RESEARCH PAPER

Recoding of the selenocysteine UGA codon by cysteine in the presence of a non-canonical tRNA^{Cys} and elongation factor SelB

Oscar Vargas-Rodriguez ^a, Markus Englert^a, Anna Merkurjev^b, Takahito Mukai^a, and Dieter Söll ^{a,b}

^aDepartment of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA; ^bDepartment of Chemistry, Yale University, New Haven, CT, USA

ABSTRACT

In many organisms, the UGA stop codon is recoded to insert selenocysteine (Sec) into proteins. Sec incorporation in bacteria is directed by an mRNA element, known as the Sec-insertion sequence (SECIS), located downstream of the Sec codon. Unlike other aminoacyl-tRNAs, Sec-tRNA^{Sec} is delivered to the ribosome by a dedicated elongation factor, SelB. We recently identified a series of tRNA^{Sec}-like tRNA genes distributed across Bacteria that also encode a canonical tRNA^{Sec}. These tRNAs contain sequence elements generally recognized by cysteinyl-tRNA synthetase (CysRS). While some of these tRNAs contain a UCA Sec anticodon, most have a GCA Cys anticodon. tRNA^{Sec} with GCA anticodons are known to recode UGA codons. Here we investigate the clostridial *Desulfotomaculum nigrificans* tRNA^{Sec}-like tRNA^{Cys}, and show that this tRNA is acylated by CysRS, recognized by SelB, and capable of UGA recoding with Cys in *Escherichia coli*. We named this non-canonical group of tRNA^{Cys} as 'tRNA^{Rec}' (Recoding with Cys). We performed a comprehensive survey of tRNA^{Rec} genes to establish their phylogenetic distribution, and found that, in a particular lineage of clostridial *Pelotomaculum*, the Cys identity elements of tRNA^{Rec} had mutated. This novel tRNA, which contains a UCA anticodon, is capable of Sec incorporation in *E. coli*, albeit with lower efficiency relative to *Pelotomaculum* tRNA^{Sec}. We renamed this unusual tRNA^{Sec} derived from tRNA^{Rec} as 'tRNA^{ReU}' (Recoding with Sec). Together, our results suggest that tRNA^{Rec} and tRNA^{ReU} may serve as safeguards in the production of selenoproteins and – to our knowledge – they provide the first example of programmed codon-anticodon mispairing in bacteria.

ARTICLE HISTORY

Received 16 March 2018
Accepted 1 May 2018

KEYWORDS

Genetic code; recoding;
selenocysteine; tRNA;
bioinformatics

1. Introduction

During mRNA translation, translational recoding allows dynamic redefinition of the meaning of a codon, or the dual use of a codon. This phenomenon is best illustrated by the reassignment of the stop codon UGA to selenocysteine (Sec). Faithful recoding requires the presence of the Sec-insertion sequence (SECIS) present in selenoprotein-encoding mRNAs and a Sec-tRNA^{Sec}-specific elongation factor, SelB (Fig. 1) [1]. In bacteria, SECIS is located in the vicinity of the UGA codon and is recognized by SelB (Fig. 1) [1–3].

Selenium, an essential micronutrient in mammals, is present in proteins as Sec, an amino acid present in organisms from all domains of life. Sec is a close relative of cysteine (Cys) whose thiol group is replaced by a selenol moiety. Sec has electronic properties different from Cys, and improves the catalytic efficiency of some redox enzymes (relative to Cys homolog). Hence, Sec is present in redox and antioxidant proteins with critical roles in the redox homeostasis of cells [1,4]. Unlike other genetically encoded amino acids, Sec has no cognate aminoacyl-tRNA synthetase (aaRS), and Sec-tRNA is formed in a tRNA-dependent manner. Its synthesis begins by attachment of Ser to selenocysteine tRNA (tRNA^{Sec}) by seryl-tRNA

synthetase (SerRS). The resultant Ser-tRNA^{Sec} is then converted to Sec-tRNA^{Sec} by selenocysteine synthase (SelA).

tRNA^{Sec} is central for Sec translation, crucially dependent on correct interactions with three different proteins: SerRS, SelA, and SelB (Fig. 1). Although tRNA^{Sec} shares significant structural homology with tRNA^{Ser}, it has unique sequence and structural elements that prevent interactions with other aaRSs and elongation factor Tu (EF-Tu) while promoting the recognition of SelA and SelB. For instance, the long variable loop, the discriminator base G73, and the G1-C72 base pair (bp) of tRNA^{Sec} are required for efficient serylation by SerRS. In contrast, the unusual 6-bp/4-nucleotide D-arm and the 13-bp acceptor branch of tRNA^{Sec} are important for the action of SelA. Finally, several bases within tRNA^{Sec} serve as antideterminants that prevent binding by EF-Tu, while facilitating the interaction with SelB (Fig. 1).

