Genetic Code Supports Targeted Insertion of Two Amino Acids by One Codon

Article *in* Science · February 2009

DOI: 10.1126/science.1164748

8 authors, including:

Alexey Lobanov
Lomonosov Moscow State University
58 PUBLICATIONS **6,050** CITATIONS
SEE PROFILE

Dmitri E Fomenko
University of Nebraska at Lincoln
101 PUBLICATIONS **2,768** CITATIONS
SEE PROFILE

Hilary G Morrison
Marine Biological Laboratory
261 PUBLICATIONS **13,546** CITATIONS
SEE PROFILE

Mitchell Sogin
MBL
548 PUBLICATIONS **49,621** CITATIONS
SEE PROFILE

Genetic Code Supports Targeted Insertion of Two Amino Acids by One Codon

Anton A. Turanov, *et al.*

Science **323**, 259 (2009);
DOI: 10.1126/science.1164748

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

tant, which has its three serine phosphorylation sites mutated to alanine, stays primarily in the nucleus (*5*, *7*) (fig. S9). The transfected cytoplasmic HDAC4-L175A mutant preserved *rd1* rods until at least P70 (Fig. 3, N and Q), whereas the nuclear HDAC4-3SA mutant failed to rescue *rd1* rods even at P50 (Fig. 3, O and R).

HIF1α plays a central role in the regulation of oxygen homeostasis (*20*). HIF1α protein is not detectable in the mature mouse retina (*21*). Exposure of retinas to hypoxia stabilizes HIF1α and protects photoreceptors from light-induced retinal degeneration (*21*). HIF1α protein stability is decreased by lysine acetylation. Acetylation of HIF1α by the acetyltransferase ARD1 enhances its degradation (*22*). HIF1α stabilization thus might provide a mechanism for HDAC4-induced photoreceptor protection in *rd1* mice. HIF1α protein was detected by immunohistochemistry in the OS of wild-type photoreceptors after HDAC4 electroporation (Fig. 4, A to D). No HIF1α immunoreactivity was detected after overexpression of HDAC6 (Fig. 4, E and F). Likewise, expression of HDAC4, but not that of HDAC6 (Fig. 4, I and J), led to the detection of HIF1α that appeared nuclear or perinuclear in the photoreceptors of the *rd1* retina (Fig. 4, G and H). Expression of a dominant negative HIF1α (dnHIF1α) construct (*23*) with pCAG-HDAC4 in the *rd1* retina negated the photoreceptor survival effect of HDAC4 (Fig. 4, K to N). A plasmid with an shRNA directed to HIF1α also blocked HDAC4-mediated photoreceptor survival (fig. S10). Thus, HIF1α appears to be required for the HDAC4 survival effect.

To determine whether acetylation of HIF1α might influence rod death in the *rd1* retina, we expressed HIF1αK532R, an acetylation mutant of HIF1α that is more stable than its wild-type form (*22*), and the wild-type HIF1α in the *rd1* retina. Wild-type HIF1α preserved a few rods (Fig. 4, P and T) and HIF1αK532R preserved more (Fig. 4, Q and U). HDAC4 was the most effective in saving rod photoreceptors (Fig. 4, R, V, and W). No additive effects were seen when HDAC4 was coexpressed with HIF1αK532R (fig. S11). The greater efficacy of HDAC4 relative to HIF1αK532R might result from HDAC4 having target(s) in addition to HIF1α, or HIF1αK532R could be less effective than deacetylated wild-type HIF1α.

In the mouse retina, HDAC4 has an essential role in neuronal survival. From a therapeutic perspective, HDAC4 prolonged photoreceptor survival in mice undergoing retinal degeneration. HDAC6 did not lead to increased abundance of HIF1α protein or promote rod survival in mice, although it rescued degeneration in *Drosophila* (*19*). Therefore, more than one pathway for neuronal survival may be regulated by HDACs.

