Sense codon emancipation for proteome-wide incorporation of noncanonical amino acids: rare isoleucine codon AUA as a target for genetic code expansion

Nina Bohlke & Nediljko Budisa
Department of Chemistry, TU Berlin, Berlin, Germany

Correspondence: Nediljko Budisa, Department of Chemistry, TU Berlin, Biocatalysis Group, Mueller-Breslau-Straße 10, D-10623 Berlin, Germany. Tel.: +49 30 314 28 821; fax: + 49 30 314 28 279; e-mail: budisa@biocat.tu-berlin.de

Received 29 October 2013; revised 20 December 2013; accepted 20 December 2013. Final version published online 27 January 2014.
DOI: 10.1111/1574-6968.12371
Editor: Jeff Cole

Keywords
genetic code; sense codons; reassignment; orthogonal pairs; synthetic biology; codon emancipation.

Abstract
One of the major challenges in contemporary synthetic biology is to find a route to engineer synthetic organisms with altered chemical constitution. In terms of core reaction types, nature uses an astonishingly limited repertoire of chemistries when compared with the exceptionally rich and diverse methods of organic chemistry. In this context, the most promising route to change and expand the fundamental chemistry of life is the inclusion of amino acid building blocks beyond the canonical 20 (i.e. expanding the genetic code). This strategy would allow the transfer of numerous chemical functionalities and reactions from the synthetic laboratory into the cellular environment. Due to limitations in terms of both efficiency and practical applicability, state-of-the-art nonsense suppression- or frameshift suppression-based methods are less suitable for such engineering. Consequently, we set out to achieve this goal by sense codon emancipation, that is, liberation from its natural decoding function – a prerequisite for the reassignment of degenerate sense codons to a new 21st amino acid. We have achieved this by redesigning of several features of the post-transcriptional modification machinery which are directly involved in the decoding process. In particular, we report first steps towards the reassignment of 5797 AUA isoleucine codons in Escherichia coli using efficient tools for tRNA nucleotide modification pathway engineering.

The genetic code: degeneracy, wobbling and codon usage

The nearly identical assignment of triplet codons to amino acids in all living organisms illustrates the universality of the genetic code and affirms the existence of a common origin. With only very few exceptions, ribosomal synthesis is restricted to 20 naturally occurring amino acids. Codon assignment in living cells relies on highly selective aminoacyl-tRNA synthetases (aaRS) which recognize both the amino acid to be charged and the related tRNA. By catalyzing a specific aminoacylation reaction, these enzymes provide the critical link between amino acids and codons and thereby establish the genetic code. As a general rule in living cells, one aminoacyl-tRNA synthetase exists for each canonical amino acid.

The most remarkable feature of the genetic code is its degeneracy. Of 64 possible triplet combinations, 61 sense codons encode the 20 amino acids, while three stop codons (UAA, ocher; UGA, opal; and UAG, amber) are used to terminate translation (see Fig. 1). With the exceptions Met and Trp, most amino acids are thus represented by more than one codon. On average, each amino acid is encoded by three different codons. For example, Ile is encoded by AUC, AUU, and AUA. Surprisingly, these three codons are decoded by only two different tRNA molecules. The reason for this is wobble base pairing of the third position as proposed by Crick (1966). Different triplets encoding the same amino acid often vary only in the third position as evident from the aforementioned Ile codons. In this case, the first two codons, AUC and AUU, are decoded by the same anticodon GAU, either involving a G-C Watson–Crick base pairing or a G-U wobble base pairing. In general, the first two bases of a triplet form Watson–Crick base pairings, whereas the third position is able to establish wobble base pairings. In case
of the third Ile codon AUA, the first base (wobble position) of the anticodon CAT is post-transcriptionally modified with a lysine to form lysidine (L), which enables A-L wobble base pairings (details are described below).

Another important feature of the genetic code is the codon usage, the abundance of a codon in a particular DNA sequence, which can vary significantly between different organisms. For Ile decoding by *Escherichia coli*, AUU and AUC are frequently used, while the AUA codon is rarely used. A comparison of AUA codons used per 1000 bases in *E. coli* (3.7), *Saccharomyces cerevisiae* (17.8), and *Homo sapiens* (7.5) highlights significant differences (Nakamura et al., 2000).

At the time of its deciphering, the genetic code was suggested to be universal in all organisms. Crick proposed a first theory on its evolution saying that the genetic code was a result of a ‘frozen accident’ unable to evolve further, since ‘no new amino acid could be introduced without disrupting too many proteins’ (Crick, 1968; Söll & RajBhandary, 2006).