We recently identified diverse groups of bacteria that encode tRNAs which structurally resemble tRNA^{Sec} but lack some tRNA^{Sec} identity elements. Instead, these tRNA^{Sec}-like species contain identity elements characteristic for aminoacylation by cysteinyl-tRNA synthetase (CysRS), including U73 and a GCA anticodon (Fig. 1) [5]. Thus, they were named SelC* tRNA^{Cys} after the *selC* gene that encodes tRNA^{Sec}. In

Figure 1. Proposed model of dual recoding of UGA with Sec or Cys in bacteria. Canonical recoding of UGA with Sec involves aminoacylation of tRNA^{Sec} with Ser by SerRS. Ser-tRNA^{Sec} is converted to Sec-tRNA^{Sec} by Sella, which uses selenophosphate synthesized by SelD as a substrate. Elongation factor SelB binds Sec-tRNA^{Sec} and delivers it to the ribosome where SECIS directs the incorporation of Sec. tRNA^{ReC} (formerly SelC* tRNA^{Cys}) is cysteinylated by CysRS, and Cys-tRNA^{ReC} is delivered to the ribosome by SelB as directed by SECIS, and inserted into proteins by UGA.

the *Deltaproteobacteria* clade, *selC** genes have evolved together with dedicated translation factors. For instance, in some lineages, *selC** genes are encoded upstream of a gene that is homologous to SelB, which we named *selB**. In other lineages of *Deltaproteobacteria*, organisms encode more complex systems that include a SelC* tRNA^{Cys}_{UCA} (with a UCA anticodon), a *selB** gene, and a second CysRS copy (*cysS**). For example, *Desulfococcus biacutus* has a SelC* tRNA^{Cys}_{UCA} and a unique A1-U72 bp (the beginning of the acceptor stem), located upstream of a *selB** and a *cysS** gene. *CysS** encodes a

non-canonical CysRS homolog that lacks an anticodon binding domain and contains mutations in the CP1 domain which recognize the G1-U72 bp of tRNA^{Cys}. With their distinct structural features (putative antideterminants against EF-Tu, and a long variable arm) SelC* tRNA^{Cys}_{UCA} species may potentially, in the presence of SelB*, recode Sec UGA codons to Cys. Indeed, we demonstrated earlier [5] that the *D. biacutus* SelC* tRNA^{Cys}_{UCA} is only cysteinylated by a CysRS variant that lacks the anticodon binding domain and has the *cysS** active site; the resulting Cys-tRNA inserted Cys at a UGA codon in the selenoprotein reporter *Desulfomonile tiedjei selD* gene in *E. coli*. However, it was unclear which elongation factor, EF-Tu or SelB, delivered this Cys-tRNA to the ribosome [5].

Although our characterization of SelC* tRNA^{Cys}_{UCA} suggested that this tRNA recodes UGA codons in its host organism, it is unknown whether the more widely distributed SelC* tRNA^{Cys}_{GCA} is capable of decoding UGA with Cys, because the A-G pairing at the wobble position of a codon is unfavorable in canonical translation. Nonetheless, since a pioneering study has shown that a tRNA^{Sec} anticodon mutation (UCA→GCA) permitted 30% recoding in *E. coli* [6], we hypothesize that SelC* tRNA^{Cys}_{GCA} can recode Sec UGA codons with Cys in the presence of SelB or SelB*. In the present work, we investigate the function of a SelC* tRNA^{Cys}_{GCA} from the clostridial *Desulfotomaculum nigrificans* using *in vivo* and *in vitro* approaches in *E. coli*.

2. Results

2.1. tRNA^{Sec}-like tRNA^{Cys} translates Sec UGA codons

To test the functionality of *Desulfotomaculum nigrificans* SelC* tRNA^{Cys}_{GCA} (Fig. 2A) *in vivo*, we used our previously developed *E. coli* ME6 strain [4], in which the genes encoding

Figure 2. Functional characterization of *D. nigrificans* tRNA^{ReC}. (A) Predicted cloverleaf structures of *D. nigrificans* canonical tRNA^{Cys}, SelC* tRNA^{Cys} (renamed tRNA^{ReC}), and canonical tRNA^{Sec}. (B) *fdhF* reporter assay. *E. coli* (Ec) or *D. nigrificans* (Dn) SelB was co-expressed with wild-type (WT) or opal anticodon (G34U variant) *D. nigrificans* tRNA^{ReC} in *E. coli* ME6 cells expressing *fdhF*. *In vitro* aminoacylation assays of *in vitro* transcribed Ec tRNA^{Cys}, wild-type (WT) and opal anticodon (G34U variant) *D. nigrificans* tRNA^{ReC} by *E. coli* CysRS (C) and *D. nigrificans* (Dn) CysRS (D).

