1-9-2009

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

Anton Turanov
*University of Nebraska-Lincoln*

Alexei Lobanov
*University of Nebraska-Lincoln*

Dmitri E. Fomenko
*University of Nebraska-Lincoln*, dfomenko2@unl.edu

Hilary Morrison
*Josephine Bay Paul Center, Marine Biological Laboratory*

Mitchell Sogin
*Josephine Bay Paul Center, Marine Biological Laboratory*

*See next page for additional authors*

# 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).

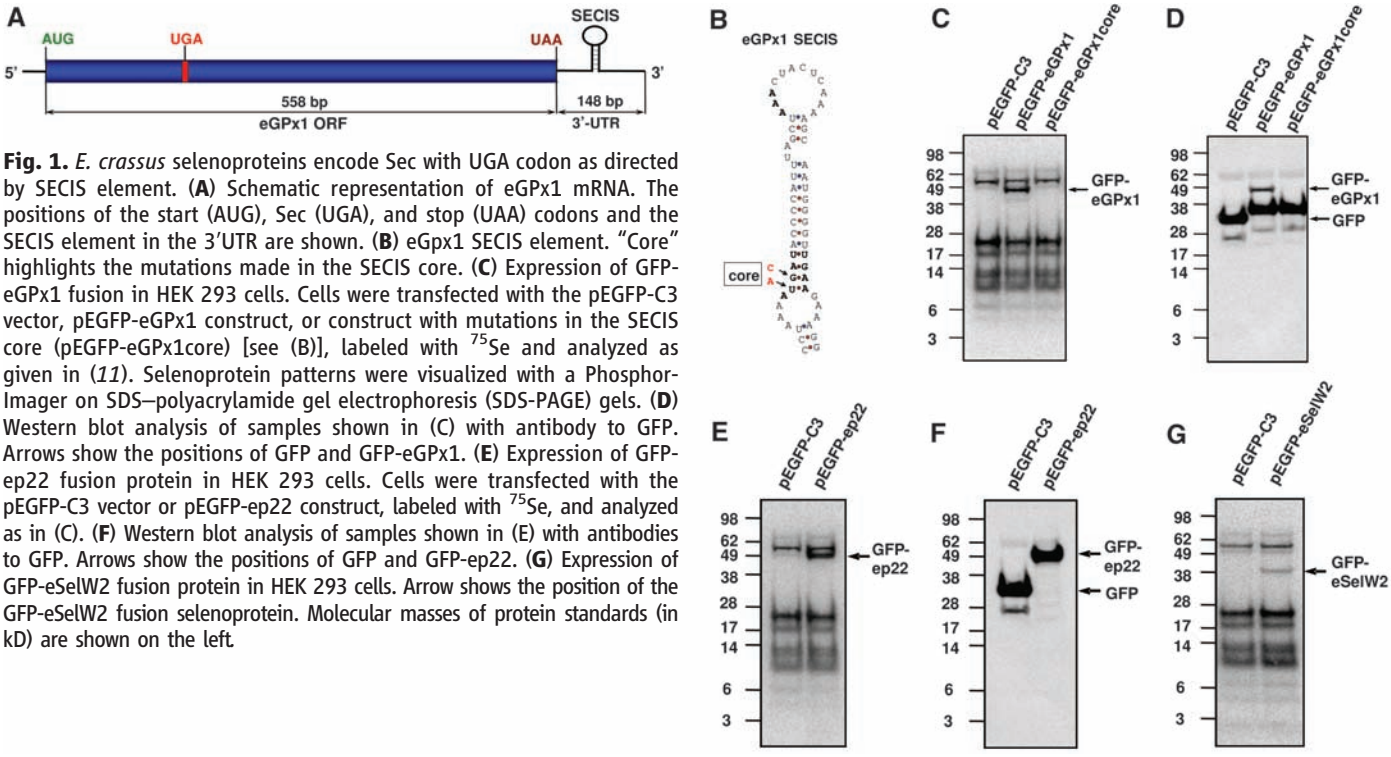



**Fig. 1.** *E. crassus* selenoproteins encode Sec with UGA codon as directed by SECIS element. (**A**) Schematic representation of eGPx1 mRNA. The positions of the start (AUG), Sec (UGA), and stop (UAA) codons and the SECIS element in the 3′UTR are shown. (**B**) eGpx1 SECIS element. "Core" highlights 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.

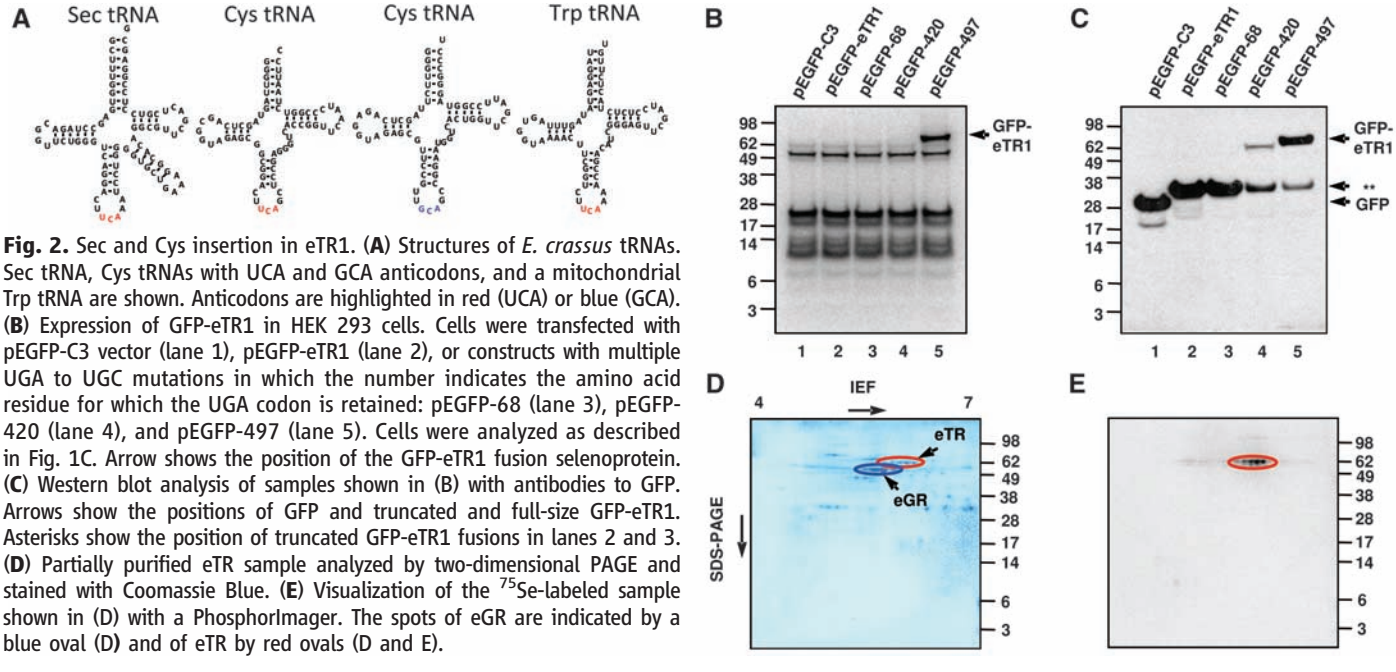



**Fig. 2.** Sec and Cys insertion in eTR1. (**A**) Structures of *E. crassus* tRNAs. Sec tRNA, Cys tRNAs with UCA and GCA anticodons, and a mitochondrial Trp tRNA are shown. Anticodons are highlighted in red (UCA) or blue (GCA). (**B**) Expression of GFP-eTR1 in HEK 293 cells. Cells were transfected 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.

