated site-specifically by suppression of stop codons (SCS) using orthogonal aminoacyl-tRNA synthetase (aaRS)/tRNA pairs (o-pairs).[3]

Although the incorporation of one single ncAA into proteins has provided fascinating insights into sequence-function relationships[2–4] as well as potential applications in enzymology,[5–7] it is often desirable to introduce simultaneously different chemical functionalities by using two or

more ncAAs.[8,9] For example, Budisa and coworkers introduced three ncAAs in parallel into the model protein barstar using the SPI method:[10] the incorporation of the Trp analogue 4-azatryptophan (4AzaTrp) endowed the protein with a blue fluorescent probe for bio-imaging,[11] the Met analogue homopropargylglycine (Hpg) equipped it with a biorthogonal handle for post-translational conjugation by copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC),[12] while the Pro analogue (4*S*)-FPro ((4*S*)-FPro) provided a stabilizing effect on its structure.[13]

Global replacement of two or more amino acids, however, requires bacterial polyauxotrophic strains, its effects on protein function are not always predictable and it is difficult to fully substitute a high number of target residues with ncAAs. Thus, it is often difficult to correlate sequence to function. In contrast, codon suppression methodologies offer the possibility to incorporate multiple ncAAs at defined sites in a more controlled manner, as well as to greatly expand the repertoire of new chemical functionalities. For example, the combination of two o-pairs, pyrrolysyl-tRNA synthetase / tRNA^{Pyl} (PylRS / tRNA^{Pyl}) and *M. jannaschii* tyrosyl-tRNA synthetase / tRNA^{Tyr} (TyrRS / tRNA^{Tyr}), allowed the simultaneous suppression of the stop codons UAG and UGA with azide- and alkyne-containing amino acids for a biorthogonal cycloaddition reaction.[14] Further, the combination of the two above o-pairs with an evolved orthogonal ribosome for specific suppression of amber and quadruplet codons on an orthogonal mRNA was reported to be efficient.[15] Generally, it was possible to insert different pairs of ncAAs and label them selectively with fluorophores for protein conformation and dynamics studies by Förster resonance energy transfer (FRET).[16,17] The problem of suppression methodologies, however, is that these often require sophisticated systems (*e.g.*, specific o-pairs for each ncAA and o-ribosomes evolved by directed evolution) for enabling an efficient labeling of proteins and usually only the successful attempts are reported (*i.e.*, permissible sites of ncAA incorporation).

Importantly, the beneficial effects provided by the SPI method can be combined with the selective insertion of useful reactive groups by SCS. For example, a lipase mutant with enhanced activity was produced by globally replacing Met residues with norleucine (Nle), while introducing site-specifically the photo-crosslinker ncAA 4-benzoylphenyl-alanine (Bpa).[18] In another experiment, all the Met residues in the model protein green fluorescent protein (GFP) were replaced with the "clickable" ncAAs Aha or Hpg, whereas the photo-crosslinker 3,4-dihydroxyphenylalanine (L-DOPA) was incorporated at a specific site.[19]

Recent improvements in the SCS methodology, *i.e.* evolving more efficient aaRSs in a genomically recoded *E. coli* strain, have greatly increased the suppression efficiency allowing the multi-site incorporation of a single ncAA at up to 30 positions in protein-polymers with high yields.[20] Multiple site-specific incorporation of different ncAAs, however, is still dependent on the number of o-pairs available and, especially, restricted to three stop codons. It is expected that this limitation will be overcome in the future by recoding sense codons.

"Rewiring" the Genetic Code: Sense Codons Reassignment

A new strategy for incorporating multiple ncAAs is gaining increasing attention. It exploits the degeneracy of the genetic code to reassign sense codons, thus requiring o-pairs whose anticodon can be changed without affecting the tRNA recognition by the cognate synthetase, while still keeping their orthogonality towards the endogenous aaRSs. Compared to the SCS method, which usually entails competition between the suppressor tRNAs and the translation release factors, in sense codon reassignment the main limitation is the competition of the exogenous tRNA with endogenous tRNAs. Since tRNA levels in the cells correlate with codon usage, rare codons are expected to be more amenable to reassignment. Furthermore, reassigning rare codons should have less disrupting effects on the proteome. Recently, the rarest *E. coli* sense codon, AGG, was recoded from arginine (Arg) to L-homoarginine (LHR) using various tricks: all the AGG codons in essential genes were replaced with synonymous ones, the cellular levels of competing tRNAs^{Arg} were eliminated/reduced, and a heterologous PylRS/tRNA$^{Pyl}_{CCU}$ mutant specific for LHR was expressed.[21] In another example, it was possible to liberate the rare isoleucine codon AUA from its natural decoding mechanism. This was achieved by eliminating the enzyme catalyzing the tRNA modification C-to-L (2-lysyl-cytidine) at the wobble position of tRNA$^{Ile}_{CAU}$, which is important for codon recognition, and by rescuing the engineered cells with an heterologous IleRS/tRNA$^{Ile}_{UAU}$ pair.[22]

Interestingly, reassignment of abundant sense codons is also feasible. Tirrell used an orthogonal yeast PheRS/tRNA$^{Phe}_{AAA}$ mutant to outcompete the endogenous tRNA$^{Phe}_{GAA}$ in reading the UUU codon, one of the two Phe codons. This allowed almost quantitative reassignment of UUU codons to the Phe analogue 2-naphthylalanine in murine dihydrofolate reductase (mDHFR), but some misincorporation at the UUC codon also occurred.[23] Recently, O'Donoghue et al. showed that the machinery of selenocysteine can be modulated to recognize 58 of the 64 naturally occurring codons, in many cases completely outcompeting the endogenous tRNA, although other codons resulted in ambiguous translation.[24] Söll and co-workers showed that the serine sense codon AGU can be reassigned to 3-iodo-L-phenylalanine (3-I-Phe),[25] while Kwon and co-workers recently showed that it is possible to achieve ambiguous reading of the leucine UUG codon with 2-naphthylalanine.[26]

In general, amino acids encoded by several codons (*e.g.*, serine, leucine and arginine) offer more chances to be recoded since the wobble pairing at the 3rd base could be outcompeted by a more energetically favorable Watson-Crick base pairing with the orthogonal tRNA. On the other hand, the reassignment of the methionine codon, AUG, the only one encoding for this amino acid and highly frequent (around 2.5 %),[27] represents a real challenge. The AUG decoding system, however, has a peculiar mechanism which could be exploited to insert two different ncAAs. The starting AUG codon and internal AUG codons are read by two distinct tRNAs, respectively, initiator tRNAi^fMet (in prokaryotes) and elongator tRNA^Met (Figure 1). Charging them by two different and mutually orthogonal MetRSs can help to deliver one ncAA site-specifically to the N-terminus and the other globally at internal methionine residues. In this article we present some preliminary results for engineering this system.

REASSIGNMENT OF THE METHIONINE CODON AUG

Global Reassignment of the AUG Codon

Met plays an important role in cellular processes. Its sulphur atom can be involved in binding to metals ions, whereas its methyl group is used in DNA methylation through its derivative *S*-adenosylmethionine (SAM). Met is also the universal starting residue in translation, even though 60 % of these residues are removed by N-terminal Methionine excision (NME) in *E. coli*.[28] It is believed that Met starts translation because it is the most metabolically expensive amino acid to synthesize.[29,30] Partial recycling would therefore help the cells to save a part of the energetic cost of translation. Furthermore, Met functions as scavenger of reactive oxygen species (ROS) to protect proteins. In presence of ROS, the thioether moiety of Met can be oxidized to its sulfoxide form, Met(O), and then be reduced by methionine sulfoxide reductase.[31] Cells exposed to oxidative stress showed mis-methionylation of specific tRNAs which increases the insertion of protective Met residues in the whole proteome.[32]

Given its physiological importance, it might be more challenging to completely replace Met with a ncAA, similarly to the recently reported complete trophic reassignment of Trp with L-β-(thieno[3,2-*b*]pyrrolyl)alanine under experimentally designed evolutionary pressure.[33,34] Nevertheless, like Trp, Met is one of the least common amino acids in proteins: it represents only about 2.5 % of all residues,[35] mostly buried in the hydrophobic core of proteins where it is rarely directly involved in catalytic function. This feature makes it a good candidate for proteome-wide substitution using the SPI method. Indeed, the relatively relaxed substrate specificity of MetRS has
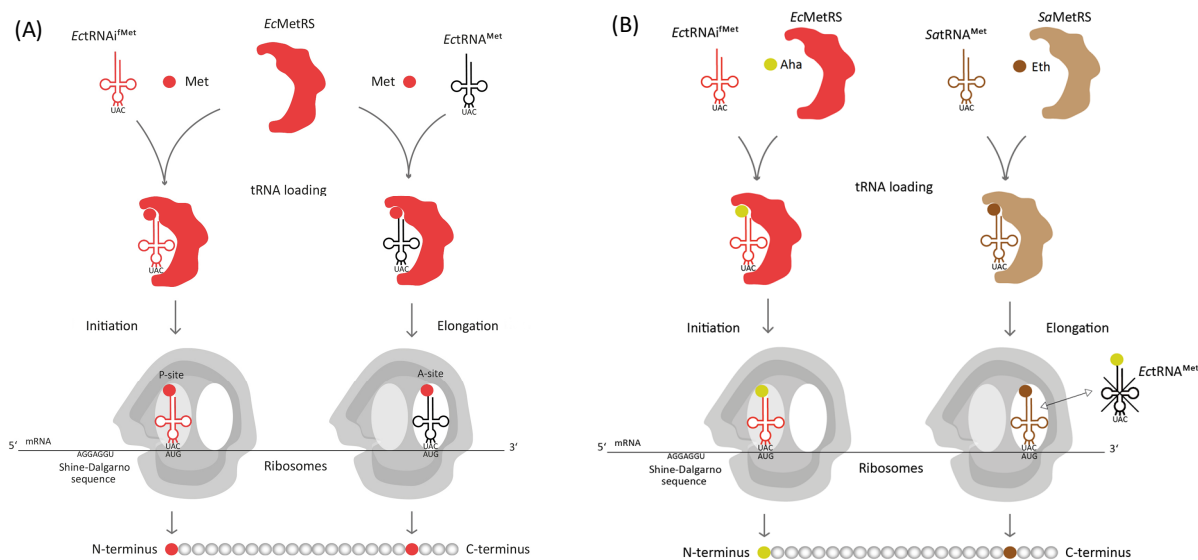
Figure 1. Natural (A) and hypothetical (B) AUG decoding system. In *E. coli,* the MetRS normally activates methionine and charges it onto both the initiator tRNAi^fMet (red) and elongator tRNA^Met (black). Initiator tRNAi^fMet is directed in the P-site of the ribosome where it participates in initiation and inserts an fMet residue at the N-terminus of the protein. Elongator tRNA^Met, instead, enters the A-site of the ribosome where it reads internal AUG codons and inserts Met in the protein sequence. Our hypothetical model includes context-dependent AUG-codon reassignments whereby one ncAA can be encoded site-specifically at the N-terminus and the other ncAAs can be inserted globally at the internal AUG positions.