SelA (*selA*), SelB (*selB*), tRNA^{Sec} (*selC*), and the selenoprotein formate dehydrogenase H (*fdhF*) were deleted. We introduced *D. nigrificans* SelC* tRNA^{Cys}_{GCA} into ME6 cells in the presence or absence of *E. coli* or *D. nigrificans* SelB, and then monitored the plasmid-based expression of *fdhF* by a traditional color assay [7]. Only full-length FDH_H with Sec or Cys in the active site (position 140) [8,9], can reduce benzyl viologen to a purple dye [7], although the Cys-containing FDH_H variant is 200-fold less active than the wildtype [9]. Remarkably, FDH_H was only expressed when SelB was co-expressed with *D. nigrificans* SelC* tRNA^{Cys}_{GCA} (Fig. 2B), suggesting that UGA is effectively recoded by the tRNA's GCA anticodon. UGA recoding by SelC* tRNA^{Cys}_{GCA} was only mediated in the presence of the SECIS element (data not shown). Interestingly, despite high sequence similarity between *E. coli* and *D. nigrificans* SelB, the latter was found to be more efficient in delivering *D. nigrificans* SelC* tRNA^{Cys}_{GCA} to the ribosome. Because ME6 cells lack the Sec-tRNA biosynthetic machinery, the UGA codon in FDH_H was likely translated as Cys, which results from the cysteinylolation of *D. nigrificans* SelC* tRNA^{Cys} by endogenous CysRS. *E. coli* CysRS uses G34 of tRNA^{Cys} as an identity element for efficient and accurate aminoacylation [10,11]. Thus, the G34→U mutation in SelC* tRNA^{Cys} prevented cysteinylolation by *E. coli* CysRS, leading to inhibited FDH_H expression (Fig. 2B). Due to the recoding functionality of *D. nigrificans* SelC* tRNA^{Cys}_{GCA}, we renamed it tRNA^{Rec}.

2.2. CysRS aminoacylates *D. nigrificans* tRNA^{Rec} in vitro

To confirm that *D. nigrificans* tRNA^{Rec} is a substrate of CysRS, we purified recombinant *E. coli* and *D. nigrificans* CysRS and tested their activity with *in vitro* transcribed *D. nigrificans* tRNA^{Rec}. Surprisingly, despite major structural differences between *E. coli* tRNA^{Cys} and tRNA^{Rec}, the *E. coli* CysRS acylated both tRNAs efficiently (Fig. 2C). A similar result was observed with *D. nigrificans* CysRS (Fig. 2D). In contrast, tRNA^{Rec}_{opal} (a G34U variant) was poorly aminoacylated compared to wild-type tRNA^{Rec} by both CysRS enzymes (Fig. 2C, D). These results imply that tRNA^{Rec} may have evolved to recode Sec UGA codons with Cys.

2.3. Phylogenetic distribution of tRNA^{Rec}

We updated our analysis of the distribution of tRNA^{Rec} genes in Bacteria to include metagenome-assembled genomes that greatly expanded the tree of life in the last two years [12–15]. tRNA^{Rec} genes are present in at least six major bacterial groups (Table 1); *Firmicutes* (including *Clostridia* subgroups and *Sulfobacillus*) (Fig. 3), a hot spring clade of *Fervidibacteria* (or *Armatimonadetes* [13]), a hot spring strain of *Chloroflexi* (probably belonging to *Dehalococcoidia*), a groundwater clade of *Actinobacteria*, *Nitrospirae* in diverse environments (Fig. 4), and *Thermodesulfobacteria* and subgroups of *Deltaproteobacteria* including *Candidatus* Desulfococcus auxilii (Fig. 5). In some cases, multiple copies of tRNA^{Rec} genes were detected. For example, three copies of tRNA^{Rec} are found in *Desulfotomaculum gibsoniae* (one of which has the rare Cys ACA anticodon), *Desulfotomaculum geothermicum* has two copies, and *Syntrophaceae* bacterium CG2_30_49_12 and *Smithella* sp. SCADC encode two tRNA^{Rec} genes within close proximity. Some organisms (e.g., *Desulfococcus multivorans* and *Carboxydotherrmus hydrogenoformans*) encode a tRNA^{Rec} pseudogene with potentially detrimental mutations [5]. All organisms encoding tRNA^{Rec} genes also encode the complete Sec-encoding machinery (*selABCD*) and canonical tRNA^{Cys} genes.

The sequences of many tRNA^{Rec} species are reasonably similar despite the diversity of their bacterial hosts, which suggests the possibility of horizontal gene transfer events. For instance, the tRNA^{Rec} gene of *Actinobacteria* is associated with *Deltaproteobacteria*-like genes. On the other hand, the structural diversity of tRNA^{Rec} implies convergent evolution of tRNA^{Rec} from tRNA^{Sec} in some bacterial lineages. For example, the *Chloroflexi* tRNA^{Rec} species is highly similar to a rare type of *Dehalococcoidia* tRNA^{Sec}, which carries a bulged pyrimidine at position 51a (Fig. 4) [4]. The uncommon base 7a in some tRNA^{Rec} species was found in a few tRNA^{Sec} species [16], while other tRNA^{Sec} species lack base 9 (Fig. 5) [4]. It is also possible that some tRNA^{Sec} genes may have evolved from tRNA^{Rec}. As previously reported, some of *Nitrospirae* and *Deltaproteobacteria* tRNA^{Rec} genes are genetically linked to a *selB* gene or a *selB* homolog gene [5]. *Nitrospirae* tRNA^{Rec} genes are followed by a typical *Nitrospirae*-type *selB* gene (Fig. 4), while the *Deltaproteobacteria* tRNA^{Rec} genes are coupled with *selB*

Table 1. Distribution of tRNA^{Rec} genes in bacteria. Members of six bacterial groups have up to three copies of tRNA^{Rec} (*selC** tRNA^{Cys}) genes in addition to canonical tRNA^{Cys} genes and a complete set of selenocysteine encoding machinery (*selABCD*). ^aThe Dissulfuribacter group belongs to genus of an unclassified order/family.