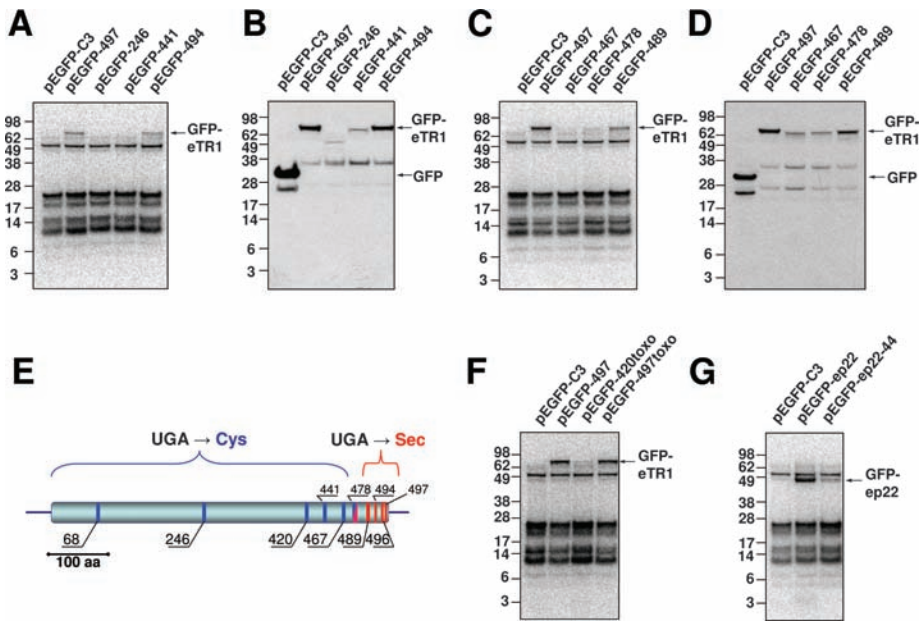



**Fig. 3.** Position-dependent Sec insertion in eTR1. (**A**) Expression of GFP-eTR1 in HEK 293 cells. Cells were transfected with pEGFP-C3 vector, a GFP-eTR1 construct containing a single UGA codon at the natural Sec position 497 (pEGFP-497), or constructs that had UGA at unnatural positions 246 (pEGFP-246), 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

# Supporting Online Material for

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

Anton A. Turanov, Alexey V. Lobanov, Dmitri E. Fomenko, Hilary G. Morrison, Mitchell L. Sogin, Lawrence A. Klobutcher, Dolph L. Hatfield, Vadim N. Gladyshev*

*To whom correspondence should be addressed. E-mail: vgladyshev1@unl.edu

Published 9 January 2009, *Science* **323**, 259 (2009)
DOI: 10.1126/science.1164748

**This PDF file includes:**

## Supporting materials

### Selenoproteins in *Euplotes*

To determine if *E. crassus* has Sec-containing proteins, we metabolically labeled this organism with $^{75}$Se, and separately labeled *Dunaliella salina*, the marine alga used as the food source for *E. crassus*. Autoradiography of *E. crassus* proteins following SDS-PAGE analysis revealed specific incorporation of radioactive Se into several proteins, and the selenoprotein pattern was similar to that observed in human HEK 293 and algal *Ostreococcus tauri* cells (Fig. S1). Humans have 25 and *O. tauri* 26 known selenoprotein genes (*S1, S2*).

### *Euplotes* genome and Sec and Cys tRNAs

The macronuclear genome of *E. crassus* is organized in the form of gene-sized chromosomes, each flanked by telomeres. To determine how widespread the use of UGA is for Cys and Sec insertion in *Euplotes*, we sequenced (*S3*) the macronuclear genome of *E. crassus* (>4x coverage), and determined coding characteristics of its codons. Consistent with previous reports (*5, S4*), we found that UGA is a common codon for Cys in this organism (e.g., we detected 445 UGA codons in ESTs and 13,206 in the genome that were predicted to code for Cys) (Fig. S2). UAA and UAG were found to be stop signals, whereas we did not detect ORFs ending with UGA. We also identified two Cys tRNAs, including one that had a UCA anticodon (i.e., corresponding to the UGA codon) (Fig. 2A). This finding shows that the previous prediction of a single Cys $tRNA_{GCA}$ serving UGA, UGC and UGU codons in *Euplotes* is incorrect (*S5*). The novel Cys $tRNA_{UCA}$ phylogenetically clustered with other Cys tRNAs (Fig. S17) suggesting it evolved by Cys tRNA gene duplication followed by a change in the anticodon. Sec tRNA also was detected. All Sec tRNAs are characterized by a long variable arm (Fig. 2A), which is missing in Cys tRNAs.

### Identification of *Euplotes* selenoproteins

As UGA codon could not be used to distinguish between Cys and Sec insertion, we relied on computational searches for SECIS elements using SECISearch (*S1*). Eight selenoprotein genes were identified, each containing a SECIS element in the 3'-UTR (Fig. S3-S16). Seven of these proteins corresponded to previously known selenoproteins and one was a new selenoprotein belonging to the Pfam09409 family (designated ep22 for *Euplotes* 22 kDa protein). Further analyses identified ep22 selenoprotein homologs in

other ciliates, *Tetrahymena* and *Paramecium*, and in a primitive animal, *Nematostella*, and in all these sequences, UGA specified Sec even though these organisms have genetic codes different than that of *Euplotes* (Fig. S15). The sequenced ep22 chromosome is illustrated in Fig. S19.

**Expression of *Euplotes* selenoproteins in mammalian cells**

To easily distinguish between Cys and Sec insertion in *Euplotes* proteins, we employed a mammalian expression system where UGA is a default stop signal and can also support SECIS-dependent Sec insertion. Due to competition of Sec insertion with termination of protein synthesis and low level of Sec insertion machinery, selenoprotein expression is accompanied by truncated protein forms.

**Purification of natural *Euplotes* selenoproteins**

Like mammalian TRs (*S6*), an abundant 55 kDa *Euplotes* selenoprotein tightly bound ADP-Sepharose and occurred in multiple spots on 2D gels (Fig. 2D, E). It was identified as eTR1 via tryptic digest and LC-MS/MS. We also partially purified eSelW2 (the major 8 kDa selenoprotein in *Euplotes*) and verified tryptic peptides from its sequence (Fig. S8, S22).

**Mechanism of alternative use of UGA codon for Sec and Cys insertion into eTR1**

It appears that eTR1 mRNA is organized such that its SECIS is unavailable for Sec insertion in most of the eTR1 sequence (Fig. S24A). However, once the translating ribosome approaches the last 20 codons in eTR1, the mRNA is restructured and the SECIS can now support Sec insertion (Fig. S24B).

Taken together, our data show that *E. crassus* utilizes UGA for insertion of both Cys and Sec, establishing it as the first known organism that utilizes one codon to code unambiguously for two different amino acids. The example of *E. crassus* shows that the genetic code can be naturally extended by recoding a subset of codons in an organism for insertion of a different amino acid by utilizing RNA elements within 3'-UTRs and controlling their availability for the translating ribosome. This mechanism could support both regulation of protein synthesis and addition of new or modified amino acids to the genetic code. For example, since Sec is made from phosphoserine on Sec tRNA, the *Euplotes* genes could be adapted for phosphoserine insertion into specific positions of protein. Finally, as the possibility of the use of one codon to code two amino acids has not been previously considered, it would be extremely important to determine whether this situation is widespread in organisms with completely sequenced genomes as well as generally in nature.

[Supplementary material is included following the article.]