**Natural variations of the genetic code**

The theory of a frozen universal genetic code is believed to be disproved by genetic code variations found in vertebrate mitochondria, where AUA encodes Met instead of Ile and UGA encodes Trp instead of being a translational termination signal (Barrell et al., 1979; Osawa et al., 1992). Today, around 20 variations of the standard genetic code are known, not only in organelles but also in free-living microorganisms (Ambrogelly et al., 2007; Campbell et al., 2013; O’Donoghue et al., 2013).
Well-documented natural variations suggest the feasibility of genetic code expansion. The design of organisms capable to decode more than the common 20 building blocks could lead to proteins or even organisms with novel activities (Budisa, 2005). One of the earliest discovered variations of the genetic code were found to be deviations in codon assignment between nuclear and mitochondrial genes (Barrell et al., 1979; Himeno et al., 1987; Osawa et al., 1992; Wolstenholme, 1992; Söll & RajBhandary, 1995).

Deviations in the codon meanings, especially in various mitochondrial genomes, clearly indicate that in the context of the smaller (‘minimal’) genome and proteome, such reassignments in the code structure are more likely, as lethal effects are less possible (Budisa, 2005). In other living beings, species-specific reassignments are sometimes employed as beneficial adaptive variations in a particular environment. One example is the Leu→Ser change for the CUG codon in some Candida species which probably leads to better survival of these mainly parasitic species in their hosts (Tuite & Santos, 1996). Disturbing effects of such changes are avoided in these cells as CUG is usually not used as a codon in the cellular mRNA. This reassignment is ‘reserved’ for specific genomic islands of this Candida pathogen; ‘island genes’ are usually involved in stress response (like heat shock) and pathogenicity (Suzuki et al., 1997; Santos et al., 2004).

Changes in codon meaning can be classified into two general categories. The first category includes reassignment of standard termination codons (UGA, UAA, and UAG) to Trp or Gln in some prokaryotes, archaea, mitochondria of several organisms, and even in nuclear genes of some protozoa (e.g. in Mycoplasma mobile the UGA termination codon encodes Trp). The second category of codon reassignment includes altered meanings in mitochondrial sense codons such as Ile→Met (Barrell et al., 1979; Budisa, 2005). The mitochondrial AUA codon is deciphered by the tRNAMet containing an unusual modified nucleotide 5-formylcytidine (FC) in the first position (number 34) of their anticodon. This tRNA translates both AUG and AUA as methionine (Takemoto et al., 2009). Sixty-two such reassignments are known in mitochondria, but only one in nuclear genes: CUG Leu→Ser in Candida cylindracea (see above) (Osawa et al., 1992; Tuite & Santos, 1996; Suzuki et al., 1997; Santos et al., 2004). As they do not affect the amino acid repertoire, such reassignments might be seen as flexibility in the ‘frozen’ structure of the genetic code.

To date, two examples for a natural expanded genetic code are known. An in-frame UGA termination codon also encodes selenocysteine (Sec), the 21st cotranslationally inserted amino acid. This recoding mechanism requires a tRNA_{UCA}, a specialized translation elongation factor (SelB) and a particular mRNA stem-loop structure, known as the selenocysteine insertion element (SECIS) (Söll, 1988; Schön et al., 1989; Cusack et al., 2005; Ambrogelly et al., 2007). UAG is ambiguous in Methanosarcinaceae where, in addition to serving as a termination codon, it also encodes pyrrolysine (Pyl), the 22nd cotranslationally inserted amino acid. In this case, a new tRNA synthetase, pyrrolysyl-tRNA synthetase (PylRS), is essential for this recoding event (Söll, 1988; Hao et al., 2002; Srinivasan et al., 2002; Blight et al., 2004; Polycarpo et al., 2004).

In fact, no known examples for codon reassignment introduce a noncanonical amino acid in response to a sense codon.