Phylum	Class	Order	Family
Firmicutes	Clostridia	Clostridiales	Peptococcaceae
Firmicutes	Clostridia	Clostridiales	Clostridiales Family XVI
Firmicutes	Clostridia	Clostridiales	Syntrophomonadaceae
Firmicutes	Clostridia	Thermoanaerobacteriales	Thermoanaerobacteraceae
Firmicutes	Bacilli	Bacillales	Alicyclobacillaceae
Fervidibacteria or Armatimonadetes			
Chloroflexi			
Actinobacteria			
Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae
Thermodesulfobacteria	Thermodesulfobacteria	Thermodesulfobacteriales	Thermodesulfobacteriaceae
Proteobacteria	Deltaproteobacteria	Desulfococcoidales	Desulfococcoidaceae
Proteobacteria	Deltaproteobacteria	Syntrophobacteriales	Syntrophaceae
Proteobacteria	Deltaproteobacteria	Syntrophobacteriales	Syntrophobacteraceae
Proteobacteria	Deltaproteobacteria	Desulfobacteriales	Desulfobacteraceae
Proteobacteria	Deltaproteobacteria	Unclassified order	Dissulfuribacter group ^a

Figure 3. Predicted cloverleaf structures of representative tRNA^{ReC} species from *Firmicutes*. Unique features are shown in red.

Figure 4. Predicted cloverleaf structures of representative tRNA^{ReC} species from *Nitrospirae*, *Actinobacteria*, *Chloroflexi* and *Feravidibacteria/Armatinomadetes*. Unique features are indicated in red. SelB next to tRNAs specifies a subgroup of *Nitrospirae* species that encode tRNA^{ReC} upstream of *Nitrospirae*-type selB.

Figure 5. Predicted cloverleaf structures of representative tRNA^{ReC} species from *Thermodesulfobacteria*, *Candidatus Desulfococcus auxilii* and *Deltaproteobacteria*. Unique features are indicated in red. The tRNA^{ReC} gene of *Deltaproteobacteria* subgroups is upstream of a *selB** gene encoding SelB*. As indicated, tRNA^{ReC} genes with an A1-U72 pair and UCA anticodon is encoded upstream a *selB** gene and a *cysS** gene encoding CysRS*.

homolog gene *selB** (Fig. 5). The *selB** gene is directly upstream a *cysS** gene encoding a CysRS homolog when the tRNA^{ReC} species has uncommon A1-U72 pair and a UCA anticodon [5] (Fig. 5). The occurrence of the $\text{tRNA}^{\text{ReC}}_{\text{UCA}}$ is limited within the *Syntrophobacteriales* and *Desulfobacteriales* groups of *Deltaproteobacteria* (Fig. 5). Therefore, it is likely that *Deltaproteobacteria* may have evolved the most intricate system among tRNA^{ReC} -utilizing organisms.

3. tRNA^{ReC} variant genes with a partial Sec identity

Our search of tRNA^{ReC} sequences revealed a novel opal suppressor tRNA species in uncultured *Pelotomaculum* organisms that also encode a canonical tRNA^{Sec} (Fig. 6A-C). These tRNA genes showed synteny and sequence similarity with the *P. thermopropionicum* tRNA^{ReC} indicating their possible evolution from a tRNA^{ReC} gene within a specific *Pelotomaculum* lineage. The predicted secondary structure of these novel tRNAs reveals interesting features, including a bulged base in the T-stem, A73 (instead of U73

of tRNA^{ReC} or G73 of tRNA^{Sec}), and a short D-arm consisting of a 5-bp stem and a 3-nucleotide loop (Fig. 6A-C). We hypothesized that this novel family of tRNA^{ReC} species may be substrates for SerRS, Sela, and SelB, instead of CysRS; thus, they may act as a non-canonical tRNA^{Sec} *in vivo*. To test this hypothesis, we used *E. coli* ME6 cells (ΔselABC) and *E. coli* FDH_H as described above. As predicted, expression of one of the tRNA^{ReC} -like tRNA species together with Sela and SelB led to functional FDH_H (Fig. 6D). However, translation in the presence of species in uncultured *Pelotomaculum* two of the tRNA^{ReC} -like tRNAs hardly expressed functional FDH_H, which may be due to the incompatibility of the tRNAs with the *E. coli* translation system. Based on this observation, we named them ' tRNA^{ReU} ' (U denotes Sec). Intriguingly, the UGA-recoding efficiency of tRNA^{ReU} was substantially lower relative to the canonical tRNA^{Sec} from same lineage of *Pelotomaculum* (Fig. 6D). This may be due to the unusual structure of tRNA^{ReU} that may render it a poor substrate for the Sec-incorporation machinery. These results suggest

