

# Lawrence Berkeley National Laboratory

## Joint Genome Institute

**Title**

Facile Recoding of Selenocysteine in Nature

**Permalink**

<https://escholarship.org/uc/item/8m84f4f1>

**Journal**

Angewandte Chemie International Edition, 55(17)

**ISSN**

1433-7851

**Authors**

Mukai, Takahito  
Englert, Markus  
Tripp, H James  
[et al.](#)

**Publication Date**

2016-04-18

**DOI**

10.1002/anie.201511657

Peer reviewed



## Facile Recoding of Selenocysteine in Nature

Takahito Mukai, Markus Englert, H. James Tripp, Corwin Miller, Natalia N. Ivanova, Edward M. Rubin, Nikos C. Kyrpides, and Dieter Söll\*

**Abstract:** Selenocysteine (Sec or U) is encoded by UGA, a stop codon reassigned by a Sec-specific elongation factor and a distinctive RNA structure. To discover possible code variations in extant organisms we analyzed 6.4 trillion base pairs of metagenomic sequences and 24903 microbial genomes for tRNA<sup>Sec</sup> species. As expected, UGA is the predominant Sec codon in use. We also found tRNA<sup>Sec</sup> species that recognize the stop codons UAG and UAA, and ten sense codons. Selenoprotein synthesis programmed by UAG in *Geodermatophilus* and *Blastococcus*, and by the Cys codon UGU in *Aeromonas salmonicida* was confirmed by metabolic labeling with <sup>75</sup>Se or mass spectrometry. Other tRNA<sup>Sec</sup> species with different anticodons enabled *E. coli* to synthesize active formate dehydrogenase H, a selenoenzyme. This illustrates the ease by which the genetic code may evolve new coding schemes, possibly aiding organisms to adapt to changing environments, and show the genetic code is much more flexible than previously thought.

The micronutrient selenium is present in proteins in the form of the versatile 21st amino acid, selenocysteine, in which the thiol moiety of cysteine (Cys) is replaced by a selenol group.<sup>[1]</sup> Selenoproteins are present in organisms from all domains of life;<sup>[2]</sup> such proteins are essential in mammalian cells,<sup>[3]</sup> yet plants and fungi lack this amino acid. Sec is present in the active site of many redox enzymes.<sup>[4]</sup> The codon for Sec is UGA which is normally a translational stop signal.<sup>[5]</sup> During translation of selenoprotein mRNAs, UGA is recoded by the interaction of a specialized elongation factor SelB (in bacteria) with a downstream Sec insertion sequence.<sup>[5,6]</sup> Recently, a synthetic biology study succeeded in reassigning Sec to a large number of sense and stop codons in *Escherichia coli*<sup>[7]</sup> demonstrating that alterations to the genetic code can be tolerated. This prompted the question, whether deviations of the standard UGA Sec assignment may naturally occur.

A computational study scanning several trillion base pairs of metagenomic data revealed a large number of stop codon

reassignments in bacteria and bacteriophages.<sup>[8]</sup> This inspired us to perform a comprehensive search of the available metagenomic and microbial genomic sequence data for anticodon variants of the typical tRNA<sup>Sec</sup><sub>UCA</sub>, the longest tRNA<sup>[9]</sup> with a tertiary structure quite different from that of canonical tRNAs.<sup>[10]</sup> A BLAST search of the tRNADB-CE database<sup>[11]</sup> revealed four tRNA<sup>Sec</sup><sub>CUA</sub> sequences, suggesting UAG codon recognition. Searching all public microbial genomes in the National Center for Biotechnology Information (NCBI) and all assembled metagenome data in the Integrated Microbial Genomes (IMG) system<sup>[12]</sup> yielded a tRNA<sup>Sec</sup><sub>GCA</sub> group, implying it was capable of UGC codon recognition. We then developed a general computational pipeline that scanned approximately 6.4 Tb of unassembled short reads, around 180 Gb of assembled contigs (> 2 kb), and 24903 microbial genomes in IMG (Figure 1A). The results affirmed UGA as the predominant Sec codon. In addition, 12 different tRNA<sup>Sec</sup> anticodon variants capable of recognizing the stop codons UAG and UAA, and 10 sense codons were discovered (Figure 1A). Further sequence validations (see Supporting Information) ascertained these tRNA<sup>Sec</sup> variants not to be sequencing artifacts.