References and Notes

1. X. J. Yang, E. Seto, *Oncogene* **26**, 5310 (2007).
2. T. A. Bolger, T. P. Yao, *J. Neurosci.* **25**, 9544 (2005).
3. T. A. Bolger, X. Zhao, T. J. Cohen, C. C. Tsai, T. P. Yao, *J. Biol. Chem.* **282**, 29186 (2007).
4. A. H. Wang, X. J. Yang, *Mol. Cell. Biol.* **21**, 5992 (2001).
5. X. Zhao *et al.*, *J. Biol. Chem.* **276**, 35042 (2001).
6. E. A. Miska *et al.*, *EMBO J.* **18**, 5099 (1999).
7. R. B. Vega *et al.*, *Cell* **119**, 555 (2004).
8. J. Lu, T. A. McKinsey, C. L. Zhang, E. N. Olson, *Mol. Cell* **6**, 233 (2000).
9. M. A. Arnold *et al.*, *Dev. Cell* **12**, 377 (2007).
10. T. Matsuda, C. L. Cepko, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16 (2004).
11. R. W. Young, *Anat. Rec.* **212**, 199 (1985).
12. J. T. Voyvodic, J. F. Burne, M. C. Raff, *Eur. J. Neurosci.* **7**, 2469 (1995).
13. D. L. Turner, C. L. Cepko, *Nature* **328**, 131 (1987).
14. R. W. Young, *J. Comp. Neurol.* **229**, 362 (1984).
15. C. Bowes *et al.*, *Nature* **347**, 677 (1990).
16. M. E. McLaughlin, M. A. Sandberg, E. L. Berson, T. P. Dryja, *Nat. Genet.* **4**, 130 (1993).
17. L. D. Carter-Dawson, M. M. LaVail, R. L. Sidman, *Invest. Ophthalmol. Vis. Sci.* **17**, 489 (1978).
18. K. Komeima, B. S. Rogers, L. Lu, P. A. Campochiaro, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11300 (2006).
19. U. B. Pandey *et al.*, *Nature* **447**, 859 (2007).
20. G. L. Semenza, *J. Appl. Physiol.* **88**, 1474 (2000).
21. C. Grimm *et al.*, *Nat. Med.* **8**, 718 (2002).
22. J. W. Jeong *et al.*, *Cell* **111**, 709 (2002).
23. B. H. Jiang, E. Rue, G. L. Wang, R. Roe, G. L. Semenza, *J. Biol. Chem.* **271**, 17771 (1996).

24. We thank J. R. Lu for providing HDAC4, HDAC5, and HDAC6 expression constructs; T. P. Yao for providing HDAC4-3SA mutant; X. J. Yang for providing HDAC4-L175A mutant; G. L. Semenza for providing dnHIF1α constructs and K. W. Kim for providing HIF1α and HIF1αK532R constructs; and T. Matsuda for assistance with various techniques. Supported by NIH grant EYO 14466, the Foundation for Retinal Research, a gift from Merck, and the Howard Hughes Medical Institute (HHMI). C.L.C. is an Investigator of HHMI and B.C. is a postdoctoral fellow of HHMI. C.L.C. and B.C. have filed a U.S. patent application, 61/088,455, based on the work in this paper.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5911/256/DC1
Materials and Methods
Figs. S1 to S11
References

22 September 2008; accepted 10 November 2008
10.1126/science.1166226

Genetic Code Supports Targeted Insertion of Two Amino Acids by One Codon

Anton A. Turanov,[1]* Alexey V. Lobanov,[1]* Dmitri E. Fomenko,[1] Hilary G. Morrison,[2] Mitchell L. Sogin,[2] Lawrence A. Klobutcher,[3] Dolph L. Hatfield,[4] Vadim N. Gladyshev[1]†

Strict one-to-one correspondence between codons and amino acids is thought to be an essential feature of the genetic code. However, we report that one codon can code for two different amino acids with the choice of the inserted amino acid determined by a specific 3′ untranslated region structure and location of the dual-function codon within the messenger RNA (mRNA). We found that the codon UGA specifies insertion of selenocysteine and cysteine in the ciliate *Euplotes crassus*, that the dual use of this codon can occur even within the same gene, and that the structural arrangements of *Euplotes* mRNA preserve location-dependent dual function of UGA when expressed in mammalian cells. Thus, the genetic code supports the use of one codon to code for multiple amino acids.