## Supporting Methods

We constructed an *E. crassus* CT5 cDNA library from total RNA isolated from cells growing in the log phase, with *Dunaliella salina* as a food source. We used the Clontech (Mountainview, CA) SMART cDNA library construction protocol and pDNR-LIB vector. Recombinant clones were picked into 96-well polypropylene 2.0-ml-deep well growth blocks containing 1.2 ml Superbroth supplemented with 30 µg/ml kanamycin and grown for 18 h at 37°C. Plasmid DNA was isolated by alkaline lysis on a Beckman Biomek FX platform. We end-sequenced clones by dideoxy chain termination using 3 µl of cleaned template and BigDye Terminator reagents (Applied Biosystems, Foster City, CA). Sequencing reactions were cleaned by isopropanol precipitation, resuspended in 7 µl of formamide, and sequenced on an Applied Biosystems 3730XL 96 capillary array sequencer. Approximately 15,000 ESTs were generated from 8,400 clones, representing nearly 2,000 distinct transcripts.

*E. crassus* strain CT5 ESTs were analyzed for selenoprotein genes by similarity against all known selenoproteins, and by SECISearch for occurrence of SECIS elements (*S1*). A partial sequence read of eGPx1 was identified, and complete sequencing of the corresponding clone indicated that it represented a full-length cDNA. All *E. crassus* ESTs were also analyzed for occurrence of UGA, UGU and UGC codons within ORFs, and the incidence of each codon was calculated.

The *E. crassus* macronuclear genome was sequenced at the 454 Life Science (www.454.com) facilities. A total of 836,065 reads (201,591,529 bp) were obtained, with an average read length of 241.1 bp. Sequences were assembled into 66,728 contigs covering 34 Mbp. Average depth of all contigs was 4.07. Both raw data (reads) and contigs were used in the computational analysis. Detailed analysis of the *E. crassus* genome will be published elsewhere.

tRNA prediction was carried out using tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/) with default settings and ARAGORN (http://130.235.46.10/ARAGORN1.1/HTML/aragornA.html). A search for Sec tRNA also included BLAST searches against known Sec tRNA sequences and use of tRNAscan-SE under "maximum sensitivity" mode. To analyze for codon frequency, EST and genome contigs were translated in 6 frames. Each candidate ORF was checked for homology to conserved protein domains with RPS-BLAST with an E-value cutoff $1e^{-6}$. ORFs containing conserved domains were analyzed for codon frequency.

Since UGA codes for both Sec and Cys in this organism, the presence of in-frame UGA does not provide sufficient criterion for recognition of selenoproteins. Therefore, another characteristic feature of selenoproteins, the presence of a SECIS element in the 3'-UTR, was used as the main criterion for recognition of selenoprotein genes. Selenoprotein searches were conducted as follows:

1) A stand-alone version of SECISearch (*S1*) was used for SECIS element identification.
2) Upstream regions of candidate SECIS elements were analyzed for similarity to known selenoproteins and to proteins from NCBI non-redundant database.
3) SECIS candidates lacking translated regions with similarity to known proteins were dismissed. Some of these sequences were partial. Therefore, some of the candidate structures might be true SECIS elements.

The searches utilized the PrairieFire Beowulf cluster at the Research Computing Facility, University of Nebraska – Lincoln.

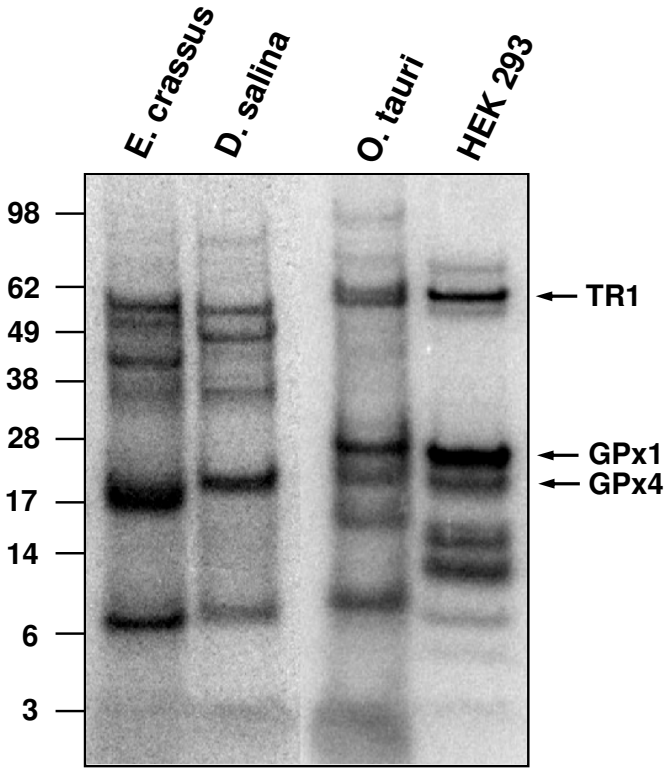
To experimentally examine *Euplotes* for the occurrence of selenoproteins, 50 ml of *E. crassus* strain CT5, as well as (as control) *D. salina*, *O. tauri* (cultured in artificial sea water) and a 10 cm plate of HEK 293 cells (cultured in DMEM medium supplemented with 10% FBS, 100 units/ml penicillin, and 100 units/ml streptomycin (Invitrogen)) were metabolically labeled with 50 μCi of $^{75}$Se ([$^{75}$Se]selenious acid (specific activity, 1,000 Ci/mmol) Research Reactor Facility, University of Missouri , Columbia, Mo.) for 48 h. Cells were collected, resuspended in PBS and sonicated. 30 μg of total soluble protein from each organism were resolved by SDS-PAGE and transferred onto a PVDF membrane (Invitrogen). Selenoprotein patterns were visualized with a PhosphorImager (*S1*).

Full-length eGPx1, ep22, eTR1, and eSelW2 cDNAs, including their 3'-UTRs, or genes, were amplified by PCR from the 077B06 EST clone (contig CL1180C1) for eGPx1 and from *E. crassus* whole cellular DNA for ep22, eTR1 and eSelW2 with Pfu Ultra DNA polymerase (Stratagene). The PCR fragments were cloned into pGEM-Teasy (Promega) and pEGFP-C3 vector (Clontech, BD Biosciences). The QuikChange™ Site-Directed Mutagenesis Kit (Stratagene) was used to generate mutations in the SECIS element of eGPx1 and mutations of in-frame codons in the eTR1 gene. Similarly, mutations were made to introduce UGA in positions 246, 441, 467, 478, 489, 494, 496 and replace the natural UGA with UGC in position 497 in the eTR1 gene. In addition, UGA was introduced into codon position 44 in the ep22 gene. A segment of *Toxoplasma* SelT 3'-UTR containing the SECIS element was amplified by PCR primer extension and cloned into the eTR1 gene using pEGFP-420 and pEGFP-497 eTR1 constructs as templates. HEK 293 cells, cultured in DMEM medium supplemented with 10% FBS, 100 units/ml penicillin, and 100 units/ml streptomycin (Invitrogen) were transfected with the resulting constructs using the calcium phosphate method. 24 h after transfection, cells on 10 cm plates were labeled by supplementing the medium with 50

μCi of $^{75}$Se for an additional 48 h. Cells were collected, resuspended in PBS and sonicated. 30 μg of total soluble protein from each transfection were resolved by SDS-PAGE and transferred onto a PVDF membrane. Selenoproteins were visualized with a PhosphorImager. Western blotting of the GFP-fusion proteins in HEK 293 cells was performed with anti-GFP antibodies (Invitrogen). The PVDF membrane was incubated with anti-GFP antibody and GFP and GFP fusion proteins were visualized with an ECL kit (GE Healthcare).