We grouped these non-canonical tRNA<sup>Sec</sup> species by anticodon type, sequence, and structural similarity (Figure 1A, and Supporting Information). The largest group (anticodon CUA) contains 366 nearly identical tRNA<sup>Sec</sup> sequences from the actinobacterial *Geodermatophilaceae* family<sup>[13]</sup> (Figure 1A,B left panel), while the other three tRNA<sup>Sec</sup><sub>CUA</sub> species may be of rhizosphere bacterial origin (Figures S1, S2). The tRNA<sup>Sec</sup><sub>CUA</sub> species amounted to 3% of the total tRNA<sup>Sec</sup> species found in a soil metagenome (3300001205).

The next group contains tRNA<sup>Sec</sup> species (with anticodon GCA able to decode Cys) from *Betaproteobacteria* and termite gut symbionts (Figure 1A, Figure S3). Two *Aeromonas salmonicida* genomes contain a tRNA<sup>Sec</sup><sub>ACA</sub>, able to decode the UGU Cys codon (Figure 1B, middle panel). The other 16 members of this group may originate from *Chloroflexi* (Figure S4) and ocean bacteria (Figure S5). The two tRNA<sup>Sec</sup><sub>UUU/CUU</sub> species (recognizing AAA/AAG Lys codons) may derive from the *Solirubrobacterales* (Figure S6). Additional tRNA<sup>Sec</sup> variants able to recognize the stop codon UAA and six sense codons (CGA, AGA, GGA, UUA, UCA, UGG) were also found (Figure S7).

We then wanted to confirm the coding properties of these non-canonical tRNA<sup>Sec</sup> species. For proof of selenoprotein synthesis three strategies were possible:

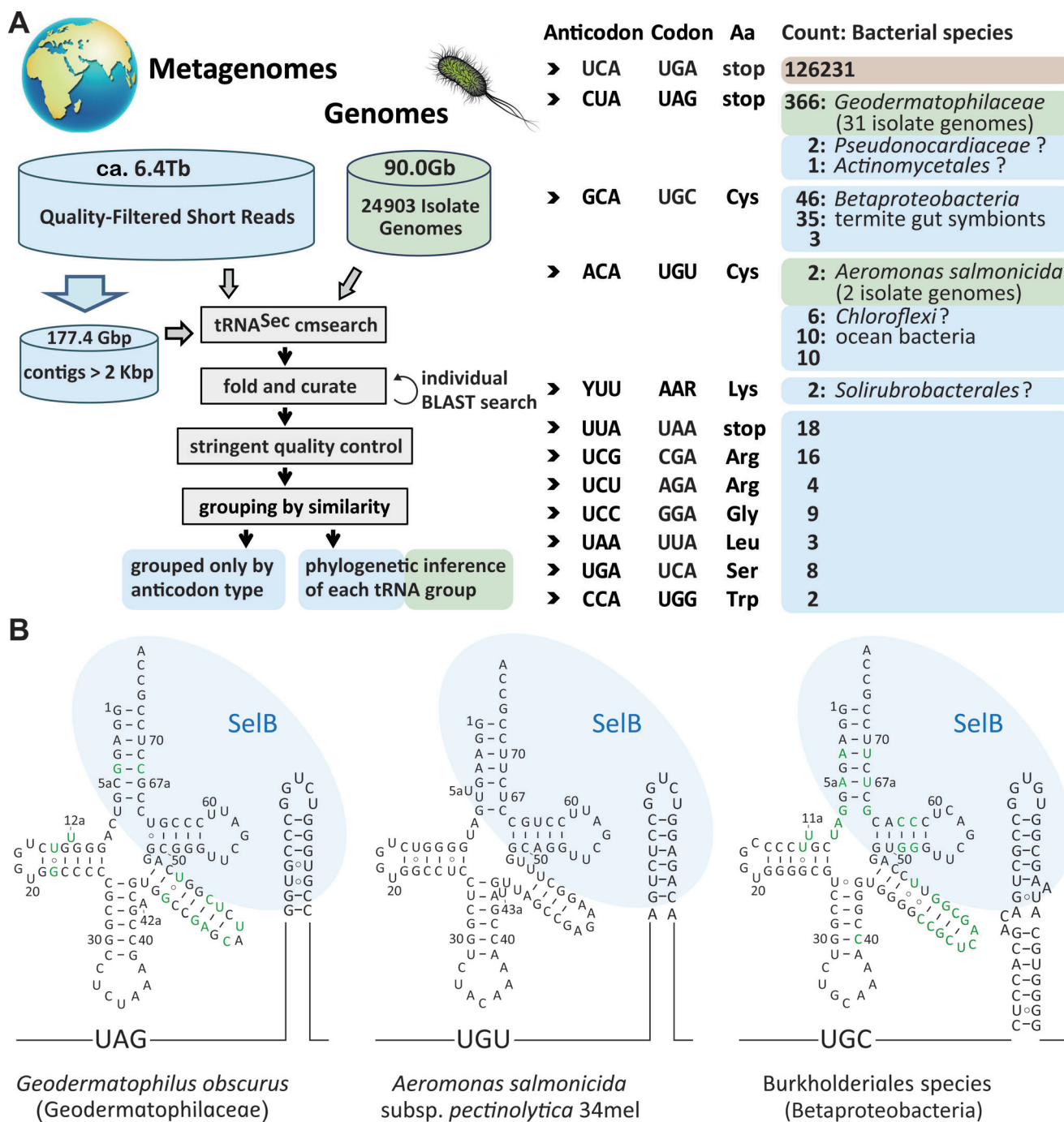
- 1) metabolic labeling of the organisms with <sup>75</sup>Se
- 2) replacing parts of the *E. coli* selenoprotein synthesis machinery with genes and tRNA<sup>Sec</sup> from our genomic or metagenomic findings