allowed the incorporation of a large repertoire of Met analogues for different purposes. Labeling of proteins with selenomethionine (SeMet) and telluromethionine (TeMet) has become a common practice for structure determination by X-ray crystallography.[36,37] Fluorinated analogues like difluoromethionine (Dfm), trifluoromethio-nine (Tfm) and 6,6,6-trifluoronorleucine (Tfn) have been used as ^{19}F-NMR probes of protein structure, to increase protein stability and to design teflon-like or "non-stick" proteins.[38–40] The more hydrophobic[41] analogues norleucine (Nle) and ethionine (Eth) were used to study overall effects on enzyme activity. Their incorporation in the Met-rich active site of Calmodulin lowered its activity,[42] whereas incorporation of Nle into Lipase from *Thermoanaerobacter thermohydro-sulfuricus* (TTL) led to an enzyme highly active in the aqueous phase without the need for thermal activation[43] and enhanced stability against denaturing agents.[44] Other examples of Met analogue applications include the use of methoxinine to study the effects of Met oxidation on prion proteins,[45] of photo-methionine in protein-protein interaction studies,[46] and of homoallylglycine (Hag) for potential use in olefin metathesis.[47] Even Met analogues which are generally translationally inactive can be incorporated by simple overexpression of MetRS, as Tirrell and coworkers showed with *trans*-crotylglycine (Tgc), *cis*-crotylglycine (Ctg), 2-aminoheptanoic acid, norvaline, 2-butynylglycine, allylglycine[48–50] and *S*-allyl-homocysteine.[51]

Met analogues containing azides and alkynes are of particular interest since they can be used as chemical handles to mimic post-translational modifications using "click chemistry". Thus far, azidohomoalanine (Aha) and homopropargylglycine (Hpg) have been incorporated using the native *Ec*MetRS,[52,53] while azidonorleucine (Anl)[54] and propargylglycine (Pra)[51] were only translationally acive using mutants of *Ec*MetRS. Especially, Aha and Hpg have been used to artificially attach to proteins post-translational modifications such as sugars[55] or biotin.[56] The recently developed BONCAT (bioorthogonal non-canonical amino acid tagging) strategy takes advantage of the incorporation of azide-bearing ncAAs in the whole proteome and selective enrichment by affinity tags (*e.g.* biotin) to study temporal and spatial characteristics of newly synthesized proteins.[57] A cell-selective labeling in co-culture was achieved expressing an *Ec*MetRS mutant specific for Anl (NNL-*Ec*MetRS), and one mutant specific for Pra (PraRS), only in a subset of cells.[58,59] Figure 2 presents a survey of various Met analogues incorporated so far in a residue-specific fashion. The extent of their incorporation in newly expressed proteins was quantified by precise analytical measurements in some instances and found to be almost quantitative.[53,60,61]

Site-Specific Reassignment of the AUG codon

Met analogues contain single atomic substitutions like H → F, C → N or S → CH/CH$_2$/O/N/Se/Te, which are interesting for studying the effects of small perturbations on protein structure and function,[62] as well as bioorthogonal reactive groups useful for post-translational modifications (Figure 2). It would be therefore interesting to site-specifically incorporate them into target proteins. As mentioned earlier, this could be achieved via the SCS method. However, this method appears to be limited to bulky (aromatic or long aliphatic) amino acids, which are the substrates of the commonly used o-pairs PylRS and *Mj*TyRS, and not available for Met analogues. The only known example of this kind is based on sense codon reassignment. Site-specific labelling of proteins at the N-terminus with azidonorleucine (Anl) was achieved in mammalian cells expressing NNL-*Ec*MetRS, since the bacterial synthetase recognizes the mammalian initiator tRNAi^{Met} but does not cross-react with the endogenous elongator tRNA^{Met}.[63]

Generally, the N-terminus appears to be a favorable target for site-specific functionalization being predominantly exposed on the protein surface and because protein initiation, in theory, could be reprogrammed separately from elongation (Figure 1). Several attempts to recode the N-terminus of proteins with ncAAs have been pursued, mostly using *in vitro* approaches.[64–68]

Interestingly, several Met analogues are known to work in protein initiation *in vivo*, indicating that they are formylated by methionyl-tRNA^{fMet} formyltransferase (FMT) and recognized by the initiation factors (IFs).[69] In addition, their N-terminal processing is attenuated[69] and can be regulated by changing the identity of the second amino acid in the protein sequence.[70,71] Even bacteria devoid of N-terminal protein formylation were developed by Mutzel and Marliere.[72] In the evolved *E. coli* strain, which has both transformylase (*fmt*) and deformylase (*fms*) deleted, protein synthesis can be initiated by unmodified methionine instead of N-formylmethionine. Based on the observation that *Pseudomonas aeruginosa* with inactivated formylase is still viable, Newton *et al*. concluded that formylation is not essential for initiation of protein synthesis in all eubacteria.[73]

In this work, we attempted the site-specific incorporation of azidohomoalanine (Aha) at the N-terminus, while simultaneously replacing the internal Met residues with its analogue ethionine (Eth) during *in vivo* protein translation. We also report the development of a genetic selection system based on the suppression of an amber stop codon instead of a Met residue in an important glycolytic gene, *pfkA*, which restores enzyme functionality and normal cellular growth.
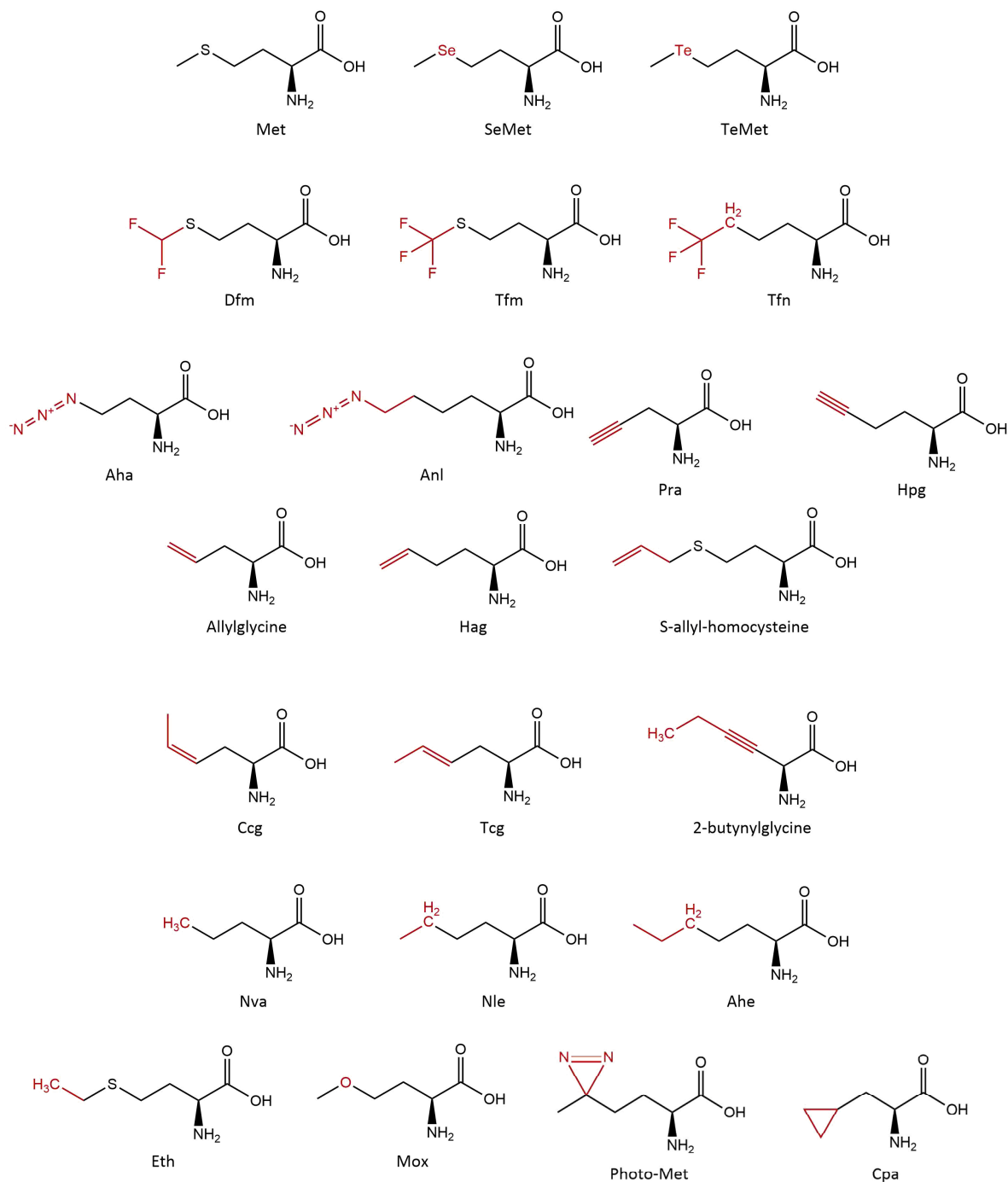
Figure 2. Met analogues incorporated residue-specifically for protein engineering. The non-canonical amino acids contain single atom or chemical group substitutions (in red). Methionine (Met); Selenomethionine (SeMet); Telluriomethionine (TeMet); Difluoromethionine (Dfm); 5,5,5-Trifluoromethionine (Tfm); 6,6,6,-Trifluoronorleucine (Tfn); Azidohomoalanine (Aha); Azidonorleucine (Anl); Propargylglycine (Pra); Homopropargylglycine (Hpg); Allylglycine; Homoallylglycine (Hag); *S*-allyl-homocysteine; *cis*-crotylglycine (Ctg); *trans*-crotylglycine (Tgc); 2-butynylglycine; Norvaline (Nva); Norleucine (Nle); 2-aminoheptanoic acid (Ahe); Ethionine (Eth); Methoxinine (Mox); Photomethionine; β-Cyclopropylalanine (Cpa).

