

## Recoding the Genetic Code with Selenocysteine\*\*

Markus J. Bröcker, Joanne M. L. Ho, George M. Church, Dieter Söll,\* and Patrick O'Donoghue\*

**Abstract:** Selenocysteine (Sec) is naturally incorporated into proteins by recoding the stop codon UGA. Sec is not hardwired to UGA, as the Sec insertion machinery was found to be able to site-specifically incorporate Sec directed by 58 of the 64 codons. For 15 sense codons, complete conversion of the codon meaning from canonical amino acid (AA) to Sec was observed along with a tenfold increase in selenoprotein yield compared to Sec insertion at the three stop codons. This high-fidelity sense-codon recoding mechanism was demonstrated for *Escherichia coli* formate dehydrogenase and recombinant human thioredoxin reductase and confirmed by independent biochemical and biophysical methods. Although Sec insertion at UGA is known to compete against protein termination, it is surprising that the Sec machinery has the ability to outcompete abundant aminoacyl-tRNAs in decoding sense codons. The findings have implications for the process of translation and the information storage capacity of the biological cell.

Sense-codon recoding is supposed to be impossible. Indeed, the fact that a codon can have more than one meaning was a dogma-breaking finding<sup>[1]</sup> and is one reason why the mechanism of selenocysteine (Sec) insertion into proteins provoked intense biochemical investigation over the last three decades. In addition, selenium is an essential micronutrient for humans.<sup>[2]</sup> Selenium in proteins is found in the form of Sec in enzymes that maintain the cell's redox balance, defending the cell against reactive oxygen species. Diseases involving Sec biosynthesis or selenoprotein malfunction have only recently surfaced because defects in these pathways are devastatingly detrimental to proper neuronal function and development.<sup>[3]</sup>

The promise of sense-codon recoding is being actively pursued as a means to further expand the genetic code and synthesize proteins with multiple noncanonical amino acids (ncAAs).<sup>[4]</sup> Nature expanded the genetic code with two amino

acids (Sec and pyrrolysine) by recoding or reassignment of stop codons.<sup>[5]</sup> Thus, an efficient, naturally evolved machinery already exists that directs recoding of particular UGA stop codons to Sec (Scheme 1 A). Although Sec is not found in all organisms (notably lacking in plants, fungi, and most Archaea), the 21st amino acid is genetically encoded in all three domains of life. Sec is biosynthesized on its tRNA,<sup>[6]</sup> and translational recoding of UGA requires the resulting Sec-tRNA<sup>Sec</sup> product, a specialized elongation factor (SelB in *E. coli*), and a downstream mRNA hairpin motif known as the Sec insertion sequence (SECIS).

We investigated the possibility to site-specifically reassign multiple sense codons using the Sec machinery. Previous attempts to encode Sec with the Leu UUA codon<sup>[7]</sup> and the Trp UGG codon<sup>[8]</sup> produced lower selenoprotein yields compared to UGA-encoded Sec and suggested significant canonical AA contamination. To systematically investigate the recoding capacity of the Sec machinery, we created a library of *E. coli* formate dehydrogenase (FDH<sub>H</sub>) mutants. Each gene variant had one of the 64 codons encoding a critical Sec residue (Sec140). In nature, FDH<sub>H</sub> is part of the membrane-associated formate hydrogen lyase (FHL) complex that decomposes formate to H<sub>2</sub> and CO<sub>2</sub> under fermentative conditions. The FHL complex shuttles electrons from formate to hydrogenase 3 which reduces protons to hydrogen molecules during anaerobic respiration.<sup>[9]</sup> Because the enzymatic activity of FDH<sub>H</sub> is dependent on the active-site residue Sec140, which coordinates an active-site molybdopterin cofactor,<sup>[10]</sup> sense-codon recoding is easily monitored in vivo and in vitro using the artificial electron acceptor benzyl viologen (BV). Active FDH<sub>H</sub> reduces BV to give a purple color<sup>[11]</sup> that is clearly visualized in living cells or monitored spectroscopically with purified FDH<sub>H</sub>.