To isolate selenoproteins, 50 ml of *E. crassus* cells were labeled with 50 μCi/g of $^{75}$Se for 72 h and the labeled cells mixed with 4 g of unlabeled cells that were cultured separately using the same procedure. Cells were washed twice in cold PBS and resuspended in PBS containing EDTA-free protease inhibitor mixture (Roche). eSelW2 was isolated on a DEAE-Sepharose column, and eTR1 was isolated on DEAE-Sepharose and then ADP-Sepharose (GE Healthcare) columns, following $^{75}$Se radioactivity in protein fractions. Based on the observed elution properties of selenoproteins, we also prepared unlabeled fractions enriched for selenoproteins (for LC-MS/MS analysis). Resulting protein fractions (100 μg of total protein) were subjected to two-dimensional electrophoresis separation. ReadyStrip IPG strips (Bio-Rad) were used for isoelectrofocusing and NOVEX gels for SDS-PAGE (Invitrogen) according to the manufacturer's protocols. Proteins were transferred onto a PVDF membrane (Invitrogen) and visualized by Coomassie Blue (Bio-Rad) staining. Sequences of proteolytic peptides of unlabeled *Euplotes* proteins were determined by LC-MS/MS at Midwest Bio Services and the proteomics facility of the Redox Biology Center, University of Nebraska.

**Supporting Figure S1. Metabolic labeling of *E. crassus*, *D. salina*, *O. tauri* and human HEK 293 cells with $^{75}$Se-selenite.** 30 µg of total protein from each organism were resolved by SDS-PAGE and transferred onto a PVDF membrane. Selenoprotein patterns were visualized with a PhosphorImager. *Arrows* indicate the positions of human thioredoxin reductase 1 (TR1), glutathione peroxidase 1 (GPx1) and glutathione peroxidase 4 (GPx4).



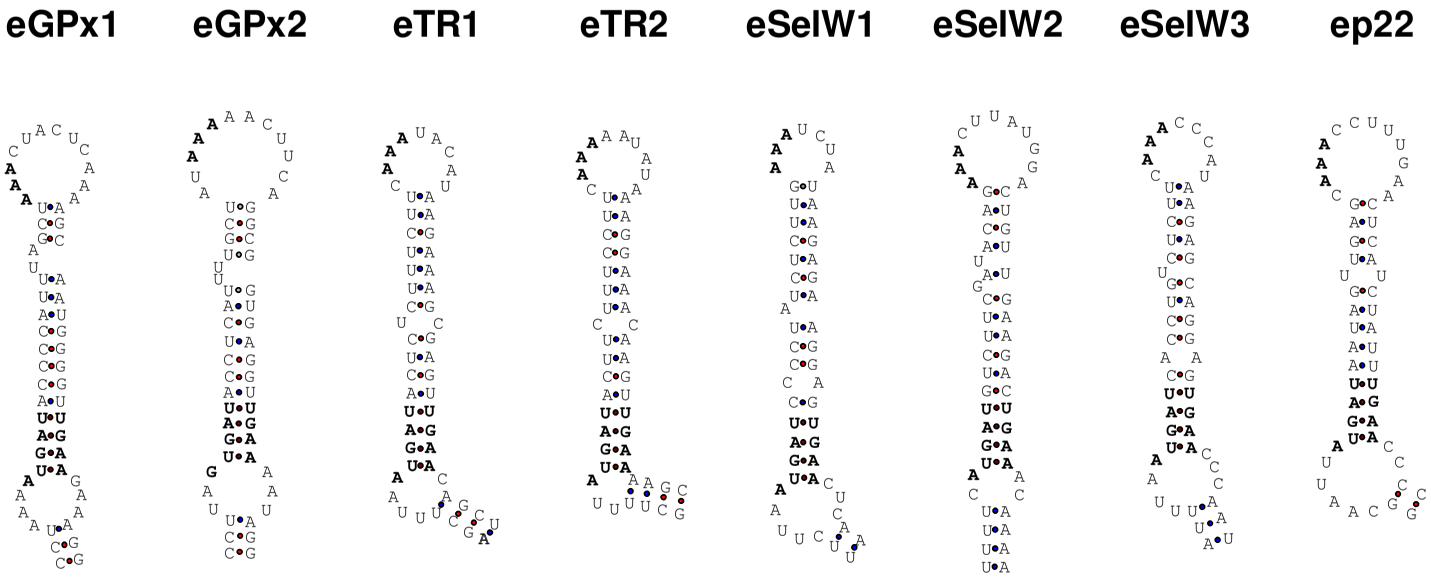
**Supporting Figure S2. Codon usage and frequencies based on EST and genome sequences.** The use of UGA is shown in red.

| AA: | EST | Genome |
|---|---|---|
| **F** | | |
| UUU | 3389 (2.0%) | 46392 (2.2%) |
| UUC | 4799 (2.8%) | 52324 (2.5%) |
| **L** | | |
| UUA | 2003 (1.1%) | 29936 (1.4%) |
| UUG | 2604 (1.5%) | 30136 (1.4%) |
| CUU | 3554 (2.1%) | 47623 (2.3%) |
| CUC | 2889 (1.7%) | 33908 (1.6%) |
| CUA | 1122 (0.6%) | 21693 (1.0%) |
| CUG | 1125 (0.6%) | 20048 (0.9%) |
| **I** | | |
| AUU | 5223 (3.1%) | 67068 (3.2%) |
| AUC | 4533 (2.7%) | 54426 (2.6%) |
| AUA | 1488 (0.9%) | 28870 (1.4%) |
| **M** | | |
| AUG | 4339 (2.6%) | 49185 (2.4%) |
| **V** | | |
| GUA | 1943 (1.1%) | 24960 (1.1%) |
| GUG | 1308 (0.9%) | 18087 (0.9%) |
| GUU | 3506 (2.1%) | 37123 (1.8%) |
| GUC | 2972 (1.8%) | 28323 (1.4%) |
| **S** | | |
| UCA | 2581 (1.5%) | 35817 (1.7%) |
| UCG | 408 (0.2%) | 7639 (0.4%) |
| UCU | 2634 (1.5%) | 40565 (1.9%) |
| UCC | 1592 (0.9%) | 17823 (0.8%) |
| AGU | 1556 (0.9%) | 25254 (1.2%) |
| AGC | 1630 (0.9%) | 20836 (1.0%) |
| **W** | | |
| UGG | 1807 (1.1%) | 20160 (1.0%) |

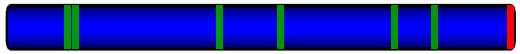
| AA: | EST | Genome |
|---|---|---|
| **P** | | |
| CCA | 3534 (2.1%) | 34051 (1.6%) |
| CCG | 293 (0.2%) | 4276 (0.2%) |
| CCU | 1901 (1.1%) | (29728 (1.4%) |
| CCC | 598 (0.3%) | 8667 (0.4%) |
| **T** | | |
| ACA | 1935 (1.1%) | 30730 (1.5%) |
| ACG | 278 (0.2%) | 5017 (0.2%) |
| ACU | 4092 (2.4%) | 45694 (2.2%) |
| ACC | 2504 (1.5%) | 21493 (1.0%) |
| **A** | | |
| GCA | 2517 (1.5%) | 31806 (1.5%) |
| GCG | 275 (0.2%) | 4493 (0.2%) |
| GCU | 5109 (3.0%) | 48893 (2.3%) |
| GCC | 2531 (1.5%) | 19566 (0.9%) |
| **Y** | | |
| UAU | 3211 (1.9%) | 44247 (2.1%) |
| UAC | 3113 (1.8%) | 33891 (1.6%) |
| **UAA SUOP** | | |
| UAA | 581 (0.3%) | 5747 (0.3%) |
| **UAG SUOP** | | |
| UAG | 109 (0.1%) | 2057 (0.1%) |
| **UGA Sec/Cys** | | |
| **UGA** | **445 (0.3%)** | **13206 (0.6%)** |
| **H** | | |
| CAU | 2131 (1.3%) | 29117 (1.4%) |
| CAC | 1482 (0.9%) | 15329 (0.7%) |
| **Q** | | |
| CAA | 3309 (1.9%) | 44881 (2.1%) |
| CAG | 2219 (1.3%) | 30538 (1.5%) |