**Experimental variations of the genetic code**

As part of a synthetic genetic code, amino acids can be replaced by noncanonical (nCAA) ones (engineering), or alternatively, an nCAA can be added to the natural repertoire (expansion). Genetic code engineering commonly involves the use of wild-type aminoacyl-tRNA synthetases (aaRSs) to incorporate nCAAs that are close structural analogs of canonical amino acids. In this approach, a strain auxotrophic for one canonical amino acid (cAA) is used for substitution of the cAA with a structurally similar nCAA analog. In the newly synthesized proteins, the canonical amino acid is efficiently replaced by its analog at all sites (global replacement). This method is often called residue-specific incorporation, as all codons coding for a given amino acids are reassigned (Budisa, 2004). Basically, genetic code engineering leads to proteins containing 19 cAA and one nCAA or 18 cAA and 2 nCAA. Up to three different nCAAs incorporated into a single protein were reported with this method (Leptien et al., 2010). Wang et al. (2004) showed that one mRNA can be translated in different ways depending on the relative rates of competing aminoacylation reactions. Depending on the intracellular nCAA levels and whether the host is engineered with an enhanced ValRS or IleRS activity, (2S, 3R)-4, 4-Trifluorovaline could be assigned either to isoleucine or valine codons (Wang et al., 2004). Using this genetic engineering approach, proteins with new activities and properties can be generated, but the amino acid repertoire is still restricted to 20 amino acids.

Genetic code expansion allows for the incorporation of a 21st amino acid. This is usually achieved by assignment of a ‘blank’ codon (nonsense or frameshift suppression). This method enables the site-specific incorporation of an nCAA within the growing polypeptide chain, with the mRNA containing a nonsense codon at the desired position. This is realized by the use of an orthogonal suppressor aaRS:tRNA pair (o-pair) capable to decode
the nonsense codon. To date, various orthogonal aaRS: tRNA pairs have been developed (Liu & Schultz, 2010) with the TyrRS from Methanocaldococcus jannaschii (mTyrRS) being most frequently used (Wang et al., 2001). More recently, pyrrolysine aaRS (PylRS) and its cognate suppressor tRNA pyT from Methanosarcinae species have been shown to be highly potent orthogonal ncAA systems (Mukai et al., 2008; Neumann et al., 2008; Huang et al., 2010). In contrast to mTyrRS, the ‘natural’ PylRS:tRNA o-pair already displays a broad substrate tolerance towards orthogonal ncAAs, but PylRS does not utilize the common 20 amino acids as substrates (Polycarpo et al., 2006; Li et al., 2009). Within a short time, a large number of different PylRS variants for incorporation of versatile ncAAs were developed (Fekner et al., 2010).

Nonsense suppression and frameshift suppression are widely used in academic research, although these approaches are still limited by several means. Both typically result in low protein yields due to cellular toxicity of the ncAA, suffer from context effects, and compete with the highly specialized endogenous termination machinery (Doerig et al., 1988; Leptien et al., 2010; Young et al., 2010). A detailed description about genetic code engineering and expansion is given elsewhere (Wang & Schultz, 2004; Wang et al., 2006; Liu & Schultz, 2010; Antonczak et al., 2011; Hoesl & Budisa, 2012).

**Sense codon-specific incorporation of ncAAs**

Ultimately, the incorporation of an ncAA typically involves outcompeting the natural aaRS substrate, recoding a stop codon, or the use of four to five base codons. As most amino acids are decoded by more than one codon (exceptions: Trp and Met), our future aim is to emancipate a codon from its natural meaning and to reassign it to an ncAA. This would allow proteome-wide incorporation of a 21st amino acid in response to a sense codon.

It is assumed that between 30 and 40 sense codons are required to encode the genetic information of an organism. Therefore, a large number of sense codons (> 20) should be available for recoding experiments (Krishnakumar et al., 2013). Correspondingly, the genetic code would be substantially expanded by such proteome-wide replacements. To our understanding, a synthetic organism with an expanded genetic code is not able to live without the ncAA, as the newly reassigned sense codon becomes not assigned.

One step towards sense codon reassignment was described by Kwon et al. (2003). In E. coli, Phe is naturally encoded by UUC and UUU codons, and a single tRNA$^{Phe}_{GAA}$ decodes both via Watson–Crick base pairing and wobble base pairing, respectively. Introduction of a heterologous tRNA with an AAA anticodon (ytRNA$^{Phe}_{AAA}$) and a corresponding aaRS (yaRS) from S. cerevisiae resulted in competition between the two tRNAs for the UUU codon. The synthetic site of yPheRS was engineered to activate the ncAA 2-naphthylalanine. Depending on experimental setup, assignment of UUC to Phe was quantitative and UUU translation into 2-naphthylalanine reached 80% (Kwon et al., 2003). Nevertheless, the cells were still able to incorporate Phe at UUU positions. In the absence of the ncAA, the cells could easily grow using their endogenous tRNA$^{Phe}_{GAA}$ for the incorporation of Phe at UUU positions. A complete sense codon-specific reassignment would lead to cell death in the absence of the ncAA. A nice overview about reassignment of sense codons in vivo and corresponding methods is given by Link & Tirrell (2005) and extensively discussed by Budisa (2005, pp. 97–100).