[\*] Dr. T. Mukai, Dr. M. Englert, Dr. C. Miller, Prof. D. Söll  
Department of Molecular Biophysics and Biochemistry  
Yale University  
New Haven, CT 06520 (USA)  
E-mail: dieter.soll@yale.edu

Prof. D. Söll  
Department of Chemistry, Yale University  
New Haven, CT 06520 (USA)

Dr. H. J. Tripp, Dr. N. N. Ivanova, Dr. E. M. Rubin, Dr. N. C. Kyrpides  
Department of Energy Joint Genome Institute (DOE JGI)  
Walnut Creek, CA 94598 (USA)

Supporting information for this article can be found under <http://dx.doi.org/10.1002/anie.201511657>.

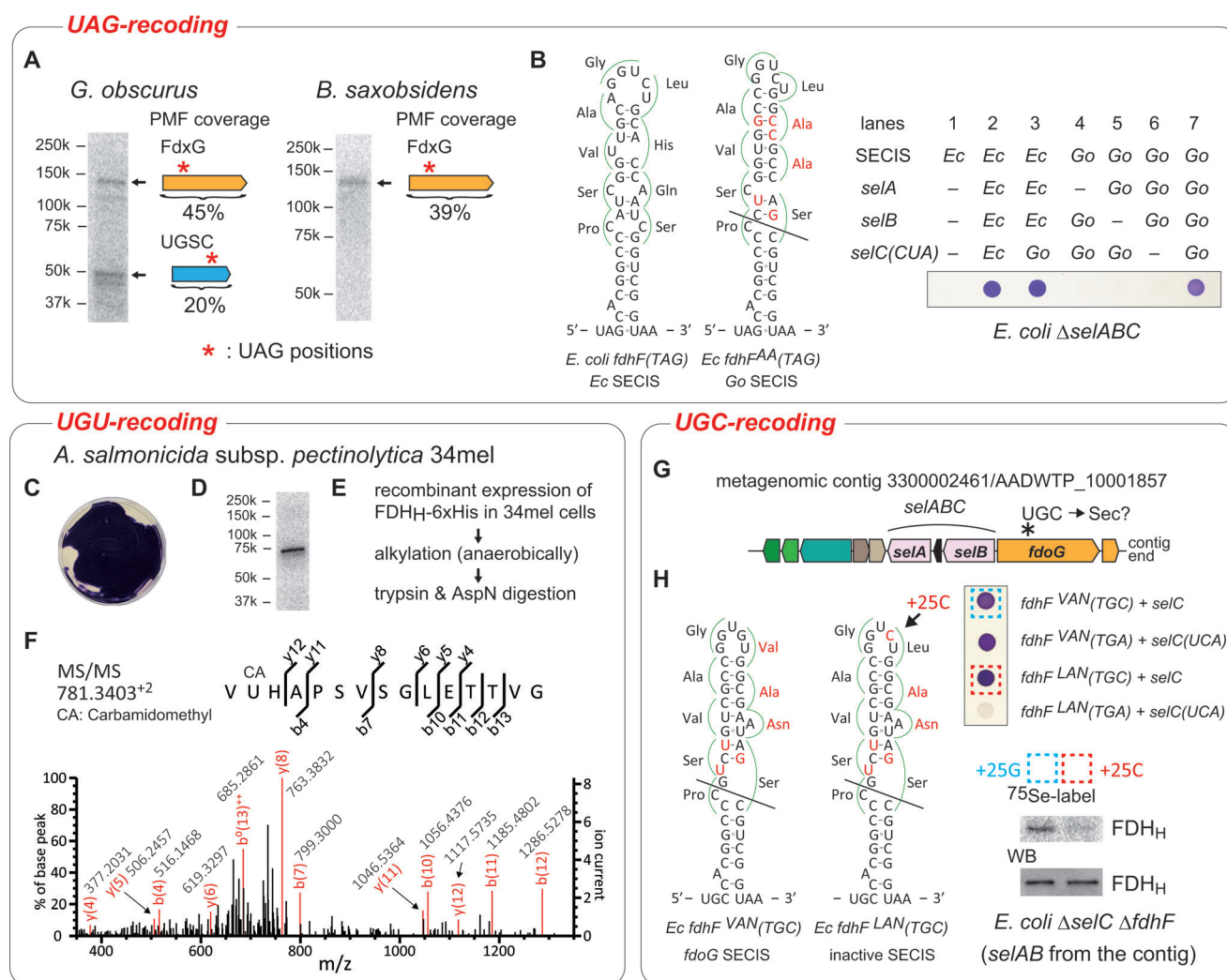


**Figure 1.** Non-canonical selenocysteine assignments in nature. A) The tRNA<sup>Sec</sup> computer-based search pipeline and the manually curated output. The non-canonical tRNA<sup>Sec</sup> sequences are grouped by codon recognition; their numbers are given with their (putative) bacterial origins from phylogenetic inference. Green indicates results from whole genomes, blue represents results from only metagenomic data. For comparison the number of canonical tRNA<sup>UCA</sup> sequences is shown (most of them were not curated). “Y” and “R” denote “C or U” and “A or G”, respectively. B) Inferred cloverleaf structures of non-canonical tRNA<sup>Sec</sup> species. The nucleotide polymorphism among the same tRNA<sup>Sec</sup> group is indicated with green letters. The Sec codons and the SECIS elements of formate dehydrogenase mRNAs are shown.

3) replacing *E. coli* tRNA<sup>Sec</sup> with the newly discovered tRNA<sup>Sec</sup> species.

In the last two strategies *E. coli* formate dehydrogenase H (FDH<sub>H</sub>, encoded by the *fdhF* gene) would serve as reporter.<sup>[7]</sup>