RESULTS AND DISCUSSION

Separating AUG and Internal AUG Codons by Two Different tRNAs

To decode differently the initial and internal AUG codons, their co-respective tRNAs^{Met} should be loaded with two different ncAAs. This can be achieved using two MetRSs, one endogenous and the other exogenous, with different substrate specificities. Most importantly, the two systems should not cross-react, *i.e.*, they need to be mutually orthogonal. Heterologous MetRS should aminoacylate its cognate tRNA^{Met} with one specific ncAA, while not interacting with endogenous *Ec*tRNA^{Met}. In addition, the heterologous tRNA^{Met} should not be recognized by endogenous *Ec*MetRS.

Cross-aminoacylation experiments with tRNAs^{Met} and MetRSs from different organisms suggest the existence of two types of MetRSs (Figure 3).[74] The "type A" from bacteria and organelles aminoacylates bacterial or organellar initiator tRNAi^{fMet} and elongator tRNA^{Met}. Moreover, it also recognizes eukaryotic and archaeal initiator tRNAi^{Met}.[75,76] In contrast, the "type B" includes archaeal or eukaryotic cytoplasmatic MetRSs that aminoacylate archaeal or eukaryotic initiator tRNAi^{Met} and elongator tRNA^{Met}. However, these MetRSs aminoacylate the bacterial or organellar tRNAs^{Met} counterparts with low efficiency.[74] Consequently, it should be feasible to use a type B MetRS/tRNA^{Met} pair to reassign AUG codons to a
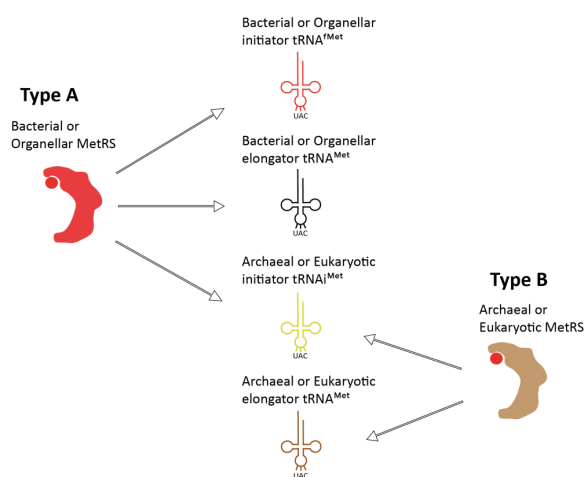
Figure 3. Two types of MetRS/tRNA^{Met} pairs. Type A includes the MetRSs from eubacteria and eukaryotic organelles, which preferentially aminoacylate their initiator tRNAi^{fMet} and elongator tRNA^{Met} counterparts, as well as archaeal or eukaryotic cytoplasmic initiator (tRNAi^{Met}). Type B includes archaeal or eukaryotic cytoplasmic MetRS, which exclusively aminoacylate their cognate elongator tRNA^{Met} and initiator tRNAi^{Met}.

Met analogue exclusively or at least preferentially during elongation in *E. coli*.

Previous experiments in our group identified a putative orthogonal MetRS / tRNA^{Met} pair from the archaeon *Sulfolobus acidocaldarius* (*Sa*).[77] In particular, *in vitro* experiments showed that *Sa*MetRS aminoacylates *Ec*tRNAi^{fMet} and elongator *Ec*tRNA^{Met} with low efficiency and that, among several Met analogues, it has a marked substrate preference towards Eth.[77] In comparison, endogenous *Ec*MetRS activates this analogue less efficiently. Eth has an ethyl group attached to the sulfur atom rather than a methyl group in Met (Figure 2). Given this slight difference, its global incorporation in target proteins is expected to have a minor impact on their structures, while it can be used to study overall effects on their biochemical properties.[42] On the other end, when Eth is incorporated in the whole proteome, it inhibits growth in many microorganisms including *E. coli*.[78]

It was found that *Ec*MetRS activates Aha with higher efficiency than Eth, while *Sa*MetRS activates this analogue 4.4 times lower than Eth.[77] *In vivo* experiments with the model protein barstar[10] containing two Met residues (b*2M), one at the N-terminus and the other at the internal position E47M, confirmed the overall preference of the endogenous system for Aha over Eth.[77] In particular, Aha was incorporated 1.8 times more at the N-terminus than at the internal position, suggesting that this ncAA works better in translation initiation than in elongation.[77] Contrarily to Eth, Aha is not toxic in mammalian cell lines, neither in whole organisms for short incubation times, and it has been used to selectively label the newly synthesized proteome via BONCAT.[57,79]

These results provided a starting point for the differential reassignment of the N-terminus to Aha using the natural substrate preference of *Ec*MetRS, while globally reassigning the internal AUG codons with Eth using the putative orthogonal *Sa*MetRS/*Sa*tRNA^{Met} pair.

Removal of metT/metU from *E. coli* and Rescue With the Heterologous SaMetRS/tRNA Pair

As discussed above, sense codon reassignment entails competition with the endogenous tRNAs. Considering that tRNA abundance correlates with codon usage[80] and that the AUG codon is highly used in *E. coli* (frequency: ~25 codons per thousand, of which ~22 at internal positions),[81,82] cellular concentrations of elongator *Ec*tRNA^{Met} might be too high to be outcompeted by an exogenous tRNA^{Met}. Therefore, we decided to eliminate this gene, which is present in the *E. coli* genome in two versions, *metT* and *metU*. Since the genomic deletion of both copies is lethal for the cells, and because the internal AUG codons become unassigned, the knockout strain was

supplemented with a rescue vector from the pMEc series[83] containing an elongator tRNA^Met (Figure S1).

Two putative elongator tRNA^Met were identified in *S. acidocaldarius*:[84] *Sa*tRNA(8) and *Sa*tRNA(16). They both differ from each other and from elongator *Ec*tRNA^Met at several positions (Figure 4). For example, they have a different base composition at conserved positions G^2 and C^3 in the acceptor stem which are determinants for aminoacylation by *Ec*MetRS.[85] A main difference is the size of the D-loop, which is composed of 10 bases in *Sa*tRNA^Met(16) compared to that of *Sa*tRNA^Met(8) and elongator *Ec*tRNA^Met which is made of 9 bases (Figure 4). It was hypothesized that this difference in size might be a determinant for orthogonality in *E. coli*,[77] similar to the way in which enlarging the anticodon loop renders elongator *Ec*tRNA^Met a good substrate for mammalian MetRS, while decreasing its interaction with bacterial MetRS.[74] Changing its size, however, does not seem to affect the rate of aminoacylation.[85] Nevertheless, in the present study we decided to test *Sa*tRNA^Met(16). A *metT/ metU*-deficient *E. coli* strain was successfully generated (see Figure S2). However, by using the pMEc vector containing different combinations of *Ec*MetRS / *Sa*MetRS and elongator *Ec*tRNA^Met / *Sa*tRNA^Met (Figure S1) we found that our system is not fully orthogonal *in vivo*: heterologous *Sa*tRNA^Met is still recognized and charged by endogenous *Ec*MetRS.

## Simultaneous Reassignment of the Initial and Internal AUG Codons to L-Azidohomoalanine and L-Ethionine

In spite of the fact that the *Sa*MetRS/*Sa*tRNA^Met pair was not fully orthogonal in *E. coli*, we set out to test how far previously observed preferential Eth-incorporation[77] can be pushed in the newly designed *metT*/*metU E. coli* strain.

Such experiment combines sense codon reassignment with the SPI-method in an *E. coli* strain auxotrophic for Met, B834(DE3), deficient in the elongator tRNA^Met (Figure S3).

Barstar (b*) congeners – proteins with the same gene sequence as wild-type protein but containing a fraction of synthetic amino acids – were expressed in the *metT /metU*-deficient B834(DE3) strain, as described in the SI. The purified congeners were analyzed by ESI mass spectrometry, in which three different species were identified: b*2Aha, b*1Aha-1Eth, and b*2Eth. The second congener can be in two forms: b*1Aha-47Eth and b*1Eth-47Aha. The distribution of each species was calculated from the peak intensity in the MS spectra (Table 1).

Notably, the relative abundances of the three b* congeners expressed in the *metT / metU*-deficient B834(DE3) strain and in the B834(DE3) wild-type strain differ from each other. Elimination of endogenous elongator *Ec*tRNA^Met led to an overall shift towards incorporation of Eth, as indicated by an increase in the abundance of b*1Aha-1Eth (+8 % and +22 %) and b*2Eth (+9 % and +19 %) in presence of *Sa*tRNA^Met and *Ec*tRNA^Met, respectively (Table 1). The observed variations can probably be related to the altered intracellular levels of elongator tRNA^Met. Analysis of congener b*1Aha-1Eth by N-terminal sequencing remains to be performed for determining the exact distribution of the two incorporated analogues at the N-terminus and at internal position 47.

Contrary to our expectations, the expression of the *Sa*MetRS/*Sa*tRNA^Met pair in the *metT/metU*-deficient strain did not alter significantly the incorporation efficiencies in favor of Eth. In fact, the distribution of Aha and Eth in b* was similar to that obtained in presence of *Ec*MetRS / *Ec*tRNA^Met. These results further confirm the fact that orthogonality is an essential prerequisite for efficient sense codon reassignment.
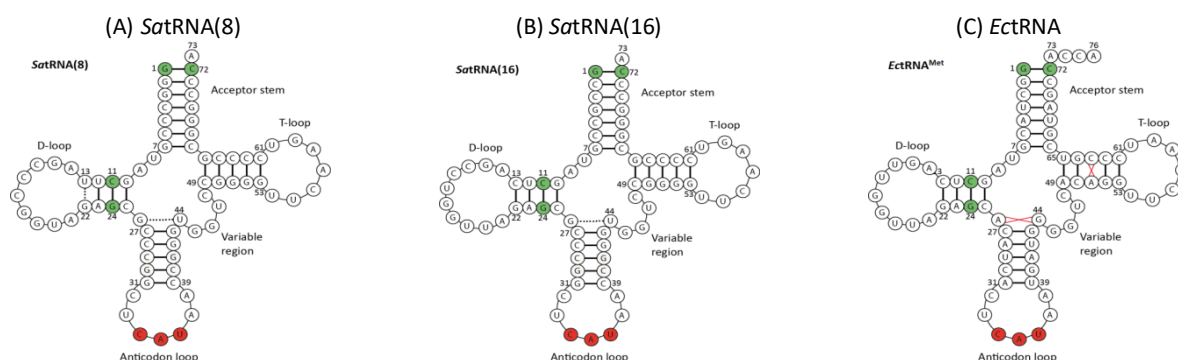
Figure 4. Structures of elongator tRNAs^Met. Secondary structure of the two putative *S. acidocaldarius* elongator tRNA^Met, (A) *tRNA8* and B) *tRNA16*,[84] and of C) *Ec*tRNA^Met (*metT*). The required identity elements for elongation (green) and aminoacylation (red) are shown. The dotted line indicates that a base modification is required for hydrogen bonding. In Archaea the CCA sequence at the 3′ terminal of pre-tRNA is added by a CCA-adding enzyme after gene transcription.