| AA: | EST | Genome |
|---|---|---|
| **N** | | |
| AAU | 4140 (2.4%) | 64286 (3.1%) |
| AAC | 4062 (2.4%) | 46994 (2.2%) |
| **K** | | |
| AAA | 6384 (3.8%) | 93376 (4.5%) |
| AAG | 8029 (4.7%) | 89123 (4.3%) |
| **D** | | |
| GAU | 6466 (3.8%) | 79897 (3.8%) |
| GAC | 4058 (2.4%) | 43684 (2.1%) |
| **E** | | |
| GAA | 7752 (4.5%) | 100748 (4.8%) |
| GAG | 5165 (3.1%) | 63905 (3.1%) |
| **C** | | |
| UGU | 1171 (0.7%) | 15327 (0.7%) |
| UGC | 1237 (0.7%) | 12701 (0.6%) |
| **R** | | |
| CGA | 135 (0.1%) | 4545 (0.2%) |
| CGG | 150 (0.1%) | 2684 (0.1%) |
| CGU | 276 (0.2%) | 4199 (0.2%) |
| CGC | 120 (0.1%) | 2442 (0.1%) |
| AGA | 5401 (3.2%) | 57712 (2.8%) |
| AGG | 1173 (0.7%) | 20305 (1.0%) |
| **G** | | |
| GGU | 2479 (1.5%) | 23908 (1.1%) |
| GGC | 1507 (0.9%) | 17529 (0.8%) |
| GGA | 6883 (4.1%) | 60322 (2.9%) |
| GGG | 1157 (0.7%) | 13672 (0.7%) |

**Supporting Figure S3. SECIS elements identified in the *Euplotes* genome.** Functionally important nucleotides are shown in bold. GPx, glutathione peroxidase; SelW, selenoprotein W; TR, thioredoxin reductase.



**Supporting Figure S4. Selenoproteins identified by searches for SECIS elements.** Organization of selenoproteins is shown on the right side of the panel. Location of Sec encoded by UGA is indicated by red lines and Cys encoded by UGA by green lines.

| Selenoprotein | Sec coded by UGA | Cys coded by UGA | Distance from Sec to SECIS core | Location of UGA encoded residues |
|---|---|---|---|---|
| Glutathione peroxidase 1 (eGPx1) | 1 | 0 | 448 |  |
| Glutathione peroxidase 2 (eGPx1) | 1 | 1 | 449 |  |
| Thioredoxin reductase 1 (eTR1) | 1 | 6 | 145 |  |
| Thioredoxin reductase 2 (eTR2) | 1 | 6 | 131 |  |
| Selenoprotein W1 (eSelW1) | 1 | 0 | 231 |  |
| Selenoprotein W2 (eSelW2) | 1 | 0 | 288 |  |
| Selenoprotein W3 (eSelW3) | 1 | 1 | 285 |  |
| Pfam09409 protein (ep22) | 1 | 0 | 92 |  |

[Supplementary material is included following the article.]

**Supporting Figure S5. Nucleotide and amino acid sequences of Euplotes thioredoxin reductase 1 (eTR1).** Initiator AUG codon is shown in green; the Sec-encoding UGA and the stop UAA codon are in bold red. The UGA codons that may correspond to Cys are shown in red. The sequences corresponding to the SECIS element are underlined and shown in blue. In the protein sequence, the initiator (M) and Sec (U) amino acid residues are indicated. The cysteine (C) amino acids encoded by UGA codons are highlighted in red, by UGU in blue and by UGC in green.

```
GTTTATAACTATAAAATATCAAATTTAGAAATGGACTATTCAGACACTCCACAAGAAGAATCCACTCATAGTTATGACTATGAT
                                  M  D  Y  S  D  T  P  Q  E  E  S  T  H  S  Y  D  Y  D

CTTTTTGTAATCGGAGGTGGTTCTGGAGGGCTTGCTTGTGCCAAGGTTGCTCAAGAGGCAGGAGCTAAAGTAGCAGTAGCAGAT
 L  F  V  I  G  G  G  S  G  G  L  A  C  A  K  V  A  Q  E  A  G  A  K  V  A  V  A  D

TTTGTAAAGCCAACTCCAAAGGGAACAAAGTGGAAAGTAGGAGGAACATGAGTGAATGTTGGTTGAATCCCCAAAAAGCTGATG
 F  V  K  P  T  P  K  G  T  K  W  K  V  G  G  T  C  V  N  V  G  C  I  P  K  K  L  M

CACTACTCCGCATTGTTAGGAAATTCATATCACGACCAAGTTGAGAGCGGATGGGAGCATGAGAAACCTTCTCATGACTGGGGT
 H  Y  S  A  L  L  G  N  S  Y  H  D  Q  V  E  S  G  W  E  H  E  K  P  S  H  D  W  G

AAAATGATTACCAATGTCAATAACCATATAAGAGGTATCAATTTTGGATACAAAGCAGATATGAGAAAGAGAGGTATAAAGTTC
 K  M  I  T  N  V  N  N  H  I  R  G  I  N  F  G  Y  K  A  D  M  R  K  R  G  I  K  F

CATGAAAAGTTTGCCTCCTTTGTCGATCCTCATACCGTACAACTAGTTGATAAGAAGGGCAAGACCGAAATGATTACTTCTAAT
 H  E  K  F  A  S  F  V  D  P  H  T  V  Q  L  V  D  K  K  G  K  T  E  M  I  T  S  N

TATTTCGTAATTGCTACTGGAGGCAGACCTCTCTATCCTGATATTCCAGGAGCCAAAGAGCATGCAATTACTAGTGATGATATT
 Y  F  V  I  A  T  G  G  R  P  L  Y  P  D  I  P  G  A  K  E  H  A  I  T  S  D  D  I

TTCTGGATGAAAGACAACCCTGGTAAAACCCTTGTGGTAGGTGCTTCCTATGTCGCGTTGGAATGAGCTGGATTTTTACATCAT
 F  W  M  K  D  N  P  G  K  T  L  V  V  G  A  S  Y  V  A  L  E  C  A  G  F  L  H  H

TTTGGAAACGAAGTTTCAGTTTGTGTCCGATCAATCTTTTTGAGAGGTTTCGATCAAGATATGGCGCAAAAGATTGCTAAAGAC
 F  G  N  E  V  S  V  C  V  R  S  I  F  L  R  G  F  D  Q  D  M  A  Q  K  I  A  K  D

ATGGAACTCAGCGGGATTAATTTCATTAGAGACTCTATGCCTACCAAAATTGAGAAAGGCGAAGAGACTGGCAAGCTCACCTGA
 M  E  L  S  G  I  N  F  I  R  D  S  M  P  T  K  I  E  K  G  E  E  T  G  K  L  T  C

TTTTTAACAGTAGGAGGCGAAGAAACTACCGTTGAAGTAGATACTGTTCTTTTTGCAATTGGTAGATATGCTGTGACAGCTGAT
 F  L  T  V  G  G  E  E  T  T  V  E  V  D  T  V  L  F  A  I  G  R  Y  A  V  T  A  D

TTAAATCTAGGTAATGCTGGACTCATTGCTGAAAAGAATGGAAAATTCATTACTGACAAATACCAGAAAACTAATGTTGACAAT
 L  N  L  G  N  A  G  L  I  A  E  K  N  G  K  F  I  T  D  K  Y  Q  K  T  N  V  D  N

ATCTATGCTATAGGGGATGTGCTTCATGGAAAATTGGAACTTACTCCAACTGCAATTCAAGCGGGAAGACTATTGGCTGATAGA
 I  Y  A  I  G  D  V  L  H  G  K  L  E  L  T  P  T  A  I  Q  A  G  R  L  L  A  D  R

CTATTTGCTGGAGGAACTACTACAATGGATTTTTATGATGTTCCTACTACTATCTTTACTCCTCTTGAGTATGGATGAGTAGGT
 L  F  A  G  G  T  T  T  M  D  F  Y  D  V  P  T  T  I  F  T  P  L  E  Y  G  C  V  G

TACTCAGAGGAAGATGCTAGGGAGGAATATGGTGACTTCATCAAAGTTTACCATACTTATTTCCAGCCATTAGAATGGAACTTT
 Y  S  E  E  D  A  R  E  E  Y  G  D  F  I  K  V  Y  H  T  Y  F  Q  P  L  E  W  N  F

GCAAAATCAATCTATAAGGAGAGGAATTGATACGTAAAAATTATAGTAAACACTGCAGATAATGACAGAGTGATCGGGTTCCAT
 A  K  S  I  Y  K  E  R  N  C  Y  V  K  I  I  V  N  T  A  D  N  D  R  V  I  G  F  H

ATTCTCTGTCCTAATGCTGGGGAGATAACACAAGGAATTGCTATTGCCATCAAAGTAGGAGTCACAAAGCCTCAGTTAGATAAC
 I  L  C  P  N  A  G  E  I  T  Q  G  I  A  I  A  I  K  V  G  V  T  K  P  Q  L  D  N

TGTGTTGGAATTCATCCTACAATTGCTGAAGAAATGACTAATCTACATATTGATAAAGCTGATAATCCAGATCCAATTAAATCG
 C  V  G  I  H  P  T  I  A  E  E  M  T  N  L  H  I  D  K  A  D  N  P  D  P  I  K  S

GATTGCTGATCTTAAAGAGTCTGTGCATATTAGGAGTTGCTAAAATTTAGTCAACTACTAATAACTGCATGGGAAAAATATAGT
 D  C  U  S

CGGTAAAGGAGACAAATCAGTAAGCAATGTTTCCAGACAATATGCTTGGTCTTTAGAAGCATTAGCTTTAATGATACTCTCTTT
CTTCAAATACATAAGAAAGCGAGTTGAACAGCTGAGGACTCTCTGAGTAGTAATACATCG
```