Each FDH<sub>H</sub> gene variant (*fdhF*<sub>140</sub>NNN) was co-expressed with *selA*, *selB*, and a *selC* mutant (expressing tRNA<sup>Sec</sup><sub>NNN</sub>)

[\*] Dr. M. J. Bröcker, Prof. D. Söll  
Department of Molecular Biophysics & 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)

J. M. L. Ho, Prof. G. M. Church  
Department of Genetics, Harvard Medical School  
Boston, MA 02115 (USA)

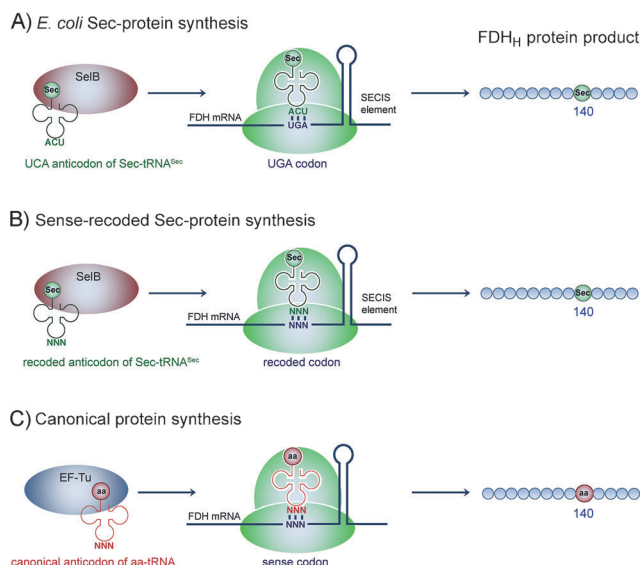
Prof. P. O'Donoghue  
Departments of Biochemistry and Chemistry  
Western University, London, ON N6A 5C1 (Canada)  
E-mail: patrick.odonoghue@uwo.ca

[\*\*] We thank Hans Aerni, Måns Ehrenberg, Ilka Heinemann, and Eric Westhof for insightful comments and discussion. We are also

grateful for the dedicated efforts of Kathryn Stone and Jean Kanyo at the W. M. Keck MS facility at Yale, and Zhengrong Wang and Ying Kiu from the Kline Geology Lab at Yale for assistance with ICP-MS analysis. This work was supported by grants to D.S. from the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the U.S. Department of Energy (DE-FG02-98ER20311; for funding the genetic experiments) and the National Institute of General Medical Sciences (GM22854), and by DARPA contract N66001-12-C-4211. M.J.B. was a Feodor Lynen Postdoctoral Fellow of the Alexander von Humboldt Foundation (Germany). J.M.L.H. was supported by a graduate fellowship from Harvard Medical School.



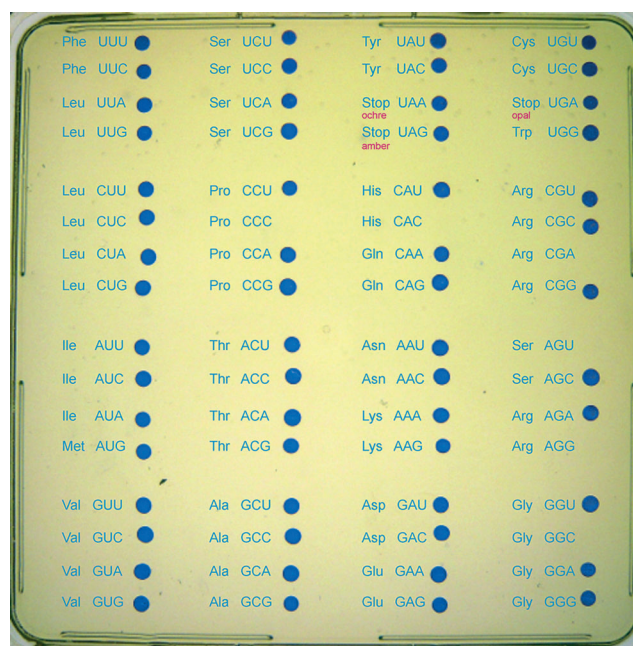
Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201308584>.