Although codons can be recoded to specify other amino acids or to have ambiguous meanings (*1*, *2*), and stop codons can be suppressed to insert amino acids (*3*), insertion of different amino acids into separate positions within nascent polypeptides by the same codeword is believed to be inconsistent with ribosome-based protein synthesis. In ciliated protozoa from the *Euplotes* genus, cysteine (Cys) is encoded by three codons, UGA, UGU, and UGC (*4*, *5*). UGA is a stop signal in the universal genetic code, and this codon can also code for the 21st amino acid, selenocysteine (Sec) (*6*).

Metabolic labeling with ^{75}Se showed that *E. crassus* contains multiple selenoproteins (fig. S1). To identify the codon used for Sec, we sequenced 15,000 *E. crassus* expressed sequence tags (ESTs) (fig. S2) and the full-length eGPx1 cDNA sequence. The eGPx1 cDNA encodes a 22-kD protein with a single in-frame UGA codon (Fig. 1A) and a Sec insertion sequence (SECIS) element (*7*) in its 3′ untranslated region (3′UTR) (Fig. 1B), which suggests that this UGA encodes Sec. Therefore, UGA may be used for both Cys and Sec insertion in *Euplotes*. Expression of eGPx1 as a fusion protein with green fluorescent protein (GFP) in human embryonic kidney (HEK) 293 cells revealed specific ^{75}Se incorporation (Fig. 1C). The corresponding full-size protein was also detected by Western blotting (Fig. 1D).

[1]Department of Biochemistry and Redox Biology Center, University of Nebraska, Lincoln, NE 68588, USA. [2]Josephine Bay Paul Center, Marine Biological Laboratory, Woods Hole, MA 02543, USA. [3]Department of Molecular, Microbial and Structural Biology, University of Connecticut Health Center, Farmington, CT 06032, USA. [4]Molecular Biology of Selenium Section, Laboratory of Cancer Prevention, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: vgladyshev1@unl.edu

Mutation of the core region of the eGPx1 SECIS element prevented ^{75}Se incorporation and protein synthesis (Fig. 1, C and D), indicating that SECIS was required for Sec insertion in response to UGA.

E. crassus genome sequencing and analysis revealed eight selenoprotein genes (figs. S3 to S16) and three tRNAs that recognize UGA codons, including Sec tRNA, mitochondrial Trp tRNA, and a novel Cys tRNA (Fig. 2A and fig. S17). A Cys tRNA recognizing UGU and UGC codons was also detected. Four of the eight selenoprotein genes contained multiple UGA codons (fig. S4). Comparison with known selenoproteins suggested the use of one codon for Sec and an additional UGA codon (or codons) within the same gene for Cys insertion. *E. crassus* thioredoxin reductases 1 (eTR1) and 2 (eTR2) had seven in-frame UGA codons, with the first six predicted to code for Cys and the last one to code for Sec (figs. S5, S6, and S18). To examine coding functions of UGA codons, we cloned a novel selenoprotein ep22, selenoprotein W2 (eSelW2), and eTR1 (figs. S5, S8, S10, and S19) and expressed them in the form of GFP-fusion proteins in HEK 293 cells. Specific ^{75}Se incorporation was observed into GFP-ep22 (Fig. 1, E and F) and GFP-eSelW2 (Fig. 1G), which had single UGA codons.

In the case of GFP-eTR1, we initially did not observe ^{75}Se incorporation (Fig. 2B, lane 2).