A Novel Genetic Selection System for AUG Reassignment *via* MmPylRS-Based Orthogonal Pairs

Due to the lack of orthogonality in the examined *Sa*MetRS/*Sa*tRNA^Met system, we turned our efforts to modify the *Methanosarcina mazei* pyrrolysyl-tRNA synthetase (PylRS) / tRNA^Pyl pair for use in the AUG decoding system. Its natural orthogonality in *E. coli* makes it a useful tool for both SCS and sense codon reassignment approaches.[86] Further-more, PylRS shows a low selectivity toward the anticodon of tRNA^Pyl, which could be mutated to recognize the Met AUG codon, and a natural broad substrate specificity, which could be evolved to accept Met analogues.
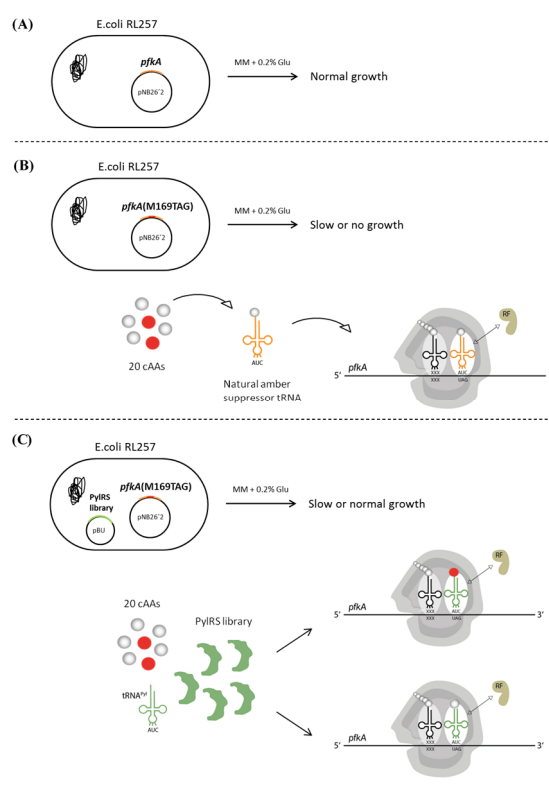
In order to identify a PylRS mutant able to charge Met, we devised a selection system based on the functionality of a glycolytic enzyme phosphofructokinase-1 (PFK-1). This enzyme catalyzes the first committed step of glycolysis, the conversion of fructose 6-phosphate to fructose 1,6-bisphosphate. Its activity, together with that of phosphofructokinase-2 (PFK-2), is essential for bacterial growth in M9 minimal medium.[87] Importantly, it contains a conserved Met residue at position 169 crucial for substrate binding. It was shown, in fact, that substitutions at this position with alanine, lysine and isoleucine drastically decrease its catalytic efficiency.[88,89] Amber suppression at this position by an *Mm*PylRS mutant able to incorporate Met is expected to restore PFK-1 functionality and, consequently, normal bacterial growth (Figure 5C).

Based on our selection experiments (depicted in Figure 5 and described in detail in the SI) we conclude that our system is reliable. Namely, the incorporation of Met at position 169 in PFK-1 can be used as selection marker, since it confers an advantage in cellular growth compared to incorporation of any of the other 19 canonical amino acids (Figure S7). However, a high background was found on the negative control plate, as several colonies grew after 1.52 days. This indicates that the system needs to be rendered more stringent in order to minimize false positives which

Table 1. Abundance of the three b* congeners obtained using pMEc1 and pMEc3 plasmids in B834(DE3)_Δ*metT*::frt_Δ*metU*::*npt* and B834(DE3) WT. The combined intensities of the corresponding peaks sum up to 100 %, from which each species percentage was calculated and annotated. Mean and standard deviation (SD) of each congener percentage from triplicate experiments with the double knockout mutant are shown. Heterologous expression of *Sa*MetRS/*Sa*tRNA^Met (highlighted in brown) did not alter significantly the distribution of the 2 Met analogues compared to the negative control represented by homologously expressed *Ec*MetRS/*Ec*tRNA^Met (highlighted in red). Natural incorporation efficiencies measured previously in the Met auxotrophic *E. coli* strain CAG18491 are also reported.[77]

Configuration, strain	Congeners		
	B*2Aha	B*1Aha-1Eth	B*2Eth
*Sa*MetRS/*Sa*tRNA^Met B834(DE3)_Δ*metT*_Δ*metU*	37 ± 6 %	45 ± 3 %	18 ± 3 %
*Ec*MetRS/*Ec*tRNA^Met B834(DE3)_Δ*metT*_Δ*metU*	29 ± 2 %	48 ± 1 %	23 ± 1 %
*Sa*MetRS/*Sa*tRNA^Met B834(DE3)	54 %	37 %	9 %
*Ec*MetRS:*Ec*tRNA^Met B834(DE3)	70 %	26 %	4 %
WT, B834(DE3)	57 %	32 %	11 %
WT, CAG18491	38 %	42 %	20 %

Figure 5. Selection system based on amber suppression for the identification of a PylRS mutant able to charge Met. (A) The *E. coli* RL257 strain expressing exogenous PFK-1 can grow on minimal medium plates; (B) When a TAG stop codon is inserted in *pfkA* at position 169 in place of a critical Met residue, cell growth is impaired. Slow growth might be restored when M169TAG is suppressed by one natural amber suppressor tRNA; (C) A PylRS library is subsequently transformed into these cells. A PylRS mutant able to suppress the M169TAG codon with Met (in red) would restore normal cell growth, whereas incorporation of any of the other 19 cAAs (in grey) would result in a slow-growth phenotype. MM = minimal medium.

may interfere with the subsequent selection. The high background might be caused by UAG read-through by one of the nonsense tRNA suppressors (tRNA^{Gln}, tRNA^{Leu}, tRNA^{Ser} and tRNA^{Tyr})[90] or near-cognate aminoacyl-tRNAs identified in *E. coli*.[91] Incorporation of one of these amino acids at position 169 might produce a full-length PFK-1 variant which is active enough to support bacterial growth, although impaired compared to cells expressing PFK-1 wild-type. To test this hypothesis, the 20 possible variants could be constructed. Nevertheless, a more probable reason for the high number of colonies on the negative control plate could be the presence of some residual LB medium after transformation. The nutritionally rich LB medium can restore normal growth because some of its components may allow the auxotrophic cells to enter glycolysis circumventing PFK-1.

To solve these problems, we have introduced a second stop codon at position M224TAG, which completely eliminated the background suppression. This position was chosen because it is not conserved and does not participate in the substrate or cofactor binding, being located rather in a loop on the protein surface. The applicability of the optimized selection system with two stop codons in identifying the desired PylRS mutant is currently being tested using different libraries constructed in our group.

CONCLUSIONS

This work describes the first attempts to reassign the sense codon AUG to two different Met analogues, Aha and Eth, in *E. coli*. Based on *in vitro* data,[77] we tried exploiting the natural substrate preference of *Ec*MetRS/*Ec*tRNAi^{fMet} pair to incorporate Aha at the N-terminus, while using an exogenous archaeal *Sa*MetRS/*Sa*tRNA^{Met} pair to insert Eth at internal AUG positions. However, it was not possible to unambiguously reassign the AUG codon in initiation and elongation *in vivo*, without strict orthogonality. Even though the elimination of the elongator *Ec*tRNA^{Met} altered the distribution of the two analogues in favor of Eth, it was not sufficient to achieve a significant AUG reassignment. Orthogonality and the differentiation of *in vitro* vs. *in vivo* data are important criteria for developing o-pairs.[92]

The orthogonality of the *Sulfolobus acidocaldarius*-based heterologous pair could be achieved by identifying and mutating positions in the *Sa*tRNA^{Met} sequence which are critical in its aminoacylation by *Ec*MetRS, as compared to *Ec*tRNA^{Met}.[85] Alternatively, the other putative elongator *Sa*tRNA^{Met} (*tRNA8*, Figure 4) could also be tested for its orthogonality *in vivo*. Furthermore, the substrate specificity of *Ec*MetRS towards Aha should be further evolved in order to achieve exclusive incorporation of this ncAA at the N-terminus. Interestingly, it is also possible to change the incorporation efficiencies at the N-terminus and internal

positions by tuning the concentration of the two Met analogues in the media.[77]

We also presented the first steps towards the design of a PylRS/tRNA^{Pyl} pair for use in the Met decoding system based on the suppression of an amber stop codon with Met in an important glycolytic gene, *pfkA*, which, in turn, restores enzyme functionality and normal cellular growth. Importantly, any orthogonal aaRS/tRNA pair that accepts Met could be tested in the *E. coli* strain lacking *Ec*tRNA^{Met}. We envision a PylRS library specifically customized for Met recognition that could be screened and identified with the here developed selection system. Once such synthetase is identified, its substrate specificity could be further evolved towards interesting Met analogues, such as Aha, Anl, Hpg, Nle and Eth and many others.

On the way to the evolution of synthetic cells with alternative biochemical building blocks, we also envision a strategy to exclusively reassign the internal or initial AUG codons but not both because Met is the major methyl group donor in cellular metabolism. Thus, its elimination from the genetic code would impair cellular growth, unless other sources of *S*-adenosylmethionine could be provided by advanced metabolic engineering. The most reasonable strategy would be to use properly configured *metT*/*metU*-deficient *E. coli* MG1655 strain which might be slowly adapted to use a different amino acid in place of Met using an evolution strategy recently reported by our group.[33,34] This configuration would also require the availability of mutually orthogonal MetRSs capable to generate the context-dependent AUG-codon reassignments: one ncAA (or Met itself) can be encoded site-specifically at the N-terminus and the other ncAAs (or Met itself) can be inserted globally at the internal AUG positions.

Supplementary Information. Supporting information to the paper is enclosed to the electronic version of the article at: http://dx.doi.org/10.5562/cca2916.

REFERENCES

[1] N. Budisa, B. Steipe, P. Demange, C. Eckerskorn, J. Kellermann, R. Huber, *Eur. J. Biochem.* **1995**, *230*, 788.

[2] J. A. Johnson, Y. Y. Lu, J.A. Van Deventer, D.A. Tirrell, *Curr. Opin. Chem. Biol.* **2010**, *14*, 774.

DOI: 10.5562/cca2915

Croat. Chem. Acta 2016, *89*(2), 243–253

[3]   C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413.

[4]   N. Budisa, *Angew. Chem. Int. Ed.* **2004**, *43*, 6426.