**Scheme 1.** Codon recoding with Sec. A) Synthesis of the *E. coli* Sec-containing formate dehydrogenase (FDH<sub>H</sub>) from the wild-type *fdhF* gene requires recoding the UGA stop codon at position 140 to Sec by Sec-tRNA<sup>Sec</sup>, elongation factor SelB, and the SECIS mRNA hairpin. Replacing the UGA with any of the 64 NNN triplets at codon 140 in *fdhF* and co-expressing a cognate Sec-tRNA<sup>Sec</sup><sub>NNN</sub> is expected to yield SelB and SECIS-dependent sense-codon-recoded Sec-containing FDH<sub>H</sub> (B) that may compete with canonical AA insertion (C) directed by native aminoacyl-tRNAs and EF-Tu dependent protein synthesis. In this case (C), SECIS is still present in the *fdhF* mRNA, but SECIS is not expected to interact with EF-Tu directed decoding.

with the cognate anticodon (Scheme 1). We anticipated that these constructs could give rise to two different protein products. Sec-tRNA<sup>Sec</sup><sub>NNN</sub> and SelB should compete with canonical aminoacyl-tRNA<sub>NNN</sub> (AA-tRNA<sub>NNN</sub>) and EF-Tu, which could give rise to Sec-containing FDH<sub>H</sub> (Scheme 1 B), canonical AA containing FDH<sub>H</sub> (Scheme 1 C), or a mixture of both protein species. The plasmid-borne FDH<sub>H</sub> and *sel* genes complemented an *E. coli*  $\Delta selA \Delta selB \Delta fdhF$  deletion strain (MH5)<sup>[12]</sup> that is otherwise unable to produce selenoproteins. Quite unexpectedly, the in vivo assay shows that the Sec recoding machinery successfully alters the meaning of all three stop codons and nearly all sense codons to Sec (Figure 1). For 58 of the possible 64 codons, the coloration obtained by Sec-dependent BV reduction is as intense as that for the UGA (wild-type) Sec codon. Although this assay is qualitative, it demonstrates for the first time that most codons are recodable. At least one codon for each of the 20 canonical amino acids is recodable to Sec; so no particular AA-tRNA species is able to completely outcompete Sec insertion. For example, the meaning of all six Leu codons, five out of six Ser codons, and four out of six Arg codons was re-assigned to Sec.

Only six codons proved refractory to Sec insertion as indicated by the lack of FDH<sub>H</sub> activity. Strikingly, four of these codons (CGA, AGU, AGG, and GGC) were found in the NGN boxes of the codon table. SDS gels and western blot analysis (Figures S1 and S2) show that full-length FDH<sub>H</sub> of 80 kDa is produced for all codons except the arginine codons



**Figure 1.** Recoding the genetic code with Sec. The canonical genetic code table is overlaid on a single agar plate spotted with 64 *E. coli* FDH<sub>H</sub> variants. In each case, an *E. coli*  $\Delta selA \Delta selB \Delta fdhF$  deletion parent strain was complemented with *E. coli* *selA*, *selB*, and each of the 64 *fdhF*<sub>140</sub>NNN codon mutants and tRNA<sup>Sec</sup><sub>NNN</sub> variants with the respective cognate anticodon. The capacity of each strain to recode the indicated codon to Sec is evidenced by an in vivo BV reduction assay in which purple colored cells express active Sec-containing FDH<sub>H</sub>.