Figure 6. tRNA^{ReU} species in *Pelotomaculum* spp. (A) *P. thermopropionicum* has a canonical tRNA^{Sec} gene and a canonical tRNA^{ReC} gene. (B) Uncultured *Pelotomaculum* species have a novel tRNA (named tRNA^{ReU}) which shows a good synteny and sequence similarity to the tRNA^{ReC} gene of *P. thermopropionicum*. (C) A few composite genomes of uncultured *Pelotomaculum* species encode a canonical tRNA^{Sec} gene and a tRNA^{ReU} gene. (D) Few tRNA^{ReU} species complemented the loss of tRNA^{Sec} in *E. coli* and expressed functional FDH_H selenoprotein.

that the redundancy of Sec-inserting tRNAs (i.e. tRNA^{Sec} and tRNA^{ReU}) exists to ensure proper selenoprotein expression. Although there was no precedent for tRNA^{Sec} genes with A73 [16], we identified tRNA^{Sec} genes with A73 in all *Sulfurimonas* species (they also encode tRNA^{Ser} with A73), *Thermicanus aegyptius* (tRNA^{Ser} isoacceptors with either A73 or G73), a few *Methylophil* spp., and some metagenomic α -proteobacteria. On the other hand, whereas all known tRNA^{Sec} species have a 4-nucleotide D-loop [16], canonical tRNA^{Sec} species in a few bacterial lineages have a 5-bp D-stem [4].

In addition to the *Pelotomaculum* tRNA^{ReU}, we also found sporadic occurrences of non-canonical variants of tRNA^{ReC} in *Clostridia* and *Deltaproteobacteria*. Because the amino acid identity of these non-canonical tRNA^{ReC} variants remains unknown, we named them 'tRNA^{ReX}'. Interestingly, *Desulfotomaculum gibsoniae* DSM 7213 encodes tRNA^{ReX} and two copies of tRNA^{ReC} genes. In contrast, Syntrophobacteraceae bacterium CG2_30_61_12 has a tRNA^{ReX}_{UCA} gene with A73 and is encoded upstream of a *selB** gene. This indicates that the amino acid identity of tRNA^{ReC} and tRNA^{ReX} species may not be fixed (to Cys or Sec) and could possibly have changed easily. The sporadic evolution of UGA suppressors, tRNA^{ReC}_{UCA}, tRNA^{ReX}_{UCA} and tRNA^{ReU}_{UCA}, from tRNA^{ReC}_{GCA} further

supports our hypothesis that tRNA^{ReC}_{GCA} is dedicated to UGA recoding and is not involved in the translation of the Cys codons.

4. Discussion

Due to the chemical similarity of Sec and Cys, they are often interchangeable and sometimes even encoded by the same codon in many redox enzyme and protein genes. In certain bacteria, Sec is read by the UGU/UGC Cys codons, which are decoded by tRNA^{Sec} species with complementary anticodons [4]. In the ciliate *Euplotes crassus*, UGA specifies the insertion of both Cys and Sec [17]. In mammals, deficiency of the micronutrient selenium causes mis-incorporation of Cys into Sec positions due to the erroneous formation of Cys-tRNA^{Sec} [18–20]. However, in bacteria, sulfur is strictly discriminated against during the synthesis of Sec-tRNA^{Sec}, preventing the formation of Cys-tRNA^{Sec} [21]. Consequently, tRNA^{ReC} genes in bacteria may have originated as an alternative mechanism to recode UGA with Cys. In archaea, near-cognate UGA suppression by Cys-tRNA^{Cys}_{GCA} was observed in selenoprotein translation when tRNA^{Sec} was deleted [22], while *E. coli* tolerates wobble decoding of UGG by Sec, a process that requires the presence of the SECIS element [23]. Therefore, although

many organisms encode multiple copies of the same gene with either Cys and Sec codons at the same position [24], codon sharing by Sec and Cys may also be a universal strategy in diverse Sec-utilizing organisms to rapidly adapt to different environmental conditions.

Additionally, tRNA^{ReC}/tRNA^{ReU} may also function as a molecular mimic of tRNA^{Sec} to maintain homeostasis between Sec-tRNA^{Sec} and SelB. Due to the high affinity of SelB for Sec-tRNA^{Sec}, overexpression of SelB can sequester Sec-tRNA^{Sec} [25]. Constitutive expression of tRNA^{ReC}/tRNA^{ReU} might mitigate the detrimental effect of tRNA^{Sec}-sequestration without strictly regulating SelB expression, conferring robustness to the Sec-encoding system. Alternatively, tRNA^{ReC}_{GCA} may translate Cys codons with Cys in selenoprotein genes in which a Sec UGA has changed to either a Cys UGU or UGC codon but have retained a SECIS element [24]. A Sec-to-Cys codon change can occur by a point mutation, but complete inactivation of a hairpin-like SECIS element requires accumulation of several mutations. Thus, tRNA^{ReC}_{GCA} may facilitate evolutionary transition over generations to allow the elimination of the SECIS element.