To confirm UAG-directed Sec incorporation we grew *Geodermatophilus obscurus* G-20 and *Blastococcus saxobidens* cells in the presence of [<sup>75</sup>Se]selenite and detected radiolabeled selenoproteins of 140 and 50 kDa size. The genome sequences predict formate dehydrogenases (FDHs) (Figure S8, Tables S1, S2) and UGSC-motif proteins,<sup>[14]</sup> (Fig-



**Figure 2.** Recoding of UAG and cysteine codons to selenocysteine. A) Metabolic <sup>75</sup>Se labeling of *G. obscurus* and *B. saxobsidens* cells. Crude extracts were resolved by SDS-PAGE, and their putative selenoproteins were visualized by PhosphorImager analysis. The results of peptide mass fingerprinting (PMF) analyses of the proteins in the excised gel bands are shown to the right of the bands. B) FDH<sub>H</sub> expression in *E. coli*  $\Delta$ selABC  $\Delta$ fdhF cells with the *G. obscurus* selABC genes and a chimeric fdhF(140TAG) gene variant having a *G. obscurus* SECIS element with a few nucleotide modifications shown in red. The selC(CUA) genes express tRNA<sup>Sec</sup><sub>CUA</sub>. The expressed selenoprotein FDH<sub>H</sub> reduced benzyl viologen, resulting in a purple color. C) FDH<sub>H</sub> activity of *A. salmonicida* subsp. *pectinolytica* 34mel cells. D) Metabolic <sup>75</sup>Se labeling of the FDH<sub>H</sub> of 34mel. E) The procedure of sample preparation for the LC-MS/MS analysis of the FDH<sub>H</sub> selenoprotein. F) PMF confirms Sec incorporation at codon 140 in the recombinant FDH<sub>H</sub>. G) tRNA<sup>Sec</sup><sub>GCA</sub> gene (selC) locus in a metagenomic contig. H) In vivo FDH<sub>H</sub> assays in *E. coli*  $\Delta$ selABC  $\Delta$ fdhF cells with the selABC genes of the metagenomic contig and a chimeric fdhF(140TGC) gene carrying the contig's SECIS element with a few nucleotide modifications shown in red. The two transformed strains boxed were metabolically labeled with <sup>75</sup>Se, and the radioactive FDH<sub>H</sub> proteins were analyzed by SDS-PAGE and autoradiography or western blotting (WB).