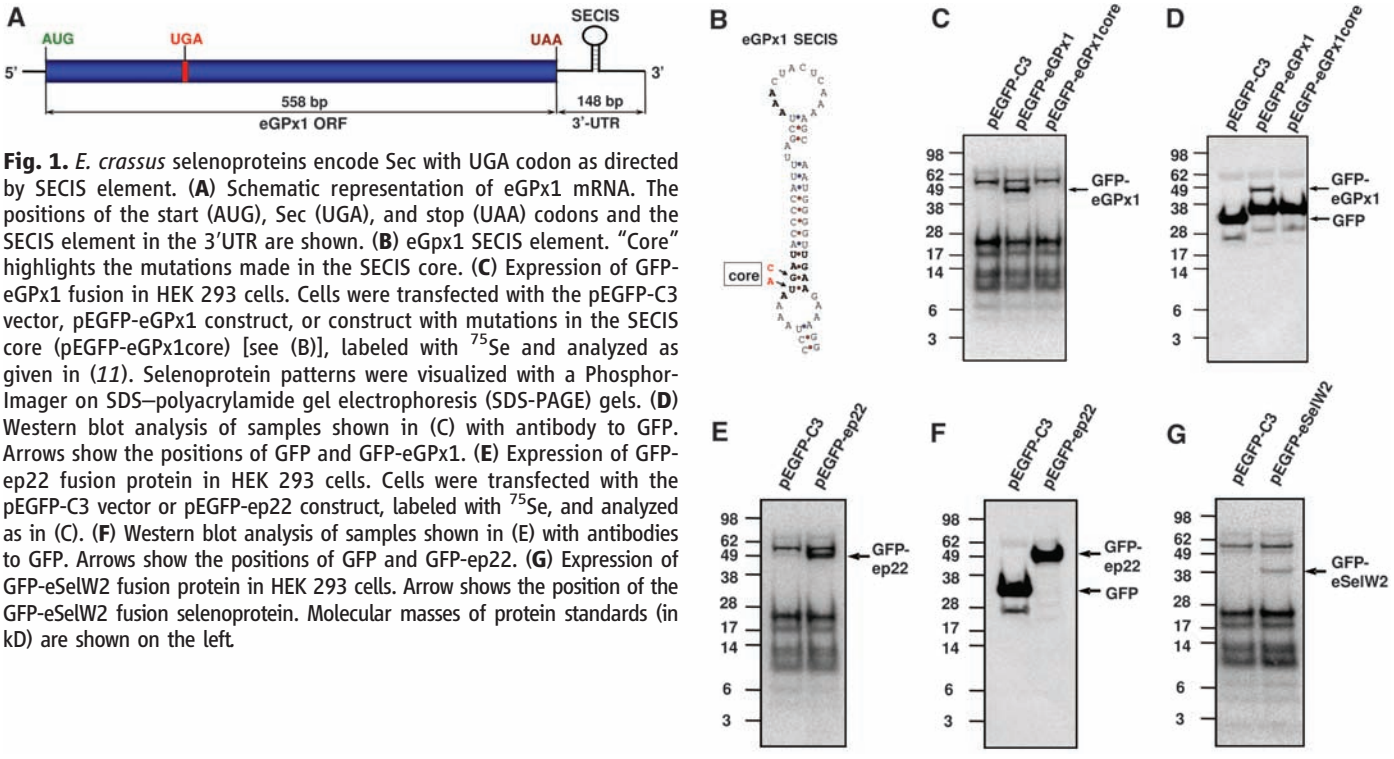

Fig. 1. *E. crassus* selenoproteins encode Sec with UGA codon as directed by SECIS element. (**A**) Schematic representation of eGPx1 mRNA. The positions of the start (AUG), Sec (UGA), and stop (UAA) codons and the SECIS element in the 3′UTR are shown. (**B**) eGpx1 SECIS element. "Core" highlights the mutations made in the SECIS core. (**C**) Expression of GFP-eGPx1 fusion in HEK 293 cells. Cells were transfected with the pEGFP-C3 vector, pEGFP-eGPx1 construct, or construct with mutations in the SECIS core (pEGFP-eGPx1core) [see (B)], labeled with ^{75}Se and analyzed as given in (*11*). Selenoprotein patterns were visualized with a PhosphorImager on SDS–polyacrylamide gel electrophoresis (SDS-PAGE) gels. (**D**) Western blot analysis of samples shown in (C) with antibody to GFP. Arrows show the positions of GFP and GFP-eGPx1. (**E**) Expression of GFP-ep22 fusion protein in HEK 293 cells. Cells were transfected with the pEGFP-C3 vector or pEGFP-ep22 construct, labeled with ^{75}Se, and analyzed as in (C). (**F**) Western blot analysis of samples shown in (E) with antibodies to GFP. Arrows show the positions of GFP and GFP-ep22. (**G**) Expression of GFP-eSelW2 fusion protein in HEK 293 cells. Arrow shows the position of the GFP-eSelW2 fusion selenoprotein. Molecular masses of protein standards (in kD) are shown on the left.