[5]   Y. Ravikumar, S. P. Nadarajan, T. Hyeon Yoo, C.-S. Lee, H. Yun, *Trends Biotechnol.* **2015**, *33*, 462.

[6]   Y. Ravikumar, S. P. Nadarajan, T. Hyeon Yoo, C.-S. Lee, H. Yun, *Biotechnol. J.* **2015**, *10*, 1862.

[7]   S. Zheng, I. Kwon, *Biotechnol. J.* **2012**, *7*, 47.

[8]   L. Merkel, M. Schauer, G. Antranikian, N. Budisa, *ChemBioChem* **2010**, *11*, 1505.

[9]   J. C. Anderson, N. Wu, S. W. Santoro, V. Lakshman, D. S. King, P. G. Schultz, *Proc. Natl. Acad. Sci.* **2004**, *101*, 7566.

[10] S. Lepthien, L. Merkel, N. Budisa, *Angew. Chem.* **2010**, *49*, 5446.

[11] S. Lepthien, M. G. Hoesl, L. Merkel, N. Budisa, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 16095.

[12] V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chem.* **2002**, *41*, 2596.

[13] C. Renner, S. Alefelder, J. H. Bae, N. Budisa, R. Huber, L. Moroder, *Angew. Chem.* **2001**, *40*, 923.

[14] W. Wan, Y. Huang, Z. Wang, W. K. Russell, P.-J. Pai, D. H. Russell, W. R. Liu, *Angew. Chem. Int. Ed.* **2010**, *49*, 3211.

[15] H. Neumann, K. Wang, L. Davis, M. Garcia-Alai, J. W. Chin, *Nature* **2010**, *464*, 441.

[16] K. Wang, A. Sachdeva, D. J. Cox, N. W. Wilf, K. Lang, S. Wallace, R. A. Mehl, *et al.*, *Nat. Chem.* **2014**, *6*, 393.

[17] A. Sachdeva, K. Wang, T. Elliott, J. W. Chin, *J. Am. Chem. Soc.* **2014**, *136*, 7785.

[18] M. G. Hoesl, N. Budisa, *ChemBioChem* **2011**, *12*, 552.

[19] A. Umeda, G. N. Thibodeaux, J. Zhu, Y. Lee, Z. J. Zhang, *ChemBioChem* **2009**, *10*, 1302.

[20] M. Amiram, A. D. Haimovich, C. Fan, Y.-S. Wang, H.-R. Aerni, I. Ntai, D. W. Moonan, *et al.*, *Nat. Biotechnol.* **2015**, *33*, 1272.

[21] T. Mukai, A. Yamaguchi, K. Ohtake, M. Takahashi, A. Hayashi, F. Iraha, S. Kira, *et al.*, *Nucleic Acids Res.* **2015**, *43*, 8111.

[22] N. Bohlke, N. Budisa, *FEMS Microbiol. Lett.* **2014**, *351*, 133.

[23] I. Kwon, K. Kirshenbaum, D. A. Tirrell, *J. Am. Chem. Soc.* **2003**, *125*, 7512.

[24] M. J. Bröcker, J. M. L. Ho, G. M. Church, D. Söll, P. O'Donoghue, *Angew. Chem. Int. Ed.* **2014**, *53*, 319.

[25] J. M. Ho, N. M. Reynolds, K. Rivera, M. Connolly, L.-T. Guo, J. Ling, D. J. Pappin, *et al.*, *ACS Synth. Biol.* **2016**, *5*, 163.

[26] I. Kwon, E. S. Choi, *Plos One* **2016**, *11*, e0152826.

[27] Codon Usage Database, http://www.kazusa.or.jp/codon/

[28] J. P. Waller, *J. Mol. Biol.* **1963**, *7*, 483.

[29] C. Kaleta, S. Schauble, U. Rinas, S. Schuster, *Biotechnol. J.* **2013**, *8*, 1105.

[30] I. G. Old, S. E. Phillips, P. G. Stockley, I. Saint Girons, *Prog. Biophys. Mol. Biol.* **1991**, *56*, 145.

[31] S. Luo, R. L. Levine, *FASEB J.* **2009**, *23*, 464.

[32] T. E. Jones, R. W. Alexander, T. Pan, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6933.

[33] M. G. Hoesl, S. Oehm, P. Durkin, E. Darmon, L. Peil, H.-R. Aerni, J. Rappsilber, *et al.*, *Angew. Chem. Int. Ed.* **2015**, *54*, 10030.

[34] C. G. Acevedo-Rocha, D. Schulze-Makuch, *ChemBioChem* **2015**, *16*, 2137.

[35] N. Budisa, *Wiley-VCH Verlag GmbH & Co.* **2005**, p. 312.

[36] W. A. Hendrickson, J. R. Horton, D. M. LeMaster, *EMBO J.* **1990**, *9*, 1665.

[37] J. O. Boles, K. Lewinski, M. Kunkle, J. D. Odom, B. Dunlap, L. Lebioda, M. Hatada, *Nat. Struct. Biol.* **1994**, *1*, 283.

[38] T. H. Yoo, D. A. Tirrell, *Angew. Chem.* **2007**, *46*, 5340.

[39] P. Cleve, V. Robinson, H. S. Duewel, J. F. Honek, *J. Am. Chem. Soc.* **1999**, *121*, 8475.

[40] N. Budisa, O. Pipitone, I. Siwanowicz, M. Rubini, P. P. Pal, T. A. Holak, M. L. Gelmi, *Chem. Biodivers.* **2004**, *1*, 1465.

[41] M. J. Wilson, D. L. Hatfield, L. A. Poirier, *FEBS Lett.* **1981**, *128*, 157.

[42] T. Yuan, H. J. Vogel, *Protein Sci.* **2008**, *8*, 113.

[43] M. G. Hoesl, C. G. Acevedo-Rocha, S. Nehring, M. Royter, C. Wolschner, B. Wiltschi, N. Budisa, *et al.*, *ChemCatChem* **2011**, *3*, 213.

[44] C. G. Acevedo-Rocha, M. G. Hoesl, S. Nehring, M. Royter, C. Wolschner, B. Wiltschi, G. Antranikian, *et al.*, *Catal. Sci. Tech.* **2013**, *3*, 1198.

[45] C. Wolschner, A. Giese, H. A. Kretzschmar, R. Huber, L. Moroder, N. Budisa, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 7756.

[46] M. Suchanek, A. Radzikowska, C. Thiele, *Nature Methods* **2005**, 261.

[47] J. C. van Hest, D. A. Tirrell, *FEBS Lett* **1998**, *428*, 68.

[48] K. L. Kiick, J. C. van Hest, D. A. Tirrell, *Angew. Chem.* **2000**, *39*, 2148.

[49] K. L. Kiick, R. Weberskirch, D. A. Tirrell, *FEBS Lett.* **2001**, *502*, 25.

[50] L. Alfonta, Z. Zhang, S. Uryu, J. A. Loo, P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125*, 14662.

[51] F. Truong, *California Institute of Technology, PhD thesis* **2013**.

[52] K. L. Kiick, E. Saxon, D. A. Tirrell, C. R. Bertozzi, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 19.

[53] J. C. M. van Hest, K. L. Kiick, D. A. Tirrell, *J. Am. Chem. Soc.* **2000**, *122*, 1282.

[54] I. C. Tanrikulu, E. Schmitt, Y. Mechulam, W. A. Goddard, D. A. Tirrell, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 15285.

[55] L. Merkel, H. S. G. Beckmann, V. Wittmann, N. Budisa, *ChemBioChem* **2008**, *9*, 1220.

[56] A. J. Link, D. A. Tirrell, *J. Am. Chem. Soc.* **2003**, *125*, 11164.

[57] D. C. Dieterich, A. J. Link, J. Graumann, D. A. Tirrell, and EM Schuman, *Proc. Natl. Acad. Sci.* **2006**, *103*, 9482.

[58] F. Truong, T. H. Yoo, T. J. Lampo, D. A. Tirrell, *J. Am. Chem. Soc.* **2012**, *134*, 8551.

[59] J. T. Ngo, J. A. Champion, A. Mahdavi, I. C. Tanrikulu, K. E. Beatty, R. E. Connor, T. H. Yoo, *et al.*, *Nat. Chem. Biol.* **2009**, *5*, 715.

[60] D. B. Cowie, G. N. Cohen, *Biochim. Biophys. Acta* **1957**, *26*, 252.

[61] N. Budisa, B. Steipe, P. Demange, C. Eckerskorn, J. Kellermann, R. Huber, *Eur. J. Biochem.* **1995**, *230*, 788.

[62] N. Budisa, R. Huber, R. Golbik, C. Minks, E. Weyher, L. Moroder, *Eur. J. Biochem.* **1998**, *253*, 1.

[63] J. T. Ngo, E. M. Schuman, D. A. Tirrell, *Proc. Natl. Acad. Sci.* **2013**, *110*, 4992.

[64] Y. Goto, A. Ohta, Y. Sako, Y. Yamagishi, H. Murakami, H. Suga, *ACS Chem. Biol.* **2008**, *3*, 120.

[65] Y. Goto, H. Murakami, H. Suga, *RNA* **2008**, *14*, 1390.

[66] Y. Goto, H. Suga, *J. Am. Chem. Soc.* **2009**, *131*, 5040.

[67] V. Ramachandiran, C. Willms, G. Kramer, B. Hardesty, *J. Biol. Chem.* **2000**, *275*, 1781.

[68] N. Muranaka, M. Miura, H. Taira, T. Hohsaka, *ChemBioChem* **2007**, *8*, 1650.

[69] L. Merkel, Y. Cheburkin, B. Wiltschi, N. Budisa, *ChemBioChem* **2007**, *8*, 2227.

[70] A. Wang, N. Winblade Nairn, R. S. Johnson, D. A. Tirrell, K. Grabstein, *ChemBioChem* **2008**, *9*, 324.

[71] C. G. Acevedo-Rocha, A. S. Geiermann, N. Budisa, L. Merkel, *Mol. Biosyst.* **2012**, *8*, 2719.

[72] P. Marliere, R. Mutzel, D. Mazel, *Google Patents* **2005**.

[73] D. T. Newton, C. Creuzenet, D. Mangroo, *J. Biol. Chem.* **1999**, *274*, 22143.

[74] T. Meinnel, Y. Mechulam, G. Fayat, S. Blanquet, *Nucleic Acids Res.* **1992**, *20*, 4741.

[75] P. Gillemaut, J. H. Weil, *Biochim. Biophys. Acta* **1975**, *407*, 240.

[76] S. Blanquet, G. Petrissant, J. P. Waller, *Eur. J. Biochem.* **1973**, *36*, 227.

[77] C. G. Acevedo-Rocha, *Technische Universität München,* PhD thesis, **2010**.