ures S1, S8, Table S3). This was the first indication for UAG read-through by Sec to form the expected FDH (FdxG) and UGSC-motif protein products (Figure 2A). Crude cell extracts were then resolved by SDS-PAGE; the proteins in the gel slices corresponding to 140 and 50 kDa were trypsinized for subsequent liquid chromatography (LC) coupled with tandem mass spectrometry (LC-MS/MS) analysis. The 140 kDa gel slices from *G. obscurus* and *B. saxobsidens* harbored full-length FdxG, while the 50 kDa gel slice from *G. obscurus* contained a UGSC-motif protein (Figure 2A, Table S4).

Like *E. coli*, the *G. obscurus* genome<sup>[13b]</sup> encodes a Sec incorporation machinery consisting of the selA, selB, selC, and

selD genes. In bacteria, SelD produces the Se donor selenophosphate, SelA converts Ser-tRNA<sup>Sec</sup> (the tRNA is the product of the selC gene) into Sec-tRNA<sup>Sec</sup>, and SelB carries Sec-tRNA<sup>Sec</sup> to the ribosome in a SECIS-dependent manner. To test their functionality, the *E. coli*  $\Delta$ selABC  $\Delta$ fdhF strain ME6 was complemented with the *G. obscurus* selABC genes. The product of the *E. coli* fdhF gene is the selenoenzyme FDH<sub>H</sub> (Table S1) whose activity requires Sec at position 140,<sup>[15]</sup> if replaced by Cys the activity drops 300-fold.<sup>[16]</sup> FDH<sub>H</sub> is readily detected by the reduction of benzyl viologen resulting in a purple color.<sup>[17]</sup> To serve as a reporter, the plasmid-encoded *E. coli* fdhF gene transformed into strain ME6 was modified to have a TAG codon in posi-



tion 140, followed by a *G. obscurus*-type SECIS element leading to an FDH<sub>H</sub> variant with two amino acid changes (see Figure 2B). Expression of the *G. obscurus selABC* genes in this modified strain produced active FDH<sub>H</sub> (Figure 2B). These data confirm that in this *E. coli* FDH<sub>H</sub> variant the UAG<sub>140</sub> codon is recoded to Sec.

UGU (Cys) recoding was confirmed in *Aeromonas salmonicida* subsp. *pectinolytica* 34mel, the type strain of the  $\gamma$ -proteobacterium *A. salmonicida* subspecies.<sup>[18]</sup> Unlike other *Aeromonas* species, the *pectinolytica* subspecies and strain Y577,<sup>[19]</sup> pair tRNA<sup>Sec</sup><sub>ACA</sub> with a UGU Cys codon in *fdhF*. In addition to the in vivo FDH activity in *pectinolytica* cells (Figure 2C), their anaerobic metabolic labeling with <sup>75</sup>Se produced radioactive Sec-containing FDH<sub>H</sub> (Figure 2D). We confirmed Sec incorporation encoded by UGU<sub>140</sub> in FDH<sub>H</sub> by overexpressing the protein from a plasmid (Figure 2E) and LC-MS/MS analysis (Table S4). The recoded selenopeptide (LC retention time 24.82 min) had the correct mass (Figure S9) and the appropriate secondary fragmentation pattern (Figure 2F). The co-eluting Cys- and Sec-peptides were detected through their different masses with ion intensities of approximately 100:1 (Figure S9A). The mass peaks of the Sec-peptides were absent in protein samples obtained from *pectinolytica* cells expressing a *fdhF* variant lacking the SECIS element (Table S4). Thus, the Cys<sub>140</sub> codon of the *fdhF* gene is translated as Sec in a SECIS-dependent manner in *A. pectinolytica* 34mel.