A previous bioinformatic study found that bacteria growing at high temperatures and low oxygen levels often have the Sec translation machinery [24]. Interestingly, we found that tRNA^{ReC/ReU}-encoding bacteria are generally thermophilic or mesothermophilic sulfate-reducing bacteria. Based on this observation, we hypothesize that tRNA^{ReC}-encoding thermophilic sulfate-reducing bacteria produce selenoproteins containing Cys instead of Sec to easily cope with oxygen toxicity and/or selenium deficiency in their niches in order to be metabolically productive. Such dynamic recoding of Sec codons with Cys may be essential for expression of selenoproteins that regulate the synthesis of other selenoproteins (e.g. selenophosphate synthase, SelD). A similar mechanism is observed in higher eukaryotes, including humans, which rely on dozens of selenoproteins. The animal Sec-encoding machinery misincorporates Cys to produce Cys-containing variants of 'selenoproteins' in a selenium-deficient diet [18]. Together, this suggests that organisms with a large selenoproteome may need an alternative mechanism to maintain production of selenoproteins under changing conditions.

Translational recoding is more dynamic and unpredictable than anticipated. Our results indicate that recoding of the Sec UGA codon is a dynamic process that involves additional factors whose functions remain to be elucidated. Nonetheless, our data provide evidence that mispairing of the UGA codon with the GCA anticodon is programmed by the genetic code, and mediated by elongation factor SelB. Critical analysis of genomics, proteomics, and tRNAomics should reveal the extent of this programmed genetic code ambiguity.

5. Materials and methods

5.1. Plasmid construction

The *D. nigrificans* CysS gene sequence was optimized for expression in *E. coli* and synthesized by Quintara Biosciences. The gene was cloned into pET20b. Complementary strands for the tRNA

genes with an upstream T7 promoter sequence were synthesized by Keck Oligonucleotide Synthesis at Yale University. The strands were annealed, and the resulting DNA fragment was cloned into pUC18. The *E. coli* selB and *D. nigrificans* selB (codon-optimized for expression in *E. coli*) genes were cloned under the arabinose promoter in pBAD-CDF (*spec*^R) [26]. tRNA^{ReC} sequences were cloned under the *lpp* promoter in pGFIB. *Pelotomaculum* tRNA^{ReU} and tRNA^{Sec} sequences were cloned under the arabinose promoter in pBAD-RSF5 (*kan*^R) [5]. All cloning was performed using DNA ligase (NEB), Gibson Assembly (New England Biolabs) or In-Fusion cloning (Clontech).

5.2. FDH_H assay for tRNA^{ReC} and tRNA^{ReU}

In vivo assays for tRNA^{ReC} were carried out in ME6 cells [4] harboring a pRSFDuet-1 plasmid [27] encoding *E. coli* serS and *fdhF*, a pBAD-CDF plasmid encoding either *E. coli* or *D. nigrificans* selB, and a pGFIB plasmid expressing *D. nigrificans* tRNA^{ReC} under an *lpp* promoter. The FDH_H assay was performed as previously described with 0.2% arabinose added to the agar plate [4]. To test the function of tRNA^{ReU} species, ME6 cells harboring pACYC184-*fdhF* and pCDFDuet-1-*EcselAB* plasmids (designated as *E. coli* Δ*selC*) were transformed with a pBAD-RSF5 plasmid [5] expressing tRNA^{ReU/Sec} using an arabinose promoter. Overnight cultures were spotted onto LB agar plates supplemented with 1 μM Na₂MoO₄, 1 μM Na₂SeO₃, 50 mM sodium formate, 10 μM isopropyl β-D-1-thiogalactopyranoside (IPTG), 0.1% arabinose, and the appropriate antibiotics and grown anaerobically at 37°C overnight. The plates were then anaerobically overlaid with a top agar containing 1 mg/mL benzyl viologen, 250 mM sodium formate, and 25 mM KH₂PO₄ (pH 7.0).

5.3. Protein purification

An ASKA collection strain harboring a plasmid for expression of the His-tagged *E. coli* CysS gene was used to purify *E. coli* CysRS [28]. Cells were grown to an A₆₀₀ of 0.5 at 37°C followed by protein overexpression with 1 mM IPTG for 4 h at 37°C. Cells were harvested by centrifugation. The cell pellet was resuspended in buffer containing 50 mM Tris (pH 7.5), 300 mM NaCl, and protease inhibitor cocktail tablets (cOmplete, Roche), followed by sonication. Lysed cells were centrifuged for 45 min at 4°C at 18,000 x g. The lysate was run through a TALON metal affinity resin (Clontech), and the protein was eluted with a linear imidazole gradient. The protein was stored at -20°C in 50 mM HEPES (pH 7.3), 50 mM KCl, 15 mM MgCl₂, 5 mM dithiothreitol, and 40% glycerol. Recombinant *D. nigrificans* CysRS was purified from *E. coli* BL21(DE3) cells using the pET20b plasmid containing the *D. nigrificans* CysS gene. Cells were grown to an A₆₀₀ of 0.6 and protein overexpression was induced with 0.1 mM IPTG overnight at 20°C. *D. nigrificans* CysRS was extracted and purified as described for *E. coli* CysRS. An *E. coli* strain from the ASKA collection was used to express His-tagged *E. coli* tRNA nucleotidyltransferase [28]. Cells were grown to an A₆₀₀ of 0.6 at 37°C followed by protein overexpression with 0.1 mM IPTG overnight at 20°C. Cells were harvested by centrifugation and resuspended with 20 mM Tris-HCl (pH 7.5), 0.5 M NaCl, 5 mM imidazole and