**Choice of the degenerate codon: AUA as a target for reassignment**

Real expansion of the genetic code will take place once when we would be able to add synthetic amino acids in a proteome-wide manner by genome-wide reassignments of target codons. Here, we elaborate the concept and report first experimental steps towards the emancipation/liberation of degenerate sense codons from their canonical function, a step towards estranging the genetic code that should enable addition of novel amino acids to the existing repertoire.

For the choice of the sense codon to be reassigned, two aspects were considered highly important. The first important point is the presence of a specific modification pathway, which is necessary for the decoding process of the codon. Through the selective interruption of this modification, the decoding mechanism can be eliminated. Thus, the codon is liberated from its original meaning and is available as an open sense codon. Now, the certain codon can be reassigned to a noncanonical amino acid. To minimize disrupting effects through the appearance of an ncAA in the whole proteome, it is favorably to select a codon rarely used. The second important selection criterion is the choice of a noncanonical amino acid whose proteome-wide replacement will be tolerated by the cell. Therefore, choosing a codon which encodes an amino acid with little metabolic involvement and catalytic functionality might be a good starting point. In the genetic code, all coding triplets with a central U (the so-called XUX group of codons) are cognate to apolar (hydrophobic) amino acids, for example Leu, Ile, and Val. These residues are chemically inert and rarely directly involved in the catalytic function of proteins, but frequently play an essential role in protein folding as well as in binding/recognition of hydrophobic ligands. Their synthetic
chemical analogs such as trifluoroleucine have been known for more than 50 years to be translationally competent, allowing their incorporation into the nascent polypeptide chain (Rennert & Anker, 1963).

With Ile rarely involved in catalytic reactions, the rare codon AUA perfectly fits for reassignment experiments. In E. coli, the codon usage of AUA is low. With only 5797 occurrences in the genome, it is usually involved in regulatory events such as ribosomal pausing (Lajoie et al., 2013a). Consequently, we speculated that replacement of Ile by ncAA at AUA positions might be tolerated by the E. coli cell.

Next, codon ambiguity has widespread effects on protein function mainly resulting in detrimental effects on cellular fitness. However, these growth defects can be compensated by multiple mutations in various proteins by improving their ability to grow in the presence of the ncAA. For example, Philippe Marlière and coworkers generated E. coli mutant strains capable of installing translational pathways specific for the incorporation of additional amino acids into proteins in vivo (Lemeignan et al., 1993). Their selection procedure resulted in stable bacterial strains with coding triplets that can be read ambiguously. Similarly, Döring & Marlière (1998) constructed stable bacterial strains with ambiguous reading of the rare coding triplet AUA by overexpression of CysRS:tRNA^{Cy}³. Experimentally designed selective pressure yielded almost full reassignment (Ile→Cys) of the AUA codon in E. coli. Such proteome-wide substitutions were possible as Ile side chains evidently do not act as catalytic residues in any enzyme of the E. coli proteome (Döring & Marlière, 1998; Budisa, 2005). Although the repertoire of amino acids was not enlarged, it could be shown that Cys miscoded at the Ile AUA codon was well tolerated by the cell and that codon reading as part of normal translation can be altered in vivo. Döring et al. (2001) were also able to globally substitute Val by the ncAA aminobutyric acid (Abu) in a quarter of the cellular proteome using an E. coli strain with defective editing function of ValRS (Döring et al., 2001). Furthermore, a nonreverting strain of E. coli with impaired editing function has been generated. Under Ile starvation conditions and relative to a wild-type E. coli strain, it achieves higher culture growth in combination with analogs such as norvaline (Pezo et al., 2004). Later, Bacher et al. (2005) abolished the editing function of IleRS and found that bacteria grow slowly even with half of their isoleucine residues being miscoded. Taken together, the rare Ile AUA codon seems to be a feasible target for sense codon reassignment experiments.

Finally, proteome-wide addition of noncanonical amino acids associated with breaking the degeneracy of the genetic code should be tolerated by correspondingly engineered cells, in an ideal case by setting up long-term cultivation experiments (Lenski et al., 1991). In the Lenski experiments or using automated devices (de Crécy-Lagard et al., 2001), such genetic estrangement through codon reassignment associated with the use of ncAAs as building blocks would be enforced towards propagation and selection of robust and viable cells with built-in alien nutrients.