[Supplementary material is included following the article.]

**Supporting Figure S6. Nucleotide and amino acid sequences of Euplotes thioredoxin reductase 2 (eTR2).** Initiator AUG codon is shown in green; and Sec-encoding UGA and stop UAG codons are in bold and red. UGA codons that correspond to Cys are shown in red. The SECIS element is shown in blue and underlined. In the protein sequence, initiator (M) and Sec (U) residues are indicated. Cysteine (C) encoded by UGA codons are highlighted in red, and by UGU in blue.

```
CCCCCAAAACCCCAAAACCCCCAAAACCCCCTTTCAAAAATATAAGCGTAGAAGTTCACATGAATCTATAATTTATAAGAATG
GCTGACTCTGAAATGAAGAGTAATGAGTCAGAAGAATATGAAGAGACTAAGCGTCGTTATGATTATGATCTCTTCGTCATTGGT
              M  K  S  N  E  S  E  E  Y  E  E  T  K  R  R  Y  D  Y  D  L  F  V  I  G

GGAGGATCAGGAGGACTTGCTTGTGCTAAAGCAGCTCAAGAGTGTGGAGCAAAGGTAGCTGTAGCTGACTTTGTCAAGCCCTCT
 G  G  S  G  G  L  A  C  A  K  A  A  Q  E  C  G  A  K  V  A  V  A  D  F  V  K  P  S

CCTCATGGATATGGAGTTGTGACTTGGGGAGTTGGTGGAACTTGTGTCAATGTGGGATGAATACCAAAGAAATTATTACATTAT
 P  H  G  Y  G  V  V  T  W  G  V  G  G  T  C  V  N  V  G  C  I  P  K  K  L  L  H  Y

TCAGCAAATTTGGGAGAAGCTTATGTTGATAGAGCTAGTAGTGGATGGGACCATGAGAAGCCAAAACACGATTGGGGTAAAATG
 S  A  N  L  G  E  A  Y  V  D  R  A  S  S  G  W  D  H  E  K  P  K  H  D  W  G  K  M

ATTTCTAATATTAATAATCATATTCGAGCTATTAATTTTAGCATCAAAACTGATTTGAGGAAGAGAGGAATAAAATTTTATGAA
 I  S  N  I  N  N  H  I  R  A  I  N  F  S  I  K  T  D  L  R  K  R  G  I  K  F  Y  E

AAATTAGCTTCTTTTGCTGATCCACATACTATTCAACTTTTAAACAAGAAAGGCAAGACAGAATTAGTGACAGCAAATCATATT
 K  L  A  S  F  A  D  P  H  T  I  Q  L  L  N  K  K  G  K  T  E  L  V  T  A  N  H  I

GTTATTGCAACTGGGGGAAGGCCTCTCTACCCTGATATCCCTGGAGCAAAGGAGTATGGTATTACAAGCGATGACATTTTCTGG
 V  I  A  T  G  G  R  P  L  Y  P  D  I  P  G  A  K  E  Y  G  I  T  S  D  D  I  F  W

CTGAAGAAAAATCCAGGTAAAACCTTGGTCATTGGCGCATCTTATATTGCACTTGAATGAGCTGGATTTTTACATAGTTTTGGT
 L  K  K  N  P  G  K  T  L  V  I  G  A  S  Y  I  A  L  E  C  A  G  F  L  H  S  F  G

AACGATGTTTCTGTGTGAGTAAGATCGGTCTTTTTGCGGGGCTTTGATCAGGATATGGCTAATATGCTTGCCAAGGATATGGAA
 N  D  V  S  V  C  V  R  S  V  F  L  R  G  F  D  Q  D  M  A  N  M  L  A  K  D  M  E

GAACATGGTGGAGTCAAATTCATTAAAAATTCAATACCTACCAAAATCGAAAAAGATGAAGAAACAGGAAAGCTCATATGATAT
 E  H  G  G  V  K  F  I  K  N  S  I  P  T  K  I  E  K  D  E  E  T  G  K  L  I  C  Y

CTCACCTCTAGAGAAGAGGAAATTACTATAGAAGTTGACACAGTTTTGTTTGCAATTGGTAGATATGCTGTTACAAAAGATCTA
 L  T  S  R  E  E  E  I  T  I  E  V  D  T  V  L  F  A  I  G  R  Y  A  V  T  K  D  L

AACCTTGAAAATGCGGGTCTCAAAGTAGAATCAAACGGTAAATTCATTACAGATGAGTTTCAACAAACTAATGTGGAGAATATC
 N  L  E  N  A  G  L  K  V  E  S  N  G  K  F  I  T  D  E  F  Q  Q  T  N  V  E  N  I

TATGCTATCGGAGATGTGATTCATGGGAAATTAGAACTAACACCCACTGCAATTCAAACAGGTAAACTACTTGCAAGAAGATTG
 Y  A  I  G  D  V  I  H  G  K  L  E  L  T  P  T  A  I  Q  T  G  K  L  L  A  R  R  L

TATGCTGGTGAAACCACAACTATGGACTTTTGTGATATTCCAACTACAATCTTCACTCCTTTAGAGTATGGATGAGTTGGATAC
 Y  A  G  E  T  T  T  M  D  F  C  D  I  P  T  T  I  F  T  P  L  E  Y  G  C  V  G  Y

TCAGAAGAAGAAGCTAAGGAAAAATATGGAGACGCCATTAAGGTATATCATACTTACTTCAAGCCATTAGAGTGGAACTATGCA
 S  E  E  E  A  K  E  K  Y  A  G  D  A  I  K  V  Y  H  T  Y  F  K  P  L  E  W  N  Y

AAATCAATTTATAAATATCGAAATTGATATGTTAAAGTAATTATAAACACTACAGAGAATGATCGGGTAATTGGCTATCATTTA
 K  S  I  Y  K  Y  R  N  C  Y  V  K  V  I  I  N  T  T  E  N  D  R  V  I  G  Y  H  L

TTGGCTCCAAATGCAGGAGAAATTACTCAAGGAATTGCAATTGCCATTAAGATTGGCCTTACTAAACACAAGTTAGATAACTGT
 L  A  P  N  A  G  E  I  T  Q  G  I  A  I  A  I  K  I  G  L  T  K  H  K  L  D  N  C

GTTGGAATCCATCCAACTGTTGCAGAAGAAGTAACGGATCTCAAGATTGATAAAGCAATCAATCCTGATCCAGTCAAGACAGAT
 V  G  I  H  P  T  V  A  E  E  V  T  D  L  K  I  D  K  A  I  N  P  D  P  V  K  T  D

TGTTGATCTTAGAGCATCTGTACATATTAGTCATTGCTTTTACAAAGGCAATCACTAATAGCAGTACGGAATAATTAATCTGAA
 C  U  S

AAGGGGAGGACCCAGTAAGCTATAGATCATACTCAAAATGAGCTATAGCTTTTATGATACTTCTTTCCTTCAAAAATATAAAGG
AAACAAGTTGAAAAGCTGGTAATTCCCTGAGCAGCTTCAGCAATAATTGTTTTTAAATTTTAAGTGTAGGTGTGATTCATCGGT
TATTGGAGATTCAATACAAAAGTTGGAGAAAGGGGTTTTGGGGTTTTGGGGTTTTGGGGG
```