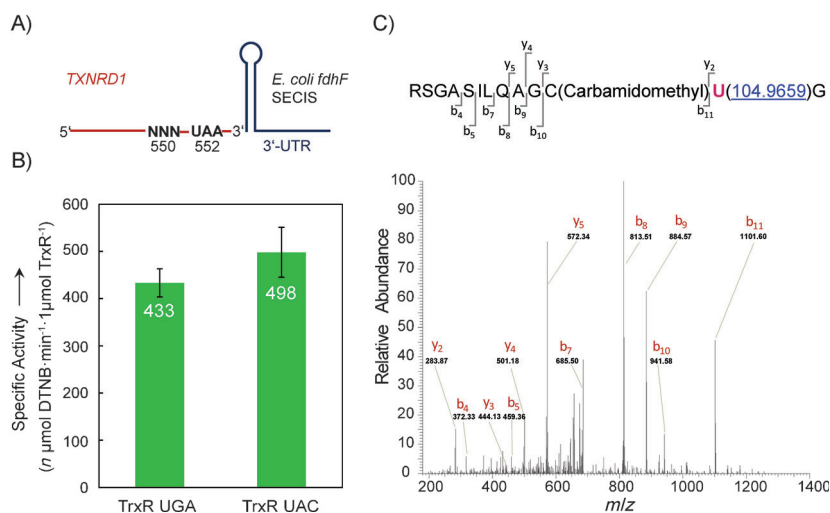
AGG and CGA. These codons show a truncated FDH<sub>H</sub> product at 15.5 kDa, suggesting that protein synthesis terminates at position 140. These data indicate that canonical AA-tRNA completely outcompetes Sec insertion (Scheme 1 C) at AGU, GGC, CAC, and CCC, but the meaning of AGG and CGA is converted to stop by an unknown mechanism.

The Sec machinery has the inherent capability to recode nearly all codons (Figure 1), but this in vivo assay does not provide quantitative information regarding the degree of recoding. Under anaerobic conditions, 64 N-terminal His-tagged FDH<sub>H</sub> proteins were over-expressed from the *fdhF*<sub>140</sub>NNN gene variants and purified (Figures S1 and S2). The Sec insertion ratio was quantitated by comparison of the initial velocity of BV reduction in sense-codon-recoded versus wild-type FDH<sub>H</sub> variants in vitro (Table S1). Although the same concentration of purified FDH<sub>H</sub> protein is used in each assay, the specific activity values differ, which indicates the amount of active (i.e., Sec140) FDH<sub>H</sub> versus the amount of inactive (i.e., canonical AA140) FDH<sub>H</sub> protein in each preparation. The assays confirmed Sec incorporation in FDH<sub>H</sub> for 58 codons with relative FDH<sub>H</sub>-specific activities ranging from 12–100 % of the wild-type. Selenoprotein yield was greater for most of the recoded sense codons compared to Sec insertion at UGA or the other stop codons, UAA and UAG (Table S2).

Plotting FDH<sub>H</sub> protein yield versus specific activity (Figure 2A) revealed six distinct types of recoding behavior.

[illegible]





**Figure 3.** Recoding the Tyr codon UAC to Sec in human TrxR1. A) Schematic representation of the expression construct for human TXNRD1. An *E. coli fdhF* SECIS element is attached directly after the UAA codon of TXNRD1 in the 3'-untranslated region (UTR). B) In vitro activity assay of recombinant human TrxR<sub>550</sub>UGA and TrxR<sub>550</sub>UAC co-expressed with tRNA<sup>Sec</sup><sub>UCA</sub> and tRNA<sup>Sec</sup><sub>GUA</sub>, respectively. Error bars show standard deviations of quadruplicate experiments. C) MS/MS identification of Sec incorporation into TrxR<sub>550</sub>UAC. The spectrum of the Sec-containing peptide, the peptide sequence and the positions of the observed b- and y-ions are indicated. Sec was identified by MASCOT as a Se-IAN-Cys modification of Cys; selenium alkylation (Se-CH<sub>2</sub>-CONH<sub>2</sub>).

therapeutic development. Similar to our observations of UAC-encoded Sec in FDH<sub>H</sub>, TrxR<sub>550</sub>UAC (4.1 mg L<sup>-1</sup> culture) leads to fourfold higher yield of pure selenoprotein compared to the UGA-encoded selenoenzyme (0.9 mg L<sup>-1</sup> culture). The specific activities are indistinguishable for TrxR produced from these two constructs (Figure 3B). No Tyr contamination was detected by MS/MS analysis (Figure 3C), and ICP-MS analysis independently confirmed 98.5% Sec incorporation in response to UGA and 95.1% Sec insertion for UAC in TrxR (Table S3).