UGC Cys→Sec recoding by tRNA<sup>Sec</sup><sub>GCA</sub> may be a common trait in the *Burkholderiales*. In some long metagenomic *Burkholderiales* contig the *selC* gene is flanked by *selB* and *selA* genes, and the *selABC* operon is located next to a formate dehydrogenase (*fdoG*) gene whose active site UGC Cys codon is followed by a putative SECIS element (Figures 1B right panel and Figure 2G). The *E. coli*  $\Delta selABC$  strain was complemented with the *Burkholderiales* contig *selAB* and *selC*-opal variant, and an *E. coli* *fdhF* variant harboring the *Burkholderiales* contig SECIS element. This strain produced active FDH<sub>H</sub> (purple color; Figure 2H, the 2nd row), but an inactive SECIS element (with a G25C mutation) did not form FDH<sub>H</sub> and the cells were colorless (Figure 2H, the 4th row). In combination with the *Burkholderiales* contig *selC*, we changed UGA<sub>140</sub> to UGC for the chimeric *fdhF* variants that carried functional or inactive SECIS elements. As the FDH<sub>H</sub> Cys<sub>140</sub> enzyme produced a purple color (Figure 2H, the 1st and 3rd rows), <sup>75</sup>Se-labeling was used to demonstrate that the functional SECIS element led to a clear signal (Figure 2H). Thus, the *Burkholderiales* contig *selA*, *selB*, tRNA<sup>Sec</sup><sub>GCA</sub>, and SECIS enabled UGC-recoding in *E. coli*.

The metagenomic tRNA<sup>Sec</sup> variants that recognize other stop and sense codons were also tested for Sec reassignment. We selected one representative metagenomic tRNA<sup>Sec</sup> species for each anticodon type and expressed them in an *E. coli*  $\Delta selC$   $\Delta fdhF$  strain, together with the *E. coli* *fdhF* variants that carry the proper cognate codons at position 140.<sup>[7]</sup> Surprisingly, all but tRNA<sup>Sec</sup><sub>UCC</sub> of the tested tRNA<sup>Sec</sup> species recoded the respective codons for Sec, as they supported the expression of active FDH<sub>H</sub> in their host *E. coli* cells (Figure S10). It should be mentioned that GGA was also poorly

recoded in our earlier Sec recoding strategy.<sup>[7]</sup> The different recoding efficiencies may result from distortions of the ideal SECIS element structure by the nature of the upstream codon.<sup>[20]</sup> In light of these results we believe that these tRNA<sup>Sec</sup> species may be used for recoding sense codons in the organisms they originate from.

What about eukaryotic organisms? Although we found nine tRNA<sup>Sec</sup> variants of algal origin (two are shown in Figure S7), they need further validation, because they are almost identical to canonical tRNA<sup>Sec</sup> species. A similar search of 92 mammalian genomes (215 Gbp) and of the *Drosophila melanogaster* genome (139 Mbp) showed no exception to the use of UGA as the Sec codon. Whether this is related to the necessity of selenoproteins in high-level redox signaling pathways<sup>[21]</sup> or due to the sophisticated backup systems<sup>[22]</sup> remains to be investigated. However, in the lower eukaryote *Euplotes crassus* UGA serves both as a Cys and also as a SECIS-dependent Sec codon.<sup>[23]</sup>

Natural reassignment of sense codons has not been seen in bacteria. But it is known in mitochondrial genomes, where a particular codon lost its original assignment and now leads to insertion of another amino acid.<sup>[24]</sup> Our case here is different; Sec insertion is mediated by a SECIS element and thus gives rise to dual use of the codon for another amino acid (through pairing with tRNA<sup>Sec</sup> variants carrying the proper anticodon).

What might account for this facile recoding to Sec? It is pertinent to note that Sec incorporation is different from that of all other amino acids; it is facilitated by its own “orthogonal” system<sup>[5]</sup> consisting of a different elongation factor (SelB), a required SECIS RNA element, a structurally unusual tRNA (tRNA<sup>Sec</sup>),<sup>[10]</sup> a dual meaning stop codon (UGA), and the use of release factor 2 (RF 2). Therefore, Sec recoding events may not have as general an effect on the protein translation machinery as might be expected from recoding canonical sense codons.<sup>[25]</sup> The ease of Cys to Sec recoding may be a consequence of the often desirable properties of selenoenzymes and selenoproteins with novel redox functions and increased enzyme activity,<sup>[4,26]</sup> while still allowing the expression of useful Cys-proteins and Cys-enzymes. Our finding of facile Sec recoding also opens our minds to the possible existence of other coding schemes. It also underscores the limitations of the current computational programs to predict selenoproteins from genome sequences, as these algorithms rest on UGA as the sole Sec codon.