**Rare isoleucine AUA codon decoding in E. coli**

With great accuracy, tRNA isoacceptors are aminocylated with their related amino acid in vivo. In E. coli, 49 tRNAs function to read the sense codons of the messenger RNA, converting them into 20 amino acids of the protein sequences (Budisa, 2004). Tuning of tRNA-mediated translation has been enabled by more than 70 chemically distinct post-transcriptional base modifications in tRNA species (Agris et al., 2007).

In E. coli, isoleucine is decoded by three different codons. The two frequently used codons, AUG and AUU, are translated by the high-abundance tRNA species, tRNA^{Ile}³, which harbors a G34AU anticodon and pairs with the two codons. The rare AUA codon is translated by low-abundance tRNA^{Ile}² that carries the same anticodon (CAU) as tRNA^{Met}. Together, the tRNA^{Ile}³ and tRNA^{Ile}² pair is able to decipher all Ile codons accurately (Muramatsu et al., 1988a, b; Soma et al., 2003; Ikeuchi et al., 2005; Nakaniishi et al., 2009).

Bacteria decode AUA via a tRNA in which the wobble position C34 is modified with a lysine, resulting in lysidine (i.e., 2-lysyl-cytidine). The enzyme catalyzing this modification is called lysidine-tRNA synthase (TltS) (Muramatsu et al., 1988a; Soma et al., 2003; Silva et al., 2006). This lysine-containing cytidine derivative occurs in almost all bacteria (Harada & Nishimura, 1974; Muramatsu et al., 1988b), whereas another chemical modification (2-agmatinylcytidine, agm2C) exists in archaea. The latter modification is introduced by agmatidine synthase TiaS (Ikeuchi et al., 2010; Mandal et al., 2010). The minor tRNA^{Ile}² lacking the lysyl/agmatidine group on C34 is aminocylated by MetRS and exclusively reads the Met codon AUG (Silva et al., 2006). An overview about Ile decoding in E. coli is given in Fig. 2.

It was shown that ErnRNA^{Ile}² bearing an LAU anticodon is charged exclusively with Ile by IleRS and recognizes only the AUA codon. According to the chemical structure of lysidine (Fig. 3a), conjugation of the lysine moiety at the C2 position of cytidine induces a tautomeric conversion with protonation of the N3 position and imino group formation at the C4 position. By this modification, the proton donor–acceptor pattern of C in hydrogen bonding is completely altered towards the function of U,
enabling L to pair with A but not with G (Suzuki & Miyauchi, 2010).

In summary, the L34 modification has a dual role: It causes the mature tRNA\textsubscript{Ile2} to be aminoacylated by IleRS and converts the codon specificity from AUG to AUA (Muramatsu et al., 1988a, b) (Figure 3b).

An essential enzyme: tRNA\textsuperscript{Ile}\textsuperscript{lysidine} synthetase

Soma et al. (2003) identified TilS as the enzyme responsible for the formation of lysidine. It could be shown that partial inactivation of \textit{tilS} results in an AUG-dependent translational defect, which supports the notion that TilS is an RNA modifying enzyme that plays a critical role in the accurate decoding of the genetic code.

Complete TilS inactivation is lethal, as the AUA codon becomes an unassigned codon without its cognate decoding tRNA (Nakanishi \textit{et al.}, 2009). Therefore, bacterial \textit{tilS} was proposed to be one of the 206 essential protein-coding genes required for maintaining bacterial cell life (Gil \textit{et al.}, 2004) and to be essential for reconstructing a minimal synthetic cell (Mushegian & Koonin, 1996; Moya \textit{et al.}, 2009; Gibson \textit{et al.}, 2010). Rescue of the translational defect caused by inactivation of TilS could be performed by a specialized tRNA able to decode the AUA codon. However, remember that codons within the Ile/Met decoding family box (e.g. AUA/AUG) are among the most difficult coding units to be discriminated accurately.

Nonetheless, the sequenced genomes of a few-living organisms appear to lack a TilS homolog and tRNA\textsubscript{Ile2}\textsubscript{CAU} (Jaffe \textit{et al.}, 2004; Silva \textit{et al.}, 2006), which suggest that alternative mechanisms of AUA translation exist. Bacterial genomes lacking \textit{tilS} are found in \textit{Bifidobacterium} (two species), \textit{M. mobile}, and \textit{Neorickettsia} (two species) (Fabret et al., 2011). These species possess a U34-containing tRNA\textsuperscript{Ile3} (tRNA\textsubscript{Ile2}\textsubscript{UAU}) for deciphering AUA codons. Notably, there are bacterial species that contain all three a U34-harboring tRNA\textsuperscript{Ile3}, a \textit{tilS}, and a C34-containing tRNA\textsuperscript{Ile2} gene pair in their genomes (\textit{Lactobacillus}, \textit{Rhodopirellula}, \textit{Planctomyces}, \textit{Ktedonobacter}, \textit{Cyanothece}) (Fabret \textit{et al.}, 2011). In this way, these species can be regarded as evolutionary intermediates between the predominant group of TilS-containing organisms and the rare group of organisms lacking the functional pair of \textit{tilS/tiaS} and tRNA\textsubscript{Ile2}\textsubscript{CAT} genes. In other words, they contain a temporary ambiguous decoding system – a prediction of the ‘ambiguous intermediate’ theory (Schultz & Yarus, 1994a, b; Jones \textit{et al.}, 2008; Fabret \textit{et al.}, 2011).