protease inhibitor cocktail tablets. Cells were lysed with 10 mg/mL lysozyme and sonication. *E. coli* tRNA nucleotidyltransferase was extracted and purified as described for *E. coli* CysRS and stored in 20 mM Tris-HCl (pH 7.5), 0.5 M NaCl, 5 mM MgCl₂, and 40% glycerol. T7 RNA polymerase was purified as previously described [29].

5.4. *In vitro* aminoacylation

100 µg of pUC18.EctRNA^{Cys} and pUC18.DntRNA^{ReC} were digested with New England Biolabs BstNI (overnight at 55°C) and NsiI-HF (5 h at 37°C), respectively. The tRNAs were transcribed *in vitro* using T7 RNA polymerase as previously described [30]. tRNAs were labeled at 3'-end adenosine with [³²P]-phosphate using *E. coli* tRNA nucleotidyltransferase and [α-³²P]-ATP [31]. Aminoacylation reactions were carried with 2 mM cysteine and 5 µM tRNA (with trace amount of [³²P]-labeled tRNA) in buffer containing 50 mM HEPES (pH 7.3), 50 mM KCl, 10 mM MgCl, 1 mM dithiothreitol, and 10 mM ATP. Reactions were initiated by adding 1 µM of CysRS. Aminoacylation was monitored by quenching 2.5 µL of the reaction mixture with 5 µL of a solution containing 0.1 U/µL P1 nuclease (MilliporeSigma) in 200 mM sodium acetate (pH 4.5) at the indicated times. The quenched solution was incubated at room temperature for 1 h, and 1 µL of solution was spotted on a PEI cellulose TLC (MilliporeSigma). Separation of [³²P]-AMP and Cys[³²P]-AMP was carried out in 100 mM sodium acetate and 5% acetic acid.

5.5. Bioinformatics analysis

Most of the tRNA^{ReC} sequences were obtained in our previous studies [4,5]. We updated our tRNA^{ReC} sequence dataset by repeatedly performing BLASTn searches of genome and metagenome data in the Integrated Microbial Genomes (IMG) system and the National Center for Biotechnology Information (NCBI). The cloverleaf structures of tRNA^{ReC/ReU} species were predicted as previously described [4,5,16,32].

Acknowledgments

We thank H. James Tripp and Natalia N. Ivanova (Joint Genome Institute) for supporting bioinformatics analysis, and Noah M. Reynolds and Anastasia Sevostyanova (Yale University) for technical support. T.M. is a Japan Society for the Promotion of Science postdoctoral fellow for research abroad. We thank Robert Kelly, Sean Caffrey, Matthew Schrenk, Ramunas Stepanauskas, Sean P. Jungbluth, Michael S. Rappé, Philip Hugenholtz, Gene W. Tyson, Jillian F. Banfield and many others for permission to use sequence data deposited in the DOE-JGI and NCBI databases.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Institute of General Medical Sciences [R35GM122560 and R01GM22854] and the U.S. Department of Energy [DE-FG02-98ER-20311].

ORCID

Oscar Vargas-Rodriguez <http://orcid.org/0000-0002-2301-2800>
Dieter Söll <http://orcid.org/0000-0002-3077-8986>