Overall our approach provides new evidence of a limited but unequivocal plasticity of the genetic code whose secrets still lie hidden in the majority of unsequenced organisms.

## Acknowledgements

We thank Hans Aerni and Jesse Rinehart for advice on LC-MS/MS and Jean Kanyo (Yale University) for the dedicated efforts on the MS analyses. We also thank Andreas Brune, Filipa Gody-Vitorino, Hans-Peter Klenk, Ryan Lynch, Katherine McMahon, Daniel Marcus, William Mohn, Len Pennacchio, and Ameet Pinto for permission to use unpublished sequence data produced through the DOE-JGI's

community sequencing program. We are grateful to Patrick O'Donoghue, Oscar Vargas-Rodriguez, Jiqiang Ling for enlightened discussions and Daniel Drell and Robert Stack for encouragement. This work was supported by grants from the National Institute for General Medical Sciences (GM22854 to D.S.) and from the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the Department of Energy (DE-FG02-98ER20311 to D.S.; for funding the genetic experiments). The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, was supported under Contract No. DE-AC02-05CH11231.

**Keywords:** genetic code · metagenome · selenocysteine · sense codon recoding · synthetic biology

- [1] J. E. Cone, R. M. Del Rio, J. N. Davis, T. C. Stadtman, *Proc. Natl. Acad. Sci. USA* **1976**, *73*, 2659–2663.
- [2] A. Ambrogelly, S. Palioura, D. Söll, *Nat. Chem. Biol.* **2007**, *3*, 29–35.
- [3] G. V. Kryukov, S. Castellano, S. V. Novoselov, A. V. Lobanov, O. Zehtab, R. Guigo, V. N. Gladyshev, *Science* **2003**, *300*, 1439–1443.
- [4] N. Metanis, D. Hilvert, *Curr. Opin. Chem. Biol.* **2014**, *18*, 27–34.
- [5] A. Böck, M. Thanbichler, M. Rother, A. Resch in *Aminoacyl-tRNA Synthetases* (Eds.: M. Ibba, C. S. Francklyn, S. Cusack), Landes Bioscience, Georgetown, TX, **2005**, pp. 320–327.
- [6] J. Heider, C. Baron, A. Böck, *EMBO J.* **1992**, *11*, 3759–3766.
- [7] M. J. Bröcker, J. M. Ho, G. M. Church, D. Söll, P. O'Donoghue, *Angew. Chem. Int. Ed.* **2014**, *53*, 319–323; *Angew. Chem.* **2014**, *126*, 325–330.
- [8] N. N. Ivanova, P. Schwientek, H. J. Tripp, C. Rinke, A. Pati, M. Huntemann, A. Visel, T. Woyke, N. C. Kyrpides, E. M. Rubin, *Science* **2014**, *344*, 909–913.
- [9] A. Schön, A. Böck, G. Ott, M. Sprinzl, D. Söll, *Nucleic Acids Res.* **1989**, *17*, 7159–7165.
- [10] a) S. Palioura, R. L. Sherrer, T. A. Steitz, D. Söll, M. Simonovic, *Science* **2009**, *325*, 321–325; b) Y. Itoh, S. Chiba, S. Sekine, S. Yokoyama, *Nucleic Acids Res.* **2009**, *37*, 6259–6268; c) Y. Itoh, S. Sekine, S. Suetsugu, S. Yokoyama, *Nucleic Acids Res.* **2013**, *41*, 6729–6738.
- [11] T. Abe, H. Inokuchi, Y. Yamada, A. Muto, Y. Iwasaki, T. Ikemura, *Front. Genet.* **2014**, *5*, 114, DOI: 10.3389/fgene.2014.00114.
- [12] V. M. Markowitz, I. M. Chen, K. Chu, E. Szeto, K. Palaniappan, M. Pillay, A. Ratner, J. Huang, I. Pagani, S. Tringe, M. Huntemann, K. Billis, N. Varghese, K. Tennessen, K. Mavromatis, A. Pati, N. N. Ivanova, N. C. Kyrpides, *Nucleic Acids Res.* **2014**, *42*, D568–D573.
