Title
Facile Recoding of Selenocysteine in Nature

Permalink
https://escholarship.org/uc/item/8m84f4f1

Journal

ISSN
1433-7851

Authors
Mukai, T
Englert, M
Tripp, Hj
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 24,903 microbial genomes for tRNA^{Sec} species. As expected, UGA is the predominant Sec codon in use. We also found tRNA^{Sec} 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 75Se or mass spectrometry. Other tRNA^{Sec} 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.[1] Selenoproteins are present in organisms from all domains of life,[2] such proteins are essential in mammalian cells,[3] yet plants and fungi lack this amino acid. Sec is present in the active site of many redox enzymes.[4] The codon for Sec is UGA which is normally a translational stop signal.[5] 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.[5, 6] Recently, a synthetic biology study succeeded in reassigning Sec to a large number of sense and stop codons in Escherichia coli,[7] demonstrating that alterations to the genetic code can possibly aide organisms to adapt to changing environments, and show the genetic code is much more flexible than previously thought.

We then wanted to confirm the coding properties of these non-canonical tRNA^{Sec} species. For proof of selenoprotein synthesis three strategies were possible:
1) metabolic labeling of the organisms with 75Se
2) replacing parts of the E. coli selenoprotein synthesis machinery with genes and tRNA^{Sec} 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.2015111657.
3) replacing *E. coli* tRNA^Sec^ with the newly discovered tRNA^Sec^ species.

In the last two strategies *E. coli* formate dehydrogenase H (FDH_H encoded by the *fdhF* gene) would serve as reporter.\[12\]

To confirm UAG-directed Sec incorporation we grew *Geodermatophilus obscurus* G-20 and *Blastococcus saxobsidens* cells in the presence of [\(^{75}\text{Se}\)]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,\[14\] (Figure 1.

**Figure 1.** Non-canonical selenocysteine assignments in nature. A) The tRNA\(^{\text{sec}}\) computer-based search pipeline and the manually curated output. The non-canonical tRNA\(^{\text{sec}}\) 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\(^{\text{sec}}\)^\text{UCA} 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\(^{\text{sec}}\) species. The nucleotide polymorphism among the same tRNA\(^{\text{sec}}\) group is indicated with green letters. The Sec codons and the SECIS elements of formate dehydrogenase mRNAs are shown.
Figure 2. Recoding of UAG and cysteine codons to selenocysteine. A) Metabolic 75Se 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) FDHH expression in E. coli ΔselABC ΔfdhF 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^Sec^CUA. The expressed selenoprotein FDHH reduced benzyl viologen, resulting in a purple color. C) FDHH activity of A. salmonicida subsp. pectinolytica 34mel cells. D) Metabolic 75Se labeling of the FDHH of 34mel. E) The procedure of sample preparation for the LC-MS/MS analysis of the FDHH, seloprotein. F) PMF confirms Sec incorporation at codon 140 in the recombinant FDHH. G) tRNA^Sec^CUX Bne (selC) locus in a metagenomic contig. H) In vivo FDHH assays in E. coli ΔselABC ΔfdhF 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 75Se, and the radioactive FDHH proteins were analyzed by SDS-PAGE and autoradiography or western blotting (WB).
Burkholderiales

selA

and

selB

nolytica

34mel.

A. pecti-

translated as Sec in a SECIS-dependent manner in

These data confirm that in this

metagenomic

Burkholderiales

D

for each anticodon type and expressed them in an

E. coli

We selected one representative metagenomic tRNASec species

stop and sense codons were also tested for Sec reassignment.

Y577,[19] pair tRNASec

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 ΔselABC strain was complemented with the Burkholderiales contig selAB and selC-opal variant, and an E. coli fdhFΔcys variant harboring the Burkholderiales contig SECIS element. This strain produced active FDH2 (purple color; Figure 2H, the 2nd row), but an inactive SECIS element (with a G25C mutation) did not form FDH2, and the cells were colorless (Figure 2H, the 4th row). In combination with the Burkholderiales contig selC, we changed UGAAG to UGC for the chimeric fdhF variant that carried functional or inactive SECIS elements. As the FDH2 Cys140 enzyme produced a purple color (Figure 2H, the 1st and 3rd rows), 75Se-labeling was used to demonstrate that the functional SECIS element led to a clear signal (Figure 2H). Thus, the Burkholderiales contig selA, selB, tRNASEC_GCA, and SECIS enabled UGC-recoding in E. coli.

The metagenomic tRNASEC variants that recognize other stop and sense codons were also tested for SEC reassignment. We selected one representative metagenomic tRNASEC species for each anticodon type and expressed them in an E. coli ΔselC ΔfdhF strain, together with the E. coli fdhF variants that carry the proper cognate codons at position 140.[7] Surprisingly, all but tRNASEC_UGG of the tested tRNASEC species recoded the respective codons for Sec, as they supported the expression of active FDH2 in their host E. coli cells (Figure S10). It should be mentioned that GGA was also poorly recoded in our earlier Sec recoding strategy.[7] The different recoding efficiencies may result from distortions of the ideal SECIS element structure by the nature of the upstream codon.[20] In light of these results we believe that these tRNASEC species may be used for recoding sense codons in the organisms they originate from.

What about eukaryotic organisms? Although we found nine tRNASEC variants of algal origin (two are shown in Figure S7), they need further validation, because they are almost identical to canonical tRNASEC 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[21] or due to the sophisticated backup systems[22] 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.[23]

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.[24] 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 tRNASEC 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[5] consisting of a different elongation factor (SelB), a required SECIS RNA element, a structurally unusual tRNA (tRNASEC),[10] 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.[25] 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,[12,26] 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 Arerni 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


Received: December 15, 2015
Published online: 0000000.

These are not the final page numbers!
Facile Recoding of Selenocysteine in Nature

Stop making sense: Selenocysteine is encoded by UGA, a stop codon reassigned 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_{Sec} species that recognize the stop codons UAG and UAA, and ten sense codons. Thus the genetic code is much more flexible than previously thought.