[78] M. H. Richmond, *Bacteriol. Rev.* **1962**, *26*, 398.

[79] S. Calve, A. J. Witten, A. R. Ocken, T. L. Kinzer-Ursem, *Sci. Rep.* **2016**, *6*, 32377.

[80] T. Ikemura, *J. Mol. Biol.* **1981**, *151*, 389.

[81] M. J. Lajoie, S. Kosuri, J. A. Mosberg, C. J. Gregg, D. Zhang, G. M. Church, *Science* **2013**, *342*, 361.

[82] M. Riley, *Nucleic Acids Res.* **2006**, *34*, 1.

[83] M. Hösl, *TU München,* PhD thesis, **2011**.

[84] F. Juhling, M. Morl, R. K. Hartmann, M. Sprinzl, P. F. Stadler, J. Putz, *Nucleic Acids Res.* **2009**, *37*, D159.

[85] T. Meinnel, Y. Mechulam, C. Lazennec, S. Blanquet, G. Fayat, *J. Mol. Biol.* **1993**, *229*, 26.

[86] T. Fekner, M. K. Chan, *Curr. Opin. Chem. Biol.* **2001**, *15*, 387.

[87] M. R. Lovingshimer, D. Siegele, G. D. Reinhart, *Protein Expr. Purif.* **2006**, *46*, 475.

[88] Y. Shirakihara, P. R. Evans, *J. Mol. Biol.* **1988**, *204*, 973.

[89] X. Wang, Z. Deng, R. G. Kemp, *Biochem. Biophys. Res. Comm.* **1998**, *250*, 466.

[90] http//www.sci.sdsu.edu/~smaloy/MicrobialGenetics/topics/rev-sup/nonsense-suppressors.html

[91] H. R. Aerni, M. A. Shifman, S. Rogulina, P. O'Donoghue, J. Rinehart, *Nucleic Acids Res.* **2014**, *43*, e8.

[92] D. Jones, S. Nehring, N. Budisa, B. Wiltschi, *PLoS ONE* **2012**, *7*, e31992.

DOI: 10.5562/cca2915

Croat. Chem. Acta **2016**, *89*(2), 243–253

Supplementary Information

Towards reassignment of the methionine codon AUG to two different noncanonical amino acids in bacterial translation

Alessandro De Simone[1], Carlos G. Acevedo-Rocha[2], Michael Georg Hoesl[3], Nediljko Budisa[1]

[1]Department of Chemistry, Biocatalysis Group, Technical University Berlin/Berlin Institute of Technology, Müller-Breslau-Str. 10, Berlin 10623, Germany.
[2]Biosyntia ApS, 2100, Copenhagen, Denmark.
[3]Clariant Produkte (Deutschland) GmbH, Semmelweisstraße 1, 82152 Planegg

Experimental Details

Chromosomal deletion of *metT* and *metU* in *E. coli*

Four rescue vectors and two controls were constructed using the pMEc plasmid as backbone, which is a shuttle vector for *E. coli* (p15A, CmR) and yeast (2u Ori and URA3)[1] to ease *in vivo* assembly (Figure S1). Plasmid pMEc1 was used to study whether the *Sa*MetRS/*Sa*tRNA^{Met} pair can rescue the *metT/metU*-deficient *E. coli* strain and reassign the internal AUG codons to Eth. Plasmids pMEc2 (*Sa*MetRS/*Ec*tRNA^{Met}) and pMEc4 (*Ec*MetRS/*Sa*tRNA^{Met}) were used to study if the endogenous and exogenous system cross-react with each other, whereas plasmids pMEc3 was used as positive control (*Ec*MetRS/*Ec*tRNA^{Met}). Finally, two other vectors, pMEc5 (*Sa*MetRS) and pMEc6 (*Ec*MetRS), carrying each synthetase gene but lacking cognate tRNA^{Met}, have served as negative controls. Expression of *Sa*MetRS and *Ec*MetRS was under control of the inducible propionate promoter, which should allow good overall synthetase production without impairing the target protein's yield [2]. The gene for tRNA^{Met} was under control of the constitutive promoter *proK*. The pMEc plasmid contains the low-copy origin of replication p15A.

The *E. coli* strain MG1655 was transformed with each of the six pMEc vectors and knockout of *metT* and *metU* was carried out using the λ-red recombination system [3]. Replacement of genomic *metT* and *metU* with *npt* (neomycin phosphotransferase) was deemed successful by the appearance of colonies on LB plus kanamycin (Kan^R) selection plates. As expected, MG1655 cells not rescued by an exogenous elongator *Ec*tRNA^{Met} (pMEc5 or pMEc6) did not grow on LB Kan^R plates. Colonies containing one of the four pMEc rescue plasmids were analyzed by PCR with different sequencing primers (Table S1). Successful recombination occurred in all the analyzed clones generating the MG1655_Δ*metT*::frt_Δ*metU*::*npt* strain (Figure S2). Importantly, cells transformed with pMEc4 survived deletion of *metT/metU*, indicating that heterologous *Sa*tRNA^{Met} can be recognized and charged by endogenous *Ec*MetRS.

Subsequently, the *npt* cassette was moved from MG1655 to the Met auxotrophic strain B834(DE3) by phage P1 transduction [4]. The successful integration of *npt* at the *metT* and *metU* loci was confirmed by PCR (Figure S3). The generated B834(DE3)_Δ*metT*::frt_Δ*metU*::*npt* strain was rescued using the plasmid for AUG codon reassignment, pMEc1, and the control plasmid pMEc3.
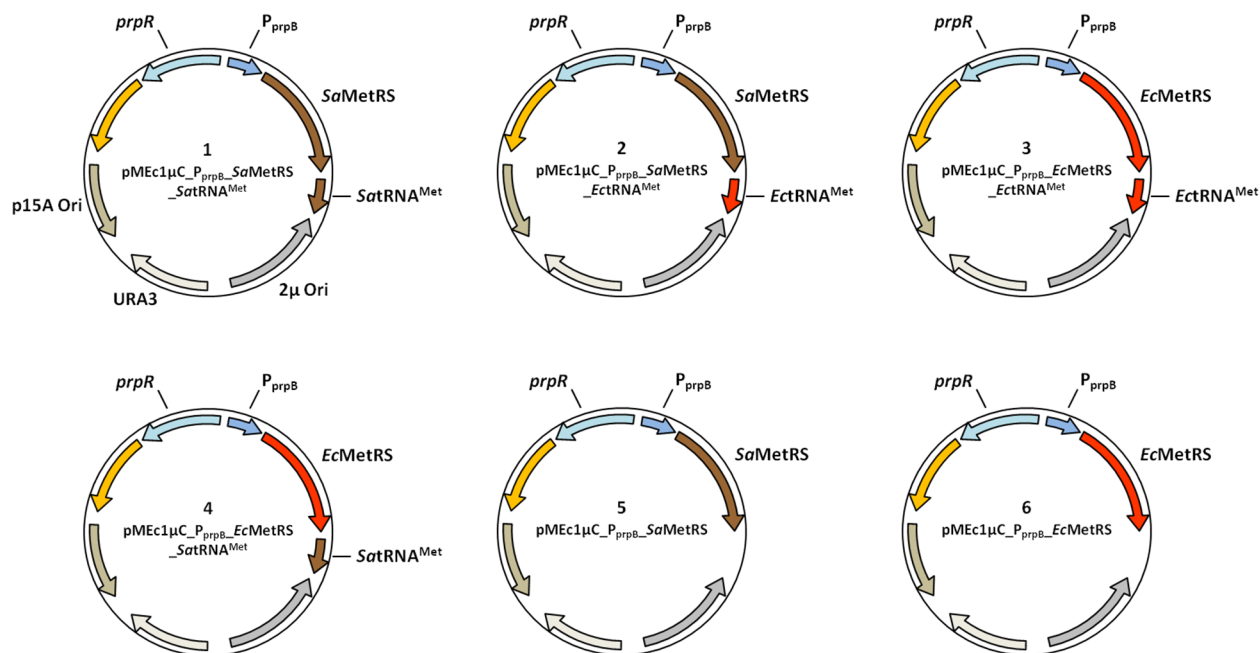
Figure S1. The six pMEc plasmids generated for AUG codon reassignment. Plasmid pMEc1 contains the putative orthogonal pair *Sa*MetRS/*Sa*tRNA[Met] used to direct the incorporation of Eth preferentially at the internal AUG codons. Plasmids pMEc2 and pMEc4 were used to test the orthogonality, respectively, of *Sa*MetRS and elongator *Sa*tRNA[Met] *in vivo*. Plasmid pMEc3 was used as negative control in comparison with pMEc1. Plasmids pMEc5 and pMEc6 were used to test the effect of MetRS overexpression. Cm[R] = chloramphenicol resistance. URA3 = uracil auxotrophic marker. Origins of replication 2μ and p15A are used for plasmid propagation in yeast and *E. coli*, respectively.

Table S1. Sequences of confirmation primers used in colony PCR.