- [13] a) G. M. Luedemann, *J. Bacteriol.* **1968**, *96*, 1848–1858; b) N. Ivanova et al., *Stand. Genomic Sci.* **2010**, *2*, 158–167.
- [14] Y. Zhang, V. N. Gladyshev, *PLoS Genet.* **2008**, *4*, e1000095.
- [15] F. Zinoni, A. Birkmann, W. Leinfelder, A. Böck, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 3156–3160.
- [16] M. J. Axley, A. Böck, T. C. Stadtman, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 8450–8454.
- [17] M. A. Mandrand-Berthelot, M. Y. K. Wee, B. A. Haddock, *FEMS Microbiol. Lett.* **1978**, *4*, 37–40.
- [18] M. E. Pavan, E. E. Pavan, N. I. Lopez, L. Levin, M. J. Pettinari, *Genome Announc.* **2013**, *1*, e00675–00613.
- [19] A. T. Vincent, M. V. Trudel, L. Freschi, V. Nagar, C. Gagne-Thivierge, R. C. Levesque, S. J. Charette, *BMC Genomics* **2016**, *17*, 44.
- [20] S. Yoshizawa, L. Rasubala, T. Ose, D. Kohda, D. Fourmy, K. Maenaka, *Nat. Struct. Mol. Biol.* **2005**, *12*, 198–203.
- [21] J. Lu, A. Holmgren, *Free Radical Biol. Med.* **2014**, *66*, 75–87.
- [22] a) H. C. Lin, S. C. Ho, Y. Y. Chen, K. H. Khoo, P. H. Hsu, H. C. Yen, *Science* **2015**, *349*, 91–95; b) A. A. Turanov, R. A. Everley, S. Hybsier, K. Renko, L. Schomburg, S. P. Gygi, D. L. Hatfield, V. N. Gladyshev, *PLoS One* **2015**, *10*, e0140353.
- [23] A. A. Turanov, A. V. Lobanov, D. E. Fomenko, H. G. Morrison, M. L. Sogin, L. A. Klobutcher, D. L. Hatfield, V. N. Gladyshev, *Science* **2009**, *323*, 259–261.
- [24] a) S. Yokobori, *Viva Origino* **2014**, *42*, 47–49; b) J. Ling, P. O'Donoghue, D. Söll, *Nat. Rev. Microbiol.* **2015**, *13*, 707–721.
- [25] a) J. M. Ho, N. M. Reynolds, K. Rivera, M. Connolly, L. T. Guo, J. Ling, D. J. Pappin, G. M. Church, D. Söll, *ACS Synth. Biol.* **2015**, *5*, 163–171; b) T. Mukai, A. Yamaguchi, K. Ohtake, M. Takahashi, A. Hayashi, F. Irahara, S. Kira, T. Yanagisawa, S. Yokoyama, H. Hoshi, T. Kobayashi, K. Sakamoto, *Nucleic Acids Res.* **2015**, *43*, 8111–8122.
- [26] a) J. Beld, K. J. Woycechowsky, D. Hilvert, in *Oxidative Folding of Peptides and Proteins*, The Royal Society of Chemistry, London, **2009**, pp. 253–273; b) G. W. Snider, E. Ruggles, N. Khan, R. J. Hondal, *Biochemistry* **2013**, *52*, 5472–5481.

Received: December 15, 2015

Published online: ■■■■■, ■■■■■

## Communications



## Genetic Code

T. Mukai, M. Englert, H. J. Tripp, C. Miller,  
N. N. Ivanova, E. M. Rubin,  
N. C. Kyrpides, D. Söll\* — ■■■-■■■

Facile Recoding of Selenocysteine in  
Nature

## Selenocysteine codons

■ major	UUA	UCA	UAA	UGU
■ minor			UAG	UGC
■ expected				UGA
■ less expected				UGG
				CGA
		AAA	AGA	
		AAG		
			GGA	

**Stop making sense:** Selenocysteine is encoded by UGA, a stop codon re-assigned by a Sec-specific elongation factor and a distinctive RNA structure. Analysis of 6.4 trillion base pairs of metagenomic sequences and about 25 000 genomes revealed tRNA<sup>Sec</sup> species that recognize the stop codons UAG and UAA, and ten sense codons. Thus the genetic code is much more flexible than previously thought.