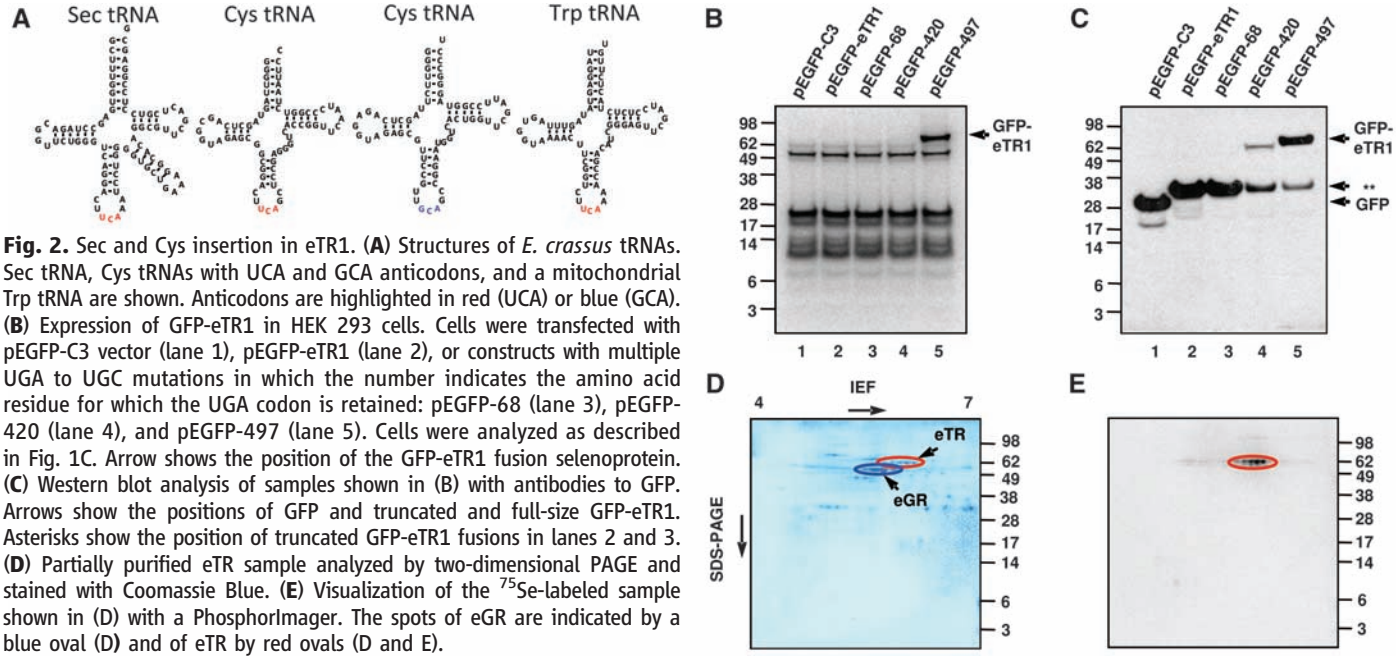

Fig. 2. Sec and Cys insertion in eTR1. (**A**) Structures of *E. crassus* tRNAs. Sec tRNA, Cys tRNAs with UCA and GCA anticodons, and a mitochondrial Trp tRNA are shown. Anticodons are highlighted in red (UCA) or blue (GCA). (**B**) Expression of GFP-eTR1 in HEK 293 cells. Cells were transfected with pEGFP-C3 vector (lane 1), pEGFP-eTR1 (lane 2), or constructs with multiple UGA to UGC mutations in which the number indicates the amino acid residue for which the UGA codon is retained: pEGFP-68 (lane 3), pEGFP-420 (lane 4), and pEGFP-497 (lane 5). Cells were analyzed as described in Fig. 1C. Arrow shows the position of the GFP-eTR1 fusion selenoprotein. (**C**) Western blot analysis of samples shown in (B) with antibodies to GFP. Arrows show the positions of GFP and truncated and full-size GFP-eTR1. Asterisks show the position of truncated GFP-eTR1 fusions in lanes 2 and 3. (**D**) Partially purified eTR sample analyzed by two-dimensional PAGE and stained with Coomassie Blue. (**E**) Visualization of the ^{75}Se-labeled sample shown in (D) with a PhosphorImager. The spots of eGR are indicated by a blue oval (D) and of eTR by red ovals (D and E).