[Supplementary material is included following the article.]

**Supporting Figure S7. Nucleotide and amino acid sequences of Selenoprotein W1 (eSelW1).** Initiator AUG codon is shown in green; and the Sec-encoding UGA and stop UAA codons are bolded and shown in red. The sequence corresponding to the SECIS element is shown in blue and underlined. In the protein sequence, initiator (M) and Sec (U) residues are highlighted. Cysteines (C) encoded by TGC codons are highlighted in green.

```
CGCGATCTTTCTGAATATACCATGAATATTCAGTACTGCGGAGGCTGATCCTATCGCCCCAAAGCTGTCTTTGTCCAGAAGGAT
                      M  N  I  Q  Y  C  G  G  U  S  Y  R  P  K  A  V  F  V  Q  K  D

ATTAAAAAGCACTTTGGTGGAAAAATTAACGTCGTCTTTGACAGAGACTCATCACTGACTGGTAACTTTGAGATTACCATCACT
 I  K  K  H  F  G  G  K  I  N  V  V  F  D  R  D  S  S  L  T  G  N  F  E  I  T  I  T

GATAAGAAGACCGGGAAATCCCAACTTCTTCACAGTAAGAAGAACGGGGACGGTTTTGTAGAGAAAGGAACTATTGATGACTTC
 D  K  K  T  G  K  S  Q  L  L  H  S  K  K  N  G  D  G  F  V  E  K  G  T  I  D  D  F

AGAGAGAAAGTCCAGAAGTTCTGCTCTTCCTAATCACTCATTCTAAGGAAAGTTCTGTTTCTAGATCTCTGCGGCATATTCTTA
 R  E  K  V  Q  K  F  C  S  S

ATGATCCCCTATCTCTTGAAATCTATAAGAGAAGGAGTGAACTCAA
```

[Supplementary material is included following the article.]

**Supporting Figure S8. Nucleotide and amino acid sequences of Selenoprotein W2 (eSelW2).** Initiator AUG codon is shown in green; and Sec-encoding UGA and stop UAA codons are in red and bolded. The sequence corresponding to the SECIS element is shown in blue and underlined. In the protein sequence, initiator (M) and Sec (U) residues are indicated. Cysteine (C) encoded by UGU is highlighted in blue.

```
CCCAAAACCCCAAAACCCCAAAACCCCTTTTAAATAAGAAATTTGGGTAGATTATTAATAATAATGGATTCAACTACAAAGGGC
                                                              M  D  S  T  T  K  G

CACATTGTGGTCAATTACTGTGGAGGTTGAGGATACCTTCCGAAAGCCCGTTACGTACAAGAGGCTGTCGAAAATAGATTCCCA
 H  I  V  V  N  Y  C  G  G  U  G  Y  L  P  K  A  R  Y  V  Q  E  A  V  E  N  R  F  P

GGAGATTTCTCCTTTGATCTCAAAGCTGATGTAGGCAAGACCGGAAGACTTGAGGTTACAGTTTTCGTTGGTGATGATACCGAA
 G  D  F  S  F  D  L  K  A  D  V  G  K  T  G  R  L  E  V  T  V  F  V  G  D  D  T  E

GGAAAGCTTGTCCACTCCAAGGATAAAGGTCAAGGCTTCGTCAAAGATTCAAACGTTGATTCAGTACTTGATTCAATCGCCGCT
 G  K  L  V  H  S  K  D  K  G  Q  G  F  V  K  D  S  N  V  D  S  V  L  D  S  I  A  A

CTCTTAGAGTAAATCATGAGCTCTAAGAGCTAAAGGCTGAAGGTAGTAGATATTCTCCCAGGTAGCCAGTATCCTTTTCATGAT
 L  L  E

GTCTTCGATACAGAAACTTATGGACTGTTGAAGACTGAAACAAAAGATTTTACTGTCTCTTTTGAGAGGAAAGTACTCCTCTGG
AGTGCTTTTTGTTCAGGTGTGAGAATACATGAACCCTGAGGCTTCCCAAATAGATATTCAAATGTTAAAGTTACTTACAAGTAG
TGTCAAAATTAGTAGAAGGATCCTGGGCAATACAGACCTAGATCAAGGTAGAAAATTATATCGTATCTTTTCCTTGGTGAGATT
CCTTAATGTCAAATCTTAAGAGCCTTTGTGGCTGCTGGATTTGAAGCGTAGTAAGGATATTTTAGAGGTGCTTAGTTAGCCAAT
TTCAGCAGCCAGAGCGGAGTTTCAAACATTTTGCTCATAGGTTCTAGATTCTAGGAAAGAGATCGTCCGTATTTTTGCTAAAGT
TCAATGTGACATCACGAAAGAGGGGTTTTGGGGGTTTTTGGGGTTTTGGGG
```

[Supplementary material is included following the article.]

**Supporting Figure S9. Nucleotide and amino acid sequences of Selenoprotein W3 (eSelW3).** Initiator AUG codon is shown in green; and Sec-encoding UGA and stop UAA codons are in bold and red. An additional UGA codon that corresponds to Cys is shown in red. The sequence corresponding to the SECIS element is shown in blue and underlined. In the protein sequence, initiator (M) and Sec (U) residues are highlighted. Cysteine (C) encoded by UGU is highlighted in blue.

```
 CCCCAAAACCCCAAAACCCCAAAACCCCTTCTTTCTTGGAAAAGTTTGAAATTTACCAAAAATAAAATATATATGTGATAAAG
ATAGGGATAATGGGAGTTTGTGGCTCGAAGTCAGAAATCTTCGGCTCCAAAGATGCTGAATTCGCTATGAAAATTCAGTTCTGT
                                                                    M  K  I  Q  F  C

GGAGGCTGATCATACCGTCCCAAAGCTGTCTATGTCCAGAAGGAGGTAGAAAAGATCTTTGGTGAAAAGTTAGCTGTTATTTTT
 G  G  U  S  Y  R  P  K  A  V  Y  V  Q  K  E  V  E  K  I  F  G  E  K  L  A  V  I  F

AAGAAAGACTTGAAGGTAACTGGGAATTTCGAAATAATCCTCTTTAATCAGAAGACGGGTGAGTCGAAGTTGGTTCATAGTAAG
 K  K  D  L  K  V  T  G  N  F  E  I  I  L  F  N  Q  K  T  G  E  S  K  L  V  H  S  K

AAGAATGGTGGTGGCTTCGTGAAGGAGGATAATTTTGATGAGTTTAAAGAAAAACTCGCCGAATTCTGATCATCTTAATCATCA
 K  N  G  G  G  F  V  K  E  D  N  F  D  E  F  K  E  K  P  T  E  F  C  S  S

ACATAGCCTCTTTCTGTTTCAACAGATATATTGCATATTTTAATGATCACCTGTCTCTTCAAACCCATAAGAGCAGGAGTGAAC
CCAATATGAATATTTCAGTTGGATTTATACGCTTCTAAGCGTAGGAAAGGGCTTTTAATCGATAATTTTCTACCTAAAATTATT
AGAAATCTTTGTCTTCTCTGATAAGAAATACTATACGATCGTAACAATACATTCAATATGGTGGTTCCCAAAAGGGGGTTTTGG
GGGTTTTTGGGGGTTTTTGGGGG
```

[Supplementary material is included following the article.]