Primer	Specification	Sequence (5′→3′)
PB327 (rev)	Primes 27 bp downstream of *metT*	GAACCTAAATCTGGTGCGTCTACC
PB328 (fwd)	Primes within *metT*/*metU*	CACATCACTCATAATGATGGGG
PB329 (fwd)	Primes within *npt*	GGCTGCATACGCTTGATCC
PB662 (rev)	Primes 371 bp downstream of *metU*	AGCCTTACAACATTGCCGG
PB698 (rev)	Primes 186 bp downstream of *metT*	ACTCGAAGAAGATGGCTGG
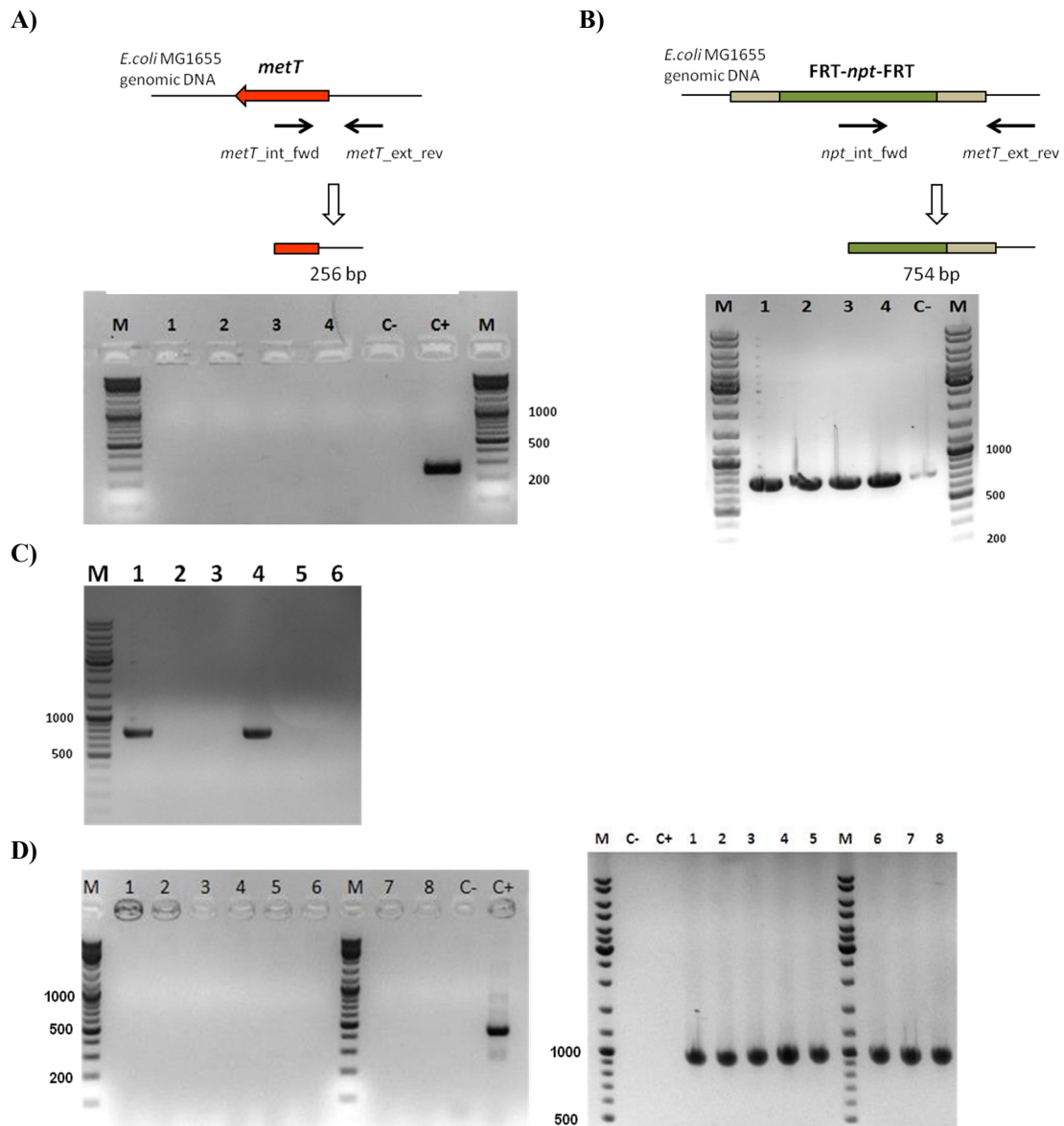
Figure S2. Verification of *metT* and *metU* knockout in MG1655. Water was used as a negative control (C-), while *E. coli* MG1655 wild-type represented the positive control (C+). **A)** Knockout of the *metT* gene in four clones (1-4) was confirmed by PCR using a forward and reverse primer that respectively amplify inside and downstream of the *metT* gene locus (PB328 + PB698, PCR product = 256 bp). **B)** Integration of the Kan^R cassette (*npt*) in place of *metT* in the same clones was verified by PCR using a primer for amplifying *npt* gene and an external primer targeting downstream of *metT* gene locus (PB327 + PB329, PCR product = 754 bp). **C)** Elimination of the Kan^R cassette was confirmed by PCR using primers PB327 + PB329. Lane 1 and 4 show the PCR of two clones before "plasmid curing", containing the Kan^R cassette (754 bp). Lane 2, 3, 5 and 6 show four clones "cured" by transformation with plasmid pCP20. The cured strain was termed MG1655_Δ*metT*::frt. **D)** Verification of knockout of *metU* and replacement with *npt* in MG1655_Δ*metT*::frt. Left: knockout of the *metU* gene in eight clones (2 clones each for pMEc1 - 4) was confirmed via PCR with oligos priming inside and downstream of the *metU* gene locus (PB328 + PB662, PCR product = 441 bp). Right: integration of the Kan^R cassette (*npt*) in place of *metU* in the same clones was verified with an internal primer directed inside the *npt* gene and an external one downstream of *metU* gene (PB329 + PB662, PCR product = 981 bp). The resulting strain is called MG1655_Δ*metT*::frt_Δ*metU*::*npt* (pMEc).
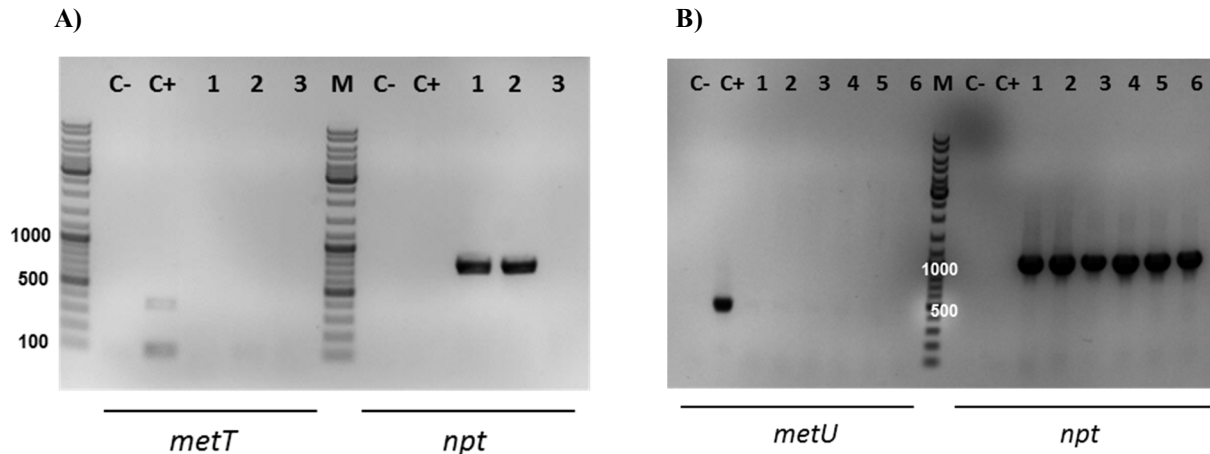
Figure S3. Verification of knockout of *metT* and *metU* in B834(DE3). Water was used as a negative control (C-), while *E. coli* MG1655 WT represented the positive control (C+). **A)** Knockout of the *metT* gene in B834(DE3) was confirmed via PCR using primers directed inside and downstream of the *metT* gene locus (PB327 + PB328, PCR product = 100 bp). Integration of the Kan^R cassette (*npt*) in place of *metT* in the same clones was verified using PCR with primers PB327 + PB329 (PCR product = 754 bp). Lanes 1-2 represent 2 knockout clones, while lane 3 shows the "cured" B834(DE3)_Δ*metT*::frt strain lacking the Kan^R cassette. **B)** Knockout of the *metU* gene in B834(DE3)_Δ*metT*::frt (pMEc) was also confirmed with PCR using the same set of primers described in Figure S2, D (*metU* PCR product = 441 bp; *npt* PCR product = 981 bp). Lanes 1-3 represent three B834(DE3)_Δ*metT*::frt_Δ*metU*::*npt* clones transformed with pMEc1. Lanes 4-6 show three B834(DE3)_Δ*metT*::frt_Δ*metU*::*npt* clones transformed with pMEc3.

Supplementation-based incorporation using the B834(DE3)_Δ*metT*::frt_Δ*metU*::*npt* strain

The generated B834(DE3)_Δ*metT*::frt_Δ*metU*::*npt* strain rescued with pMEc1 and control plasmid pMEc3 was further transformed with the expression vector pQE80L carrying the model protein barstar (b*2M). Expression of b*2M is controlled under the IPTG-inducible T5 promoter and contains an internal methionine at position 47.