This was likely due to termination at UGAs coding for Cys in *Euplotes*, which were recognized as stop signals in mammalian cells. We therefore prepared mutant forms of GFP-eTR1, in which six of the seven UGA codons were replaced with UGC, leaving single UGA at positions 68, 420, or 497. Of these, amino acids 68 and 420 corresponded to Cys, and 497 corresponded to Sec in other TRs. We found that ^{75}Se (and, therefore, Sec) could be inserted only at position 497 (Fig. 2B, lane 5). Western blotting confirmed the synthesis of truncated proteins when UGA was at positions 68 and 420, and of the full-size protein at position 497 (Fig. 2C). Thus, Sec was only inserted into the classical Sec site in eTR1, whereas other UGA positions were not served by SECIS for Sec insertion and instead supported termination of translation in mammalian cells (in *Euplotes,* Cys would be inserted).

To confirm Cys insertion at UGA codons other than codon 497 in eTR1, we purified the ^{75}Se-labeled 55-kD selenoprotein band from *E. crassus* after a series of chromatographic steps (Fig. 2, D and E). Liquid chromatography–tandem mass spectrometry sequencing revealed peptides corresponding to eTR1 and a more abundant glutathione reductase (eGR) (figs. S20 to S22). This analysis identified eTR1 peptides containing Cys in positions 63, 68, 208, and 270, which are encoded by UGA codons (figs. S5 and S18), whereas peptides containing Sec at these positions were not detected. Thus, UGA differentially codes for Cys and Sec in different positions within the *E. crassus* eTR1 gene.

To determine whether Cys and/or Sec insertion is associated with UGA position within the gene, we prepared GFP-eTR1 mutants containing single UGA codons in unnatural codon positions: 246, 441, 467, 478, 489, 494, or 496. ^{75}Se-labeling and Western blotting revealed that UGA terminated translation in positions 246, 441, and 467 but inserted Sec in positions 489 and 494 (Fig. 3, A to E). Sec was also inserted at position 496 (fig. S23), whereas position 478 was intermediate, supporting a low level of Sec insertion (Fig. 3, C and D). Thus, Sec insertion was restricted to approximately the last 20 codons, whereas the region upstream supported termination by UGA in mammalian cells (and, therefore, Cys insertion in *E. crassus*).

We replaced a segment corresponding to part of the eTR1 3′UTR, including the entire SECIS element, with the 3′UTR region of *Toxoplasma* selenoprotein T (SelT), which also has a SECIS element (*8*). In this mutant, Sec insertion was detected at position 420, that is, upstream of codon 478 (Fig. 3F), indicating that replacement of the functional 3′UTR region changed the coding function of UGA. Similarly, Sec could be inserted in the N-terminal region of ep22, in addition to its natural C-terminal penultimate position (Fig. 3G), which suggests a model wherein Sec insertion is dependent on an RNA structure (fig. S24).

We have demonstrated that UGA can designate different amino acids within the same gene, with the choice of the amino acid inserted determined by availability of the functional element within the 3′UTR and the location of UGA within the gene. Although dual functions of stop codons have previously been described, they support the insertion of single amino acids (e.g., Sec or pyrrolysine) in competition with termination (*9*) or ambiguous codon function due to dual specificity of a particular tRNA (*10*). Here, we show that one codon supports specific insertion of multiple amino acids, indicating that evolutionary expansion of the genetic code is possible.