It remains unclear why certain codons are more “recodable” than others. It is reasonable to assume that codon usage might correlate with recodability, because translation is believed to be less efficient with lower abundant codons. The intracellular concentration of AA-tRNA<sup>[6]</sup> is another likely explanation. Perhaps codons read by less abundant AA-tRNA species are more easily recoded. Despite the logic of these ideas, there is no general correlation between recodability and codon usage or AA-tRNA concentration (Figure S5). There are potential complicating factors; perhaps the Sec-tRNA<sup>Sec</sup><sub>NNN</sub> species are not decoded on the ribosome with equal efficiency due to compatibility of the tRNA body and anticodon loop. There is also the question of whether the nature of the mRNA codon (position 140) affects the structural integrity of the SECIS element. Furthermore, differences between AA-tRNA levels may not be sufficient enough to fundamentally alter competition with Sec-tRNA<sup>Sec</sup><sub>NNN</sub>. Based on crystal structures, a model of the ribosome-bound Sec-decoding complex (i.e., ribosome, SelB, Sec-tRNA<sup>Sec</sup>, mRNA with SECIS) indicates extensive interactions between SelB and the 30S subunit in regions that are known to alter the accuracy of mRNA decoding.<sup>[17]</sup> Kinetic

and structural studies of sense-codon decoding with Sec will help define the mechanism discovered here.

In the normal Sec-decoding situation, the release factor (RF2 in *E. coli*) competes with Sec insertion,<sup>[18]</sup> leading to truncated FDH<sub>H</sub> (Figure S4). Encoding Sec with sense codons escapes competition with RFs, which leads to the significantly enhanced selenoprotein yield for all but ten codons. It is unclear why nature selected UGA as the “Sec” codon when most other codons lead to more selenoprotein production. Perhaps the scarcity of Se in the environment induced organisms to naturally limit the level of selenoprotein synthesis with RF competition. RF interaction may, nevertheless, affect the recodability of certain sense codons. Due to its similarity with UGA, UGG is a hotspot for premature termination by RF2,<sup>[19]</sup> which may explain why this codon produces less Sec-containing FDH<sub>H</sub> than all other group II codons (Figure 2).

Encoding Sec with sense codons will become a useful method for efficiently producing selenoproteins; yet this discovery opens the door to recode many codons, not

only with Sec, but possibly also with a diverse array of ncAAs of biomedical and biotechnological interest. The human proteome contains many essential modified proteins that arise by posttranslational modification (PTM), and mismodified proteins may lead to defects in cellular signaling and protein aggregates that form the molecular basis for diseases. In order to biochemically define the role of PTMs and to produce therapeutic agents that specifically target mismodified proteins, there is an urgent need to synthesize proteins containing several different, distinct ncAAs. Progress in the last decade allowed site-specific incorporation of one or two ncAAs into recombinant proteins by reassignment of non-sense codons.<sup>[4b]</sup> In vitro protein synthesis methods were recently enhanced by creating dual-meaning initiation and sense codons to synthesize potential chemotherapeutic peptides with two ncAAs,<sup>[20]</sup> and amino acid starvation methods permitted production of proteins with three ncAAs for biological imaging applications.<sup>[21]</sup> Furthermore, unlike the Sec machinery that only recodes codons associated with SECIS, current genetic code expansion methods are not site-specific because they lead to global reassignment of stop codons; this contaminates the natural proteome with ncAAs, resulting in growth defects.<sup>[22]</sup>

In order to synthesize proteins with more than 23 amino acids, additional “recodable” codons are needed.<sup>[23]</sup> We anticipate that mutagenesis of the Sec insertion machinery will enable protein synthesis with multiple ncAAs, which will require engineering SelB, SECIS, and also orthogonal tRNA synthetases to acylate tRNA<sup>Sec</sup> and tRNA<sup>Sec</sup>-like molecules with ncAAs. The fact that SepCysS forms Cys-tRNA<sup>Sec</sup>, which actively decodes UGA with Cys instead of Sec, is an encouraging example.<sup>[24]</sup>