References

- Labunskyy VM, Hatfield DL, Gladyshev VN. Selenoproteins: molecular pathways and physiological roles. *Physiol Rev*. 2014;94:739–777.
- Fischer N, Neumann P, Böck LV, et al. The pathway to GTPase activation of elongation factor SelB on the ribosome. *Nature*. 2016;540:80–85.
- Rodnina MV, Fischer N, Maracci C, et al. Ribosome dynamics during decoding. *Philos Trans R Soc Lond B Biol Sci*. 2017;372. DOI:10.1098/rstb.2016.0182
- Mukai T, Englert M, Tripp HJ, et al. Facile recoding of selenocysteine in nature. *Angew Chem Int Ed Engl*. 2016;55:5337–5341.
- Mukai T, Vargas-Rodriguez O, Englert M, et al. Transfer RNAs with novel cloverleaf structures. *Nucleic Acids Res*. 2017;45:2776–2785.
- Baron C, Heider J, Böck A. Mutagenesis of selC, the gene for the selenocysteine-inserting tRNA-species in *E. coli*: effects on *in vivo* function. *Nucleic Acids Res*. 1990;18:6761–6766.
- Lacourciere GM, Levine RL, Stadtman TC. Direct detection of potential selenium delivery proteins by using an *Escherichia coli* strain unable to incorporate selenium from selenite into proteins. *Proc Natl Acad Sci U S A*. 2002;99:9150–9153.
- Zinoni F, Birkmann A, Leinfelder W, et al. Cotranslational insertion of selenocysteine into formate dehydrogenase from *Escherichia coli* directed by a UGA codon. *Proc Natl Acad Sci U S A*. 1987;84:3156–3160.
- Axley MJ, Böck A, Stadtman TC. Catalytic properties of an *Escherichia coli* formate dehydrogenase mutant in which sulfur replaces selenium. *Proc Natl Acad Sci U S A*. 1991;88:8450–8454.
- Pallanck L, Li S, Schulman LH. The anticodon and discriminator base are major determinants of cysteine tRNA identity *in vivo*. *J Biol Chem*. 1992;267:7221–7223.
- Komatsoulis GA, Abelson J. Recognition of tRNA^{Cys} by *Escherichia coli* cysteinyl-tRNA synthetase. *Biochemistry*. 1993;32:7435–7444.
- Bowers RM, Kyrpides NC, Stepanauskas R, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol*. 2017;35:725–731.
- Eloe-Fadros EA, Paez-Espino D, Jarett J, et al. Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. *Nat Commun*. 2016;7:10476
- Jungbluth SP, Amend JP, Rappe MS. Metagenome sequencing and 98 microbial genomes from Juan de Fuca Ridge flank subsurface fluids. *Sci Data*. 2017;4:170037.
- Parks DH, Rinke C, Chuvochina M, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol*. 2017;2:1533–1542.
- Santesmasses D, Mariotti M, Guigo R. Computational identification of the selenocysteine tRNA (tRNA^{Sec}) in genomes. *PLoS Comput Biol*. 2017;13:e1005383.
- Turanov AA, Lobanov AV, Fomenko DE, et al. Genetic code supports targeted insertion of two amino acids by one codon. *Science*. 2009;323:259–261.
- Xu XM, Turanov AA, Carlson BA, et al. Targeted insertion of cysteine by decoding UGA codons with mammalian selenocysteine machinery. *Proc Natl Acad Sci U S A*. 2010;107:21430–21434.
- Turanov AA, Everley RA, Hybsier S, et al. Regulation of selenocysteine content of human selenoprotein P by dietary selenium and Insertion of cysteine in place of selenocysteine. *PLoS One*. 2015;10:e0140353.

20. Gladyshev VN, Carlson BA, Hatfield DL. Pathways in de novo biosynthesis of selenocysteine and cysteine in eukaryotes. *Selenium*: Springer; 2016. p. 39–45.
21. Tormay P, Wilting R, Lottspeich F, et al. Bacterial selenocysteine synthase—structural and functional properties. *Eur J Biochem*. 1998;254:655–661.
22. Seyhan D, Jehmlich N, Von Bergen M, et al. Selenocysteine-independent suppression of UGA codons in the archaeon *Methanococcus maripaludis*. *Biochim Biophys Acta*. 2015;1850:2385–2392.
23. Xu J, Croitoru V, Rutishauser D, et al. Wobble decoding by the *Escherichia coli* selenocysteine insertion machinery. *Nucleic Acids Res*. 2013;41:9800–9811.
24. Zhang Y, Romero H, Salinas G, et al. Dynamic evolution of selenocysteine utilization in bacteria: a balance between selenoprotein loss and evolution of selenocysteine from redox active cysteine residues. *Genome Biol*. 2006;7:R94.
25. Xu YW, Jiang ZH, Mu Y, et al. Effects of combinatorial expression of selA, selB and selC genes on the efficiency of selenocysteine incorporation in *Escherichia coli*. *Chem Res Chin Univ*. 2013;29:87–94.
26. Haruna K, Alkazemi MH, Liu Y, et al. Engineering the elongation factor Tu for efficient selenoprotein synthesis. *Nucleic Acids Res*. 2014;42:9976–9983.
27. Aldag C, Bröcker MJ, Hohn MJ, et al. Rewiring translation for elongation factor Tu-dependent selenocysteine incorporation. *Angew Chem Int Ed Engl*. 2013;52:1441–1445.
28. Kitagawa M, Ara T, Arifuzzaman M, et al. Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res*. 2005;12:291–299.
29. Ellinger T, Ehricht R. Single-step purification of T7 RNA polymerase with a 6-histidine tag. *Biotechniques*. 1998;24:718–720.
30. Melton DA, Krieg PA, Rebagliati MR, et al. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res*. 1984;12:7035–7056.
31. Ledoux S, Uhlenbeck OC. [3'-32P]-labeling tRNA with nucleotidyltransferase for assaying aminoacylation and peptide bond formation. *Methods*. 2008;44:74–80.
32. Itoh Y, Sekine S, Suetsugu S, et al. Tertiary structure of bacterial selenocysteine tRNA. *Nucleic Acids Res*. 2013;41:6729–6738.