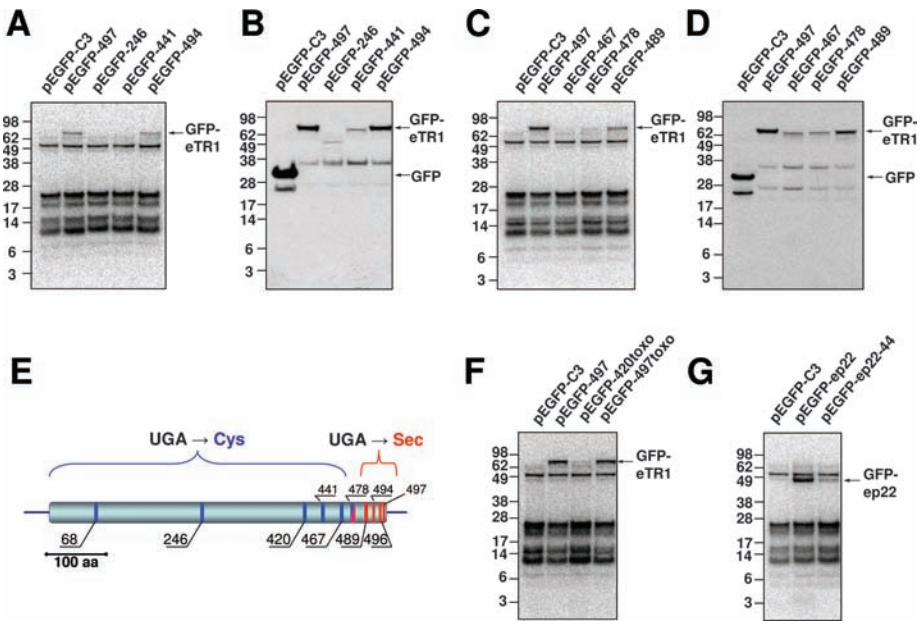

Fig. 3. Position-dependent Sec insertion in eTR1. (**A**) Expression of GFP-eTR1 in HEK 293 cells. Cells were transfected with pEGFP-C3 vector, a GFP-eTR1 construct containing a single UGA codon at the natural Sec position 497 (pEGFP-497), or constructs that had UGA at unnatural positions 246 (pEGFP-246), 441 (pEGFP-441), or 494 (pEGFP-494). Cells were analyzed as described in Fig. 1C. (**B**) Western blot analysis of samples shown in (A) with antibodies to GFP. (**C**) Cells were transfected with pEGFP-C3 vector, a GFP-eTR1 construct containing a single UGA codon at the natural Sec position 497 (pEGFP-497), or constructs that had UGA at unnatural positions 467 (pEGFP-467), 478 (pEGFP-478), or 489 (pEGFP-489). (**D**) Western blot analysis of samples shown in (C) with antibodies to GFP. (**E**) Summary of experimental evidence for Sec and Cys insertion in eTR1. The positions of Cys insertion (corresponding to termination in mammalian cells) are shown by blue lines, and Sec insertion by red lines. Position 478 supported low-level Sec insertion. (**F**) Cells were transfected with pEGFP-C3 vector, a GFP-eTR1 construct containing a single UGA codon at the natural Sec position 497 (pEGFP-497), or constructs containing a 3′UTR segment of *Toxoplasma* SelT and UGA at position 420 (pEGFP-420toxo) or 497 (pEGFP-497toxo). (**G**) Cells were transfected with pEGFP-C3 vector or with ep22 constructs in which UGA corresponded to positions 190 (pEGFP-ep22) or 44 (pEGFP-ep22-44). Arrows show the positions of GFP and full-size GFP-eTR1 or GFP-ep22.

References and Notes

1. O. Namy, J. P. Rousset, S. Napthine, I. Brierley, *Mol. Cell* **13**, 157 (2004).
2. J. F. Atkins, P. V. Baranov, *Nature* **448**, 1004 (2007).
3. J. C. Anderson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 7566 (2004).
4. S. Osawa, T. H. Jukes, K. Watanabe, A. Muto, *Microbiol. Rev.* **56**, 229 (1992).
5. F. Meyer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 3758 (1991).
6. D. L. Hatfield, V. N. Gladyshev, *Mol. Cell. Biol.* **22**, 3565 (2002).
7. M. J. Berry *et al.*, *Methods Enzymol.* **347**, 17 (2002).
8. S. V. Novoselov *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7857 (2007).
9. G. Srinivasan, C. M. James, J. A. Krzycki, *Science* **296**, 1459 (2002).
10. T. Suzuki, T. Ueda, K. Watanabe, *EMBO J.* **16**, 1122 (1997).
11. Materials and methods are available as supporting material on *Science* Online.

12. We thank K. Hammar for constructing the EST library and I. Sorokina (Midwest Bio Services) for protein sequencing. This work was supported by NIH GM061603 and GM065204 to V.N.G., AI058054 to M.L.S., NSF 0343813 to L.A.K., and the Intramural Research Program, National Cancer Institute, NIH, to D.L.H. Sequences for tRNA and selenoprotein genes have been deposited in the GenBank database under accession numbers FJ440148 to FJ440159.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5911/259/DC1
Materials and Methods
Figs. S1 to S24
References

18 August 2008; accepted 11 November 2008
10.1126/science.1164748