The fact that the Sec insertion machinery has the intrinsic capability to alter the meaning of 58 of the 64 codons indicates that the biological cell will be able to genetically encode far more amino acids than previously recognized. Although the Sec insertion machinery recodes some codons poorly, it is possible that the recoding fidelity may be enhanced through selection. Given that certain codons are recodable at high fidelity, it is conceivable that in nature Sec is encoded by codons other than UGA, suggesting the existence of unknown selenoproteins. Indeed, SECIS-like structures that are not associated with in-frame UGA codons have been identified bioinformatically and dismissed as evolutionary remnants.<sup>[25]</sup> If other RNA signals, like SECIS, exist in nature or if they can be engineered in the laboratory, there is potentially no limit to the number of amino acids that the cell can encode and no reason to stop at 20 or even 23.

Received: October 2, 2013

Published online: ■ ■ ■ ■ ■, ■ ■ ■ ■ ■

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

- [1] F. Zinoni, A. Birkmann, W. Leinfelder, A. Böck, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 3156–3160.
- [2] M. P. Rayman, *Lancet* **2000**, *356*, 233–241.
- [3] a) O. Agamy, B. B. Zeev, D. Lev, B. Marcus, D. Fine, D. Su, G. Narkis, R. Ofir, C. Hoffmann, E. Leshinsky-Silver, H. Flusser, S. Sivan, D. Söll, T. Lerman-Sagie, O. S. Birk, *Am. J. Hum. Genet.* **2010**, *87*, 538–544; b) S. Seeher, Y. Mahdi, U. Schweizer, *Curr. Protein Pept. Sci.* **2012**, *13*, 337–346.
- [4] a) R. Krishnakumar, L. Prat, H. R. Aerni, J. Ling, C. Merryman, J. I. Glass, J. Rinehart, D. Söll, *ChemBioChem* **2013**, *14*, 1967–1972; b) P. O'Donoghue, J. Ling, Y. S. Wang, D. Söll, *Nat. Chem. Biol.* **2013**, *9*, 594–598.
- [5] A. Ambrogelly, S. Palioura, D. Söll, *Nat. Chem. Biol.* **2007**, *3*, 29–35.
- [6] J. Yuan, P. O'Donoghue, A. Ambrogelly, S. Gundllapalli, R. L. Sherrer, S. Palioura, M. Simonovic, D. Söll, *FEBS Lett.* **2010**, *584*, 342–349.
- [7] M. J. Berry, J. W. Harney, T. Ohama, D. L. Hatfield, *Nucleic Acids Res.* **1994**, *22*, 3753–3759.
- [8] J. Xu, V. Croitoru, D. Rutishauser, Q. Cheng, E. S. Arnér, *Nucleic Acids Res.* **2013**, DOI: 10.1093/nar/gkt764.
- [9] G. Sawers, *Antonie van Leeuwenhoek* **1994**, *66*, 57–88.
- [10] J. C. Boyington, V. N. Gladyshev, S. V. Khangulov, T. C. Stadtman, P. D. Sun, *Science* **1997**, *275*, 1305–1308.
- [11] M. J. Axley, D. A. Grahame, T. C. Stadtman, *J. Biol. Chem.* **1990**, *265*, 18213–18218.
- [12] C. Aldag, M. J. Bröcker, M. J. Hohn, L. Prat, G. Hammond, A. Plummer, D. Söll, *Angew. Chem.* **2013**, *125*, 1481–1485; *Angew. Chem. Int. Ed.* **2013**, *52*, 1441–1445.
- [13] G. W. Snider, E. Ruggles, N. Khan, R. J. Hondal, *Biochemistry* **2013**, *52*, 5472–5481.
- [14] Y. C. Chen, K. S. Prabhu, A. M. Mastro, *Nutrients* **2013**, *5*, 1149–1168.
- [15] J. Lu, A. Holmgren, *Antioxid. Redox Signaling* **2012**, *17*, 1738–1747.
- [16] a) H. Dong, L. Nilsson, C. G. Kurland, *J. Mol. Biol.* **1996**, *260*, 649–663; b) K. A. Dittmar, M. A. Sorensen, J. Elf, M. Ehrenberg, T. Pan, *EMBO Rep.* **2005**, *6*, 151–157.
- [17] S. Yoshizawa, L. Rasubala, T. Ose, D. Kohda, D. Fourmy, K. Maenaka, *Nat. Struct. Mol. Biol.* **2005**, *12*, 198–203.
- [18] J. B. Mansell, D. Guevremont, E. S. Poole, W. P. Tate, *EMBO J.* **2001**, *20*, 7284–7293.
- [19] a) D. V. Freistrotter, M. Kwiatkowski, R. H. Buckingham, M. Ehrenberg, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 2046–2051; b) J. Sund, M. Ander, J. Aqvist, *Nature* **2010**, *465*, 947–950.
- [20] Y. Goto, M. Iseki, A. Hitomi, H. Murakami, H. Suga, *ACS Chem. Biol.* **2013**, DOI: 10.1021/cb400549p.
- [21] S. Lepthien, L. Merkel, N. Budisa, *Angew. Chem.* **2010**, *122*, 5576–5581; *Angew. Chem. Int. Ed.* **2010**, *49*, 5446–5450.
- [22] I. U. Heinemann, A. J. Rovner, H. R. Aerni, S. Rogulina, L. Cheng, W. Olds, J. T. Fischer, D. Söll, F. J. Isaacs, J. Rinehart, *FEBS Lett.* **2012**, *586*, 3716–3722.
- [23] M. J. Lajoie, A. J. Rovner, D. B. Goodman, H. R. Aerni, A. D. Haimovich, G. Kuznetsov, J. A. Mercer, H. H. Wang, P. A. Carr, J. A. Mosberg, N. Rohland, P. G. Schultz, J. M. Jacobson, J. Rinehart, G. M. Church, F. J. Isaacs, *Science* **2013**, *342*, 357–360.
- [24] J. Yuan, M. J. Hohn, R. L. Sherrer, S. Palioura, D. Su, D. Söll, *FEBS Lett.* **2010**, *584*, 2857–2861.
- [25] Y. Zhang, H. Romero, G. Salinas, V. N. Gladyshev, *Genome Biol.* **2006**, *7*, R94.