\textbf{Mycoplasma mobile tRNA\textsuperscript{Ile3}\textsubscript{UAU}:IleRS pair is able to decipher AUA codons in \textit{E. coli}}

As mentioned before, \textit{M. mobile} does not carry the \textit{tilS} gene but a tRNA\textsubscript{Ile3}\textsubscript{UAU} without any modification. The
acquisition of a tRNA$_{\text{UAU}}^{\text{Ile3}}$ gene by M. mobile has led to the loss of both the tRNA$_{\text{CAT}}^{\text{Ile}}$ and the tilS gene (Silva et al., 2006). Very recently, Taniguchi et al. (2013) demonstrated that MnmtIleRS recognizes the UAU anticodon without any modification, whereas E. coli IleRS did not efficiently aminoclylate tRNA$_{\text{UAU}}^{\text{Ile3}}$ in vitro. They found Arg865 in MnmtIleRS to be critical for the specificity to recognize the AUA codon and showed that when the corresponding site in EcIleRS W905 is mutated to Arg, EcIleRS was also able to charge tRNA with a UAU anticodon efficiently. Furthermore, they found that MnmtRNA$_{\text{UAU}}^{\text{Ile3}}$ cannot distinguish between AUA and AUG codons on E. coli ribosomes. This finding suggests that in E. coli, tRNA$_{\text{UAU}}^{\text{Ile3}}$ might also read AUG codons as Ile to an extent that could lead to translational defects. As tRNA$_{\text{CAT}}^{\text{Met}}$ strongly recognizes the AUG codon in the cell, potential misreading might be reduced by competition of MnmtRNA$_{\text{UAU}}^{\text{Ile3}}$ with endogenous tRNA$_{\text{CAT}}^{\text{Met}}$ (Taniguchi et al., 2013).

**Escherichia coli Atils strain rescued with MnmtRNA$_{\text{UAU}}^{\text{Ile3}}$:IleRS capable to read rare AUA codons**

We have now tested the ability of M. mobile tRNA$_{\text{UAU}}^{\text{Ile3}}$:IleRS pair (MmPair) to rescue the lethal tilS deletion in E. coli through direct AUA codon reading. The ability to decode the AUA codon with a heterologous tRNA$_{\text{UAU}}^{\text{Ile3}}$:IleRS pair – which can later be modified to incorporate an ncAA at AUA positions – is the first step towards AUA sense codon reassignment.

To analyze whether the MnmtRNA$_{\text{UAU}}^{\text{Ile3}}$:IleRS pair (MmPair) is able to read the AUA codon in a tilS-depleted E. coli strain, we constructed several rescue vectors. First, tRNA$_{\text{UAU}}^{\text{Ile3}}$ and MnmtIleRS were cloned separately as well as together into plasmid pNB26’2. Furthermore, tilS was cloned into pNB26’1 (Figure S1) as a positive control. Both plasmids carry a p15A origin of replication and pLac promoter. These four different plasmids were used in tilS deletion experiments designed to rescue AUA codon readout. The chromosomal tilS in E. coli MG1655 was deleted following the method of Datsenko & Wanner (2000). Obtained colonies were analyzed by PCR with different confirmation primers (Fig. 4a; Oligonucleotide sequences are given in Supporting Information, Table S1). For experimental details see Data S1.