**Supporting Figure S10. Nucleotide and amino acid sequences of ep22.** Initiator AUG codon is shown in green; and Sec-encoding UGA and stop UAG codons are in bold and red. The SECIS element is shown in blue and underlined. In the protein sequence, initiator (M) and Sec (U) residues are highlighted. Cysteine (C) encoded by UGA codons are highlighted in red and by UGC in green.

```
  CCCCAAAACCCCCAAAAACCCCCAAAAACCCCTACCTTAGACTAATAATATTGACATCAACCAAAATTATAAAATAATGGAA
                                                                          M  E
TCAAGTGACGATAAGGTTGGCTGCGTGCAGTCTATTCTAGTAGTCCTAGAAGGCCTGAATGATGATTCTTCCATCATCACAGGT
 S  S  D  D  K  V  G  C  V  Q  S  I  L  V  V  L  E  G  L  N  D  D  S  S  I  I  T  G

CTAGAAATCTTAATAAAGCTGATTAAGAACATACTAAAGTCTCCTCATGAAGAGAAATTTAGAAACATTAAGAAGACTAACAAG
 L  E  I  L  I  K  L  I  K  N  I  L  K  S  P  H  E  E  K  F  R  N  I  K  K  T  N  K

GCTATTTCCACAAAGCTGTTGTCCCTCAGTGGAATCGAAGATTTGATCCTCGCCCTTGGTTACAAAGATGATAATGATGAGTTC
 A  I  S  T  K  L  L  S  L  S  G  I  E  D  L  I  L  A  L  G  Y  K  D  D  N  D  E  F

TATGTATTCGACATTGACAAGTACTCTGACCTCTACAAACTAAAGAGAGCTATCCAAGAGTTCCACGATGAGAAAAGAAAGAAG
 Y  V  F  D  I  D  K  Y  S  D  L  Y  K  L  K  R  A  I  Q  E  F  H  D  E  K  R  K  K

TACATGACTCCAGAAGAACTTGAGAAATTCGAAATCCTCCAAGAGCAGAAGAGAAAGTTCTACGAAGATAACAAGAAAAAGGCT
 Y  M  T  P  E  E  L  E  K  F  E  I  L  Q  E  Q  K  R  K  F  Y  E  D  N  K  K  K  A

AAAGCTCGTAAAGATCTTGAGAATGGCATGAAGTTCGACCGTGAAGAGAAGAATCAAGAAGAAATCAAGTCTTCTAAGGCTAAT
 K  A  R  K  D  L  E  N  G  M  K  F  D  R  E  E  K  N  Q  E  E  I  K  S  S  K  A  N

CACCTAAACTTTGGAGCTAATGTGGTTAAATTCCAACCACCAGCTCCAGCCTCTCGTTGAGGTTAGATGCCCTGCTCCGATGTC
 H  L  N  F  G  A  N  V  V  K  F  Q  P  P  A  P  A  S  R  U  G

CGGAGTGGGAATACCTTAAATCAAGGGAATCTGAGATTGGTTGAGAGAAATGGAGCCAATAGGCAATTATGATAATAGTTGAGC
AAACCTTTGAACTCATCTATTTGAACCCCATTATCTTTCCCTTCTAAATATTTCCTTATATGTATCTATATTTTGCTTGTAAAA
AGGGGTTTTGGGGTTTTGGGGTTTTGGGG
```

[Supplementary material is included following the article.]

**Supporting Figure S11. Nucleotide and amino acid sequences of glutathione peroxidase 1 (eGPx1).** Initiator AUG codon is shown in green; and Sec-encoding UGA and stop UAG codons are in bold and red. UGA codons that correspond to Cys are shown in red. The sequence corresponding to the SECIS element is underlined. In the protein sequence, initiator (M) and Sec (U) residues are highlighted. Cysteine (C) encoded by UGA codons are highlighted in red, by UGC in green and by UGU in blue.

```
AAAATAAATATTGAGCAATTTTATTATAAATATAAAAAGTAACTTCACCTTTATAAGGAATTATGGGACAAGTCTTCTTCAAA
                                                              M  G  Q  V  F  F  K

TCTAAGAAGGAGAAGCTAGCAACCACGGTGAAGTCACTCTTTGAAATATCCGCCAAGGATATTGACGGGCAGACTCACCTGCTA
 S  K  K  E  K  L  A  T  T  V  K  S  L  F  E  I  S  A  K  D  I  D  G  Q  T  H  L  L

GCGGATCTAGCAGAAGGCCGTAAATGTACTATGGTCGTGAATGTAGCCTCAAAATGAGGATTGACCAAGACTCACTACAAACAG
 A  D  L  A  E  G  R  K  C  T  M  V  V  N  V  A  S  K  U  G  L  T  K  T  H  Y  K  Q

ATGGTCAAGATTCACAATAAATACAGGGATCATGGATTTGAGATCTTTGCTTTCCCTTGCAACCAGTTCATGAGCCAAGAGCCA
 M  V  K  I  H  N  K  Y  R  D  H  G  F  E  I  F  A  F  P  C  N  Q  F  M  S  Q  E  P

GGAACCCACGAACAGATCAAGAAATTTGCTCAGGAGAAGTATGGTGCTGAATTCCCACTCTTCTCTAAGGTAGACGTCAATGGC
 G  T  H  E  Q  I  K  K  F  A  Q  E  K  Y  G  A  E  F  P  L  F  S  K  V  D  V  N  G

CCTGACACTCATGAAGTGTTCAAGTTCTGCAGAAGACACTCACCATTGTATGATGCTGAGAAGGATGTCGTGCAGAATATCCCT
 P  D  T  H  E  V  F  K  F  C  R  R  H  S  P  L  Y  D  A  E  K  D  V  V  Q  N  I  P

TGGAACTTCGCTAAGTTTTTGATTGATAACAAAGGGACAAGTTGTTGAGTATTACACTCCCAAGCAGAACCCAGATCTCTGCGT
 W  N  F  A  K  F  L  I  D  N  K  G  T  S  C  C  V  L  H  S  Q  A  E  P  R  S  L  R

GCCAAAGATCGAAGAGATGCTTGGATTGTAACTCAGCAGATAAAATTTTAGAAATGGCTCGAGCTCAACTATGCTTATTGCCTT
 A  K  D  R  R  D  A  W  I  V  T  Q  Q  I  K  F

AGGTGATACCTCATTTTGCTATAAAAACTTCAGGCGGTGAGGTTGAAAATAGGTGATGACATCTTTCGAGATCGCCATTTTATG
TTAAAATTTTATTAATATGGAAATCTTACGAAAATGATTTACCCATGGATATTCTTTGGGAAGGTATCCTCACCCATTTTATAT
GAGATTTGGGGGTTTTTGGGGGTTTTTGGGGGTTTTTGGGG
```