## Communications



### RNA Engineering



M. J. Bröcker, J. M. L. Ho, G. M. Church,  
D. Söll,\*  
P. O'Donoghue\* ———— ■■■■-■■■■

Recoding the Genetic Code with  
Selenocysteine

Pro UUU	Ser UCU	Tyr UAU	Cys UGU
Pro UUC	Ser UCC	Tyr UAC	Cys UGC
Leu UUA	Ser UCA	Stop UAA	Stop UGA
Leu UUG	Ser UCG	Stop UAG	Sec UGG
Leu GUU	Pro GCU	His CAU	Arg CGU
Leu GUC	Pro GCC	His CAC	Arg CGC
Leu GUA	Pro GCA	Gln CAA	Arg CGA
Leu GUG	Pro GCG	Gln CAG	Arg CGG
Ile AUU	Thr ACU	Asn AAU	Ser AGU
Ile AUC	Thr ACC	Asn AAC	Ser AGC
Ile AUA	Thr ACA	Lys AAA	Arg AGA
Met AUG	Thr ACG	Lys AAG	Arg AGG
Val GUU	Ala GCU	Asp GAU	Gly GGU
Val GUC	Ala GCC	Asp GAC	Gly GGC
Val GUA	Ala GCA	Glu GAA	Gly GGA
Val GUG	Ala GCG	Glu GAG	Gly GGG

The selenocysteine insertion machinery was engineered to recode multiple sense codons. This was demonstrated for *Escherichia coli* formate dehydrogenase and recombinant human thioredoxin reductase. In the picture the canonical genetic code table is overlaid with a single assay in which purple-colored cells express active Sec-containing enzymes.