For pNB26’1_tils, genomic tilS was successfully replaced by cat (chloramphenicol acetyl-transferase) as shown in Fig. 4b (1 and 2). Colonies obtained for pNB26’2_MnmtRNA$_{\text{UAU}}^{\text{Ile3}}$ and for pNB26’2_MnmtIleRS retained the tilS gene and were most probably spontaneous Cm-resistant colonies (Figure 4b (2 and 3)). These findings suggest that E. coli IleRS is not able to charge MnmtRNA$_{\text{UAU}}^{\text{Ile3}}$ in vitro, which is consistent with the *in vitro* results of Taniguchi et al. (2013). Regarding the deletion experiment with pNB26’2_MnmtIleRS, two explanations are plausible. First, MnmtIleRS might not able to charge EcRNA$_{\text{CAT}}^{\text{Ile}}$ and EcRNA$_{\text{CAT}}^{\text{Ile3}}$. Alternatively, MnmtIleRS could charge one or both EcRNA, whereas the aminoclylated tRNA is not able to decipher the AUA codon on the E. coli ribosome. As Taniguchi et al. (2013) could show *in vitro* aminoclylation of EcRNA$_{\text{CAT}}^{\text{Ile3}}$ by MnmtIleRS, the second option is most probably correct. Nevertheless, using tRNA or IleRS separately, it was not possible to completely remove tilS from the E. coli genome. In contrast, the combination of MnmtRNA$_{\text{UAU}}^{\text{Ile3}}$ and MnmtIleRS was able to rescue the lethal tilS deletion in E. coli MG1655, rendering the cells viable without the essential enzyme TilS.

The generated E. coli MG1655 ΔtilS::cat strain (pNB26’2_MmPair) grows slower than the wild-type equivalent (Fig. 5 and Fig. S2). The calculated doubling times are 38 and 28 min, respectively (Fig. S3). The lower growth rate might be due to interference with the AUG codon by MnmtRNA$_{\text{UAU}}^{\text{Ile3}}$ or due to low intracellular levels of MnmtRNA$_{\text{UAU}}^{\text{Ile3}}$. Currently, we are investigating this phenomenon in more detail.

Taken together, our results show that it is possible to liberate the AUA codon from its natural deciphering mechanism in E. coli. Using MnmtIleRS and MnmtRNA$_{\text{UAU}}^{\text{Ile3}}$, we were able to restore AUA decoding and to rescue the lethal chromosomal tilS deletion (see Fig. 4). Figure 6 represents the suggested rescue mechanism of the MmPair. In future, we will engineer an orthogonal aaRS to specifically incorporate a 21st ncAA at AUA positions to generate an E. coli strain with a proteome-wide expanded genetic code through sense codon reassignment.

**Outlook: towards genetic code expansion via experimental reassignment of AUA codons**

Currently available methodologies for genetic code expansion are well suited for small-scale experiments focused on particular academic questions (e.g. those using fluorescence tags) but are of quite limited use for the design and creation of robust artificial diversity in synthetic cells. Thus, we believe that it is necessary to replace the traditional suppression-based approaches with reassignments of sense (preferably rare) codons. However, we are well aware that rare codons such as AUA are often part of translational fine-tuning systems such as ribosomal pausing (Zhang et al., 2009). Nonetheless, we succeeded to develop a rescue strategy that compensated for the translational defects by importing the AUA decoding system from M. mobile, which consists of MnmtIleRS and MnmtRNA$_{\text{UAU}}^{\text{Ile3}}$.
In the next step, we will use an engineered isoleucyl-tRNA synthetase (IleRS) from an evolutionary distant organism to reassign the AUA codon with an ncAA. Just like related enzymes for branched-chain amino acids (Thr, Val, Leu), this IleRS exhibits a hydrolytic editing that is capable to dissociate noncognate amino acids. In conjunction with the substrate recognition and activation by the active site architecture, this leads to the ‘double sieve’ concept (Fersht, 1977; Fersht & Dingwall, 1979). The availability of high-resolution X-ray crystal structures of different IleRSs (in apo-forms or in tRNA/ATP/amino acid complexes) is thus of enormous importance for the identification of residues crucial for amino acid recognition as well as for aaRS:tRNA interactions. The basic enzyme engineering strategy will rely on well-established mutagenesis approaches to redesign the binding pocket and to attenuate the aaRS editing activity. For example, Mursinna & Martinis (2002) reported a rational design that blocked the amino acid editing reaction in EcLeuRS. In particular, a