Both clones were cultured in presence of limiting concentration of Met (45 µM) leading to its exhaustion in the medium during the fermentation, at which point, 0.5 mM L-Aha and 1 mM DL-Eth were added. Subsequently, expression of b*2M and MetRS was induced using 1 mM IPTG and 20 mM propionate, respectively. Expression of the two MetRSs was detected after enrichment of the synthetases from cell lysates using a Strep-Tactin® column (Figure S4). Barstar congeners were purified by ion exchange chromatography and analyzed by ESI mass spectrometry (Figure S5).
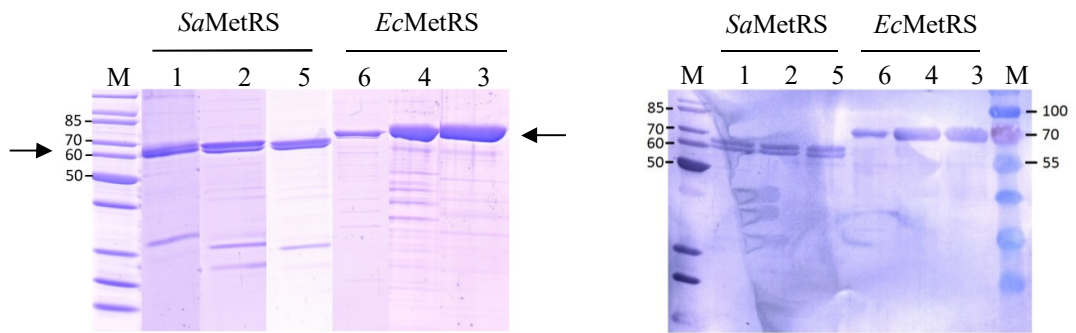
Figure S4. MetRSs purification. A) 12% SDS-PAGE and B) corresponding Western Blot of Strep-Tactin® purified *Ec*MetRS (77 kDa) and *Sa*MetRS (67 kDa) expressed under the propionate promoter. Numbers correspond to the pMEc plasmids described in Figure S1. The black arrow indicates the height were MetRSs are expected. M = protein marker.
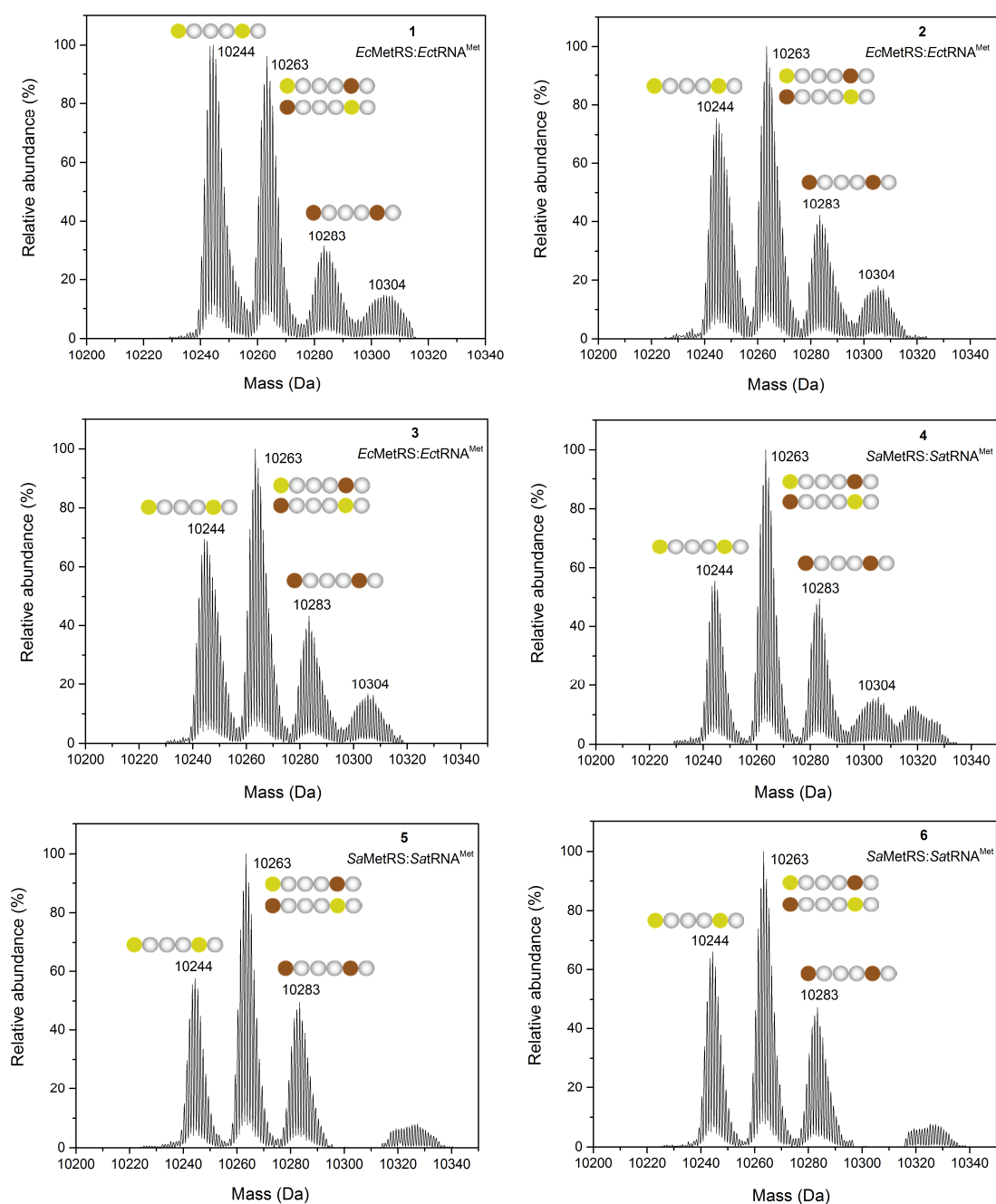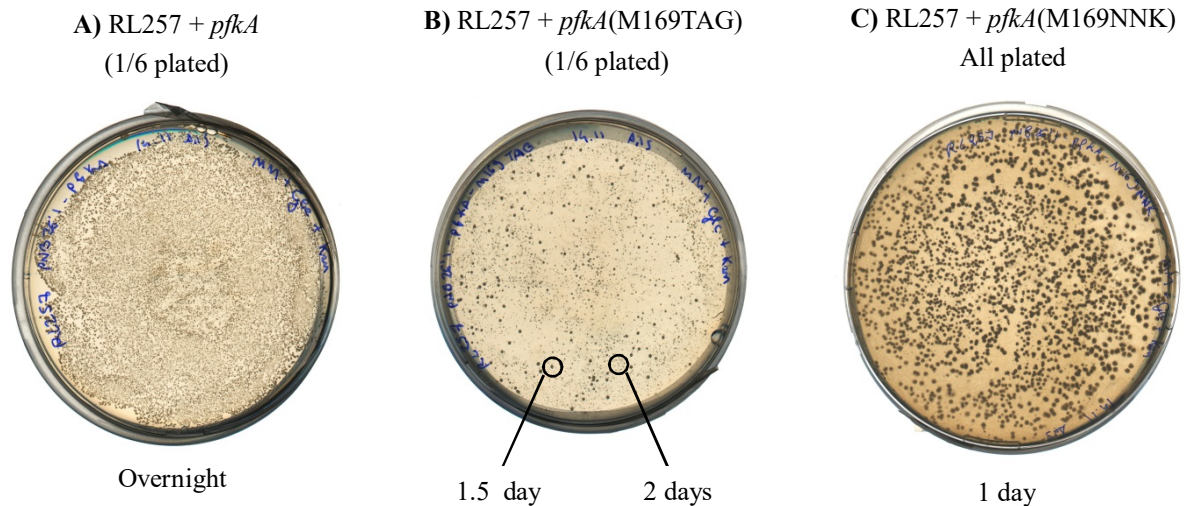
Figure S5. ESI-MS spectra of the purified b* congeners. SPI experiments in presence of two pMEc configurations are shown: *Ec*MetRS/*Ec*tRNA^Met (1-3) and *Sa*MetRS/*Sa*tRNA^Met (4-6). Every spectrum shows three different peaks corresponding to: b*2Aha (10244 Da), b*1Aha-1Eth (10263 Da) and b*2Eth (10283 Da). It was not possible to define the nature of the fourth peak (10304 Da). Aha is shown in yellow, Eth in brown. The peak intensity is relative to the most abundant congener. The experiments were done in triplicate (*n*=3).

PFKI-based selection system

As recipient for genetic selection, *E. coli* strain RL257, which is deficient in PFK-1 and PFK-2 and contains the *lacI^q* allele [5], was used. This strain does not grow in M9 minimal medium containing glucose as the sole carbon source over a period of 25 hours [5] and longer, as observed in our experiments. As selection plasmid, we used the low-copy number pNB26'1_*pfkA*(M169TAG) carrying the gene for PFK-1 (*pfkA*) under the *lac* promoter and having an amber stop codon at position 169. To test the stringency of the selection system, a library of *pfkA* variants containing the random mutation M169NNK (N = any base, K = G or T) was generated by site-saturation mutagenesis using primer X (FW) and Y (RV). With the NNK degenerate codon, all 20 cAAs are theoretically covered by just 32 of the 64 codons. The diversity of the saturation library was estimated by sequencing (Figure S6). Subsequently, the *pfkA*(M169NNK) library was transformed by electroporation into RL257 cells. In parallel, RL257 cells were transformed with plasmid pNB26´1_*pfkA* or plasmid pNB26´1_*pfkA*(M169TAG) to server as positive and negative control, respectively. The transformants were plated onto M9 plates containing 0.2% glucose and kanamycin and incubated at 37 °C. Interestingly, RL257 colonies transformed with the positive control grew overnight despite PFK-1 expression was not induced (Figure S7, A). This finding suggests that leaky expression of *pfkA* in presence of the LacI repressor produces enough PFK-1 to rescue the host. The negative control, instead, showed colonies after 1.5 days. After 2 days, satellite colonies appeared on the same plate (Figure S7, B). Finally, in case of the RL257 strain transformed with the *pfkA*(M169NNK) library, the first colonies appeared after 1 day incubation followed by smaller ones after 2 days (Figure S7, C). Six colonies were picked up from the plate containing the library and the pNB26´1_*pfkA*(M169NNK) plasmid was isolated from them. The *pfkA* gene was then sequenced and in all the 6 cases it was found to carry the methionine codon, ATG, at position 169. A second analysis was made by pooling all the colonies on the plate and isolating again the pNB26´1_*pfkA*(M169NNK) plasmid. Also in this case sequencing results showed the presence of the ATG codon at position 169 (Figure S8).
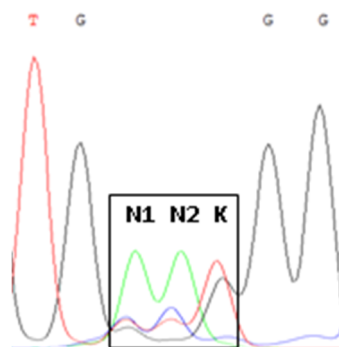
Figure S6. Sequencing results of site-directed saturation mutagenesis at position M169 in *pfkA*. The M169NNK mutation is highlighted in the box. The sequencing chromatogram is represented using the software Ape (by M. Wayne Davis). Green = A; Blue = C; Red = T; Black = G. The randomization frequencies at positions N1 and N2 are higher for the nucleotide "A".

A) RL257 + *pfkA*
(1/6 plated)

B) RL257 + *pfkA*(M169TAG)
(1/6 plated)

C) RL257 + *pfkA*(M169NNK)
All plated

Overnight

1.5 day   2 days

1 day

Figure S7. Complementation of PFK-1-deficient *E. coli* strain RL257. A) Positive control: RL257 strain expressing *pfkA* WT grew on M9 plate + 0.2% glucose + Kan after o/n incubation at 37 °C. **B)** Negative control: RL257 strain expressing *pfkA*(M169TAG) grew on M9 plate + 0.2% glucose + Kan after 1.5 day incubation at 37 °C. **C)** RL257 strain expressing the *pfkA*(M169NNK) library grew on Ø 20 cm M9 plate + 0.2% glucose + Kan after 1 day incubation at 37 °C. Smaller colonies appeared after 2 days incubation.
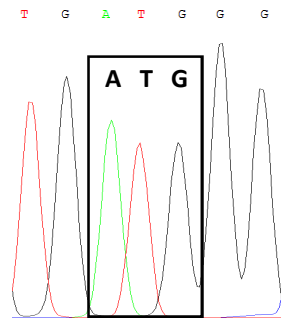
Figure S8. Sequencing results of the *pfkA*(M169NNK) library extracted from RL257 colonies grown on M9 + 0.2% Glu. The randomized codon (169NNK) in the *pfkA* library corresponded to the methionine ATG codon in all the pooled RL257 colonies. The sequencing chromatogram is represented using the software Ape (by M. Wayne Davis).

References

1.   M Hösl, *TU München* PhD thesis (2011).
2.   SK Lee and JD Keasling, *Applied and Environmental Microbiology* **71** (2005) 6856-6862.
3.   KA Datsenko and BL Wanner, *Proceedings of the National Academy of Sciences* **97** (2000) 6640-6645.
4.   LC Thomason, N Costantino, and DL Court, *Curr Protoc Mol Biol* **79** (2014) 1.17.1-1.17.8.
5.   MR Lovingshimer, D Siegele, and GD Reinhart, *Protein expression and purification* **46** (2006) 475-82.