**Fig. 4.** (a) Confirmation primers used to verify tilS deletion in genomic DNA of *Escherichia coli* MG1655. Primer combination 1kb_up_tilS_for and tilS_conf_rev amplifies a part of tilS and 1 kb upstream of genomic *E. coli* DNA. Combination tilS_for and tilS_rev amplifies tilS in genomic or plasmid DNA. Combination 1kb_up_tilS_for and Cmr_conf_rev amplify part of cat and 1 kb upstream of genomic *E. coli* DNA to verify that cat replaced tilS. (b) PCR verification of ΔtilS:cat genotype in *E. coli* MG1655. 1) PCR results with primers 1kb_up_tilS_for and Cmr_conf_rev. For single *Mm* tRNA³UAU, no replacement of tilS with cat was observed. As a control, also *E. coli* MG1655 WT showed no PCR product. For *Mm* IleRS, tilS on a plasmid and the combination of *Mm* tRNA³UAU with *Mm* IleRS, tilS replacement with cat is observed. 2) PCR results with primers 1kb_up_tilS_for and tilS_conf_rev. For all variants except *Mm* tRNA³UAU together with *Mm* IleRS and tilS on a plasmid, tilS was found not to be deleted. 3) PCR results with primers tilS_for and tilS_rev. These results show that only for the *Mm* tRNA³UAU/*Mm* IleRS combination, *E. coli* is able to live without tilS.

**Fig. 5.** Comparison of growth curves between *E. coli* MG1655 WT (black line) and *E. coli* MG1655 ΔtilS::cat (pNB26'2_MmPair) (red line).
single Thr at position 252 was found to play a critical role, with its mutation to bulky amino acids like Phe or Tyr resulting in the loss of the editing function (Mursinna & Martinis, 2002). Tang & Tirrell (2002) have identified three LeuRS mutants (Thr252Tyr, Thr252Leu, Thr252Phe) with attenuated editing activities that allow successful translation of norvaline, allylglycine, homoallyglycine, homopropargylglycine, and 2-butynylglycine at Leu positions of the model protein (Tang & Tirrell, 2002). This approach should be applicable to all 'double sieve' aaRS for branched-chain amino acids such as ThrRS, ValRS, LeuRS, and IleRS.

By a variety of mechanisms, long-term cultivation of bacteria can lead to increased mutation rates—the so-called adaptive mutation phenomenon (Marlière et al., 2011). Under such conditions, these organisms need to explore both a larger genetic space and a large protein folding space to accommodate the desired ncAAs in their proteome and intermediary metabolism. Thus, we anticipate long-term cultivation experiments using a special turbidostat system as a promising tool for creating codon-emancipated cells (de Crécy-Lagarde et al., 2001). Via a controllable and defined selective pressure, codon removal will be actively enforced in microbial cells, allowing the reassignment of AUA codons to suitable Ile analogs. Subjected to successive selection cycles in continuous culture, *E. coli* strains with ambiguous reading of this codon will prevail. Eventually, sufficiently stable variants will be perpetuated.

We are fully aware that the reassignment of AUA codons to ncAAs potentially causes detrimental effects on cell survival due to their partial localization in essential genes. However, the same argument also applies to the elimination or reassignment of amber stop codons (UGA) in the genome of *E. coli* (321 ORFs). Nonetheless, the groups of Church, Sakamoto, and Yokoyama have shown very recently that the latter codon can be either fully replaced via oligonucleotide-mediated MAGE technology or reassigned by simple knock out of release factor 1 (RF1), compensated by the presence of a suppressor tRNA capable to decode the UAG codon (Mukai et al., 2010; Isaacs et al., 2011; Johnson et al., 2011). Together, the replacement of all UAG codons with UAA codons in *E. coli* and the deletion of RF1 allowed for the reassignment of UAG translation function (Lajoie et al., 2013b). Recently, George Church and coworkers analyzed the possibility of replacing degenerate rare codons in essential genes with a synonymous codon. They selected 42 essential genes and removed all instances of 13 rare codons from these genes indicating that genome-wide removal of rare codons is feasible (Lajoie et al., 2013a). In light of these results, our strategy of proteome-wide AUA codon reassignment appears feasible.

**Acknowledgements**

We are indebted to Dr Tobias Baumann for critically reading this manuscript. We gratefully acknowledge financial support from the DFG Sonderforschungsbereich 765 (SFB 765; Multivalenz als chemisches Organisations- und Wirkprinzip), FP7 EU-funded METACODE-Consortium, UniCat Excellence Cluster of TU Berlin, and from DFG Forschergruppe 1805 (Ribosome Dynamics in Regulation of Speed and Accuracy of Translation).

**References**


Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. Experimental details.

Fig. S1. Schematic presentation of the pNB26′2 cargo site.

Fig. S2. Comparison of growth curves between Escherichia coli MG1655 WT (black line) and E. coli MG1655 AtilS: cat (pNB26′2_MmPair) (red line) including standard deviation error bars.

Fig. S3. Semi-logarithmic plot of OD600 nm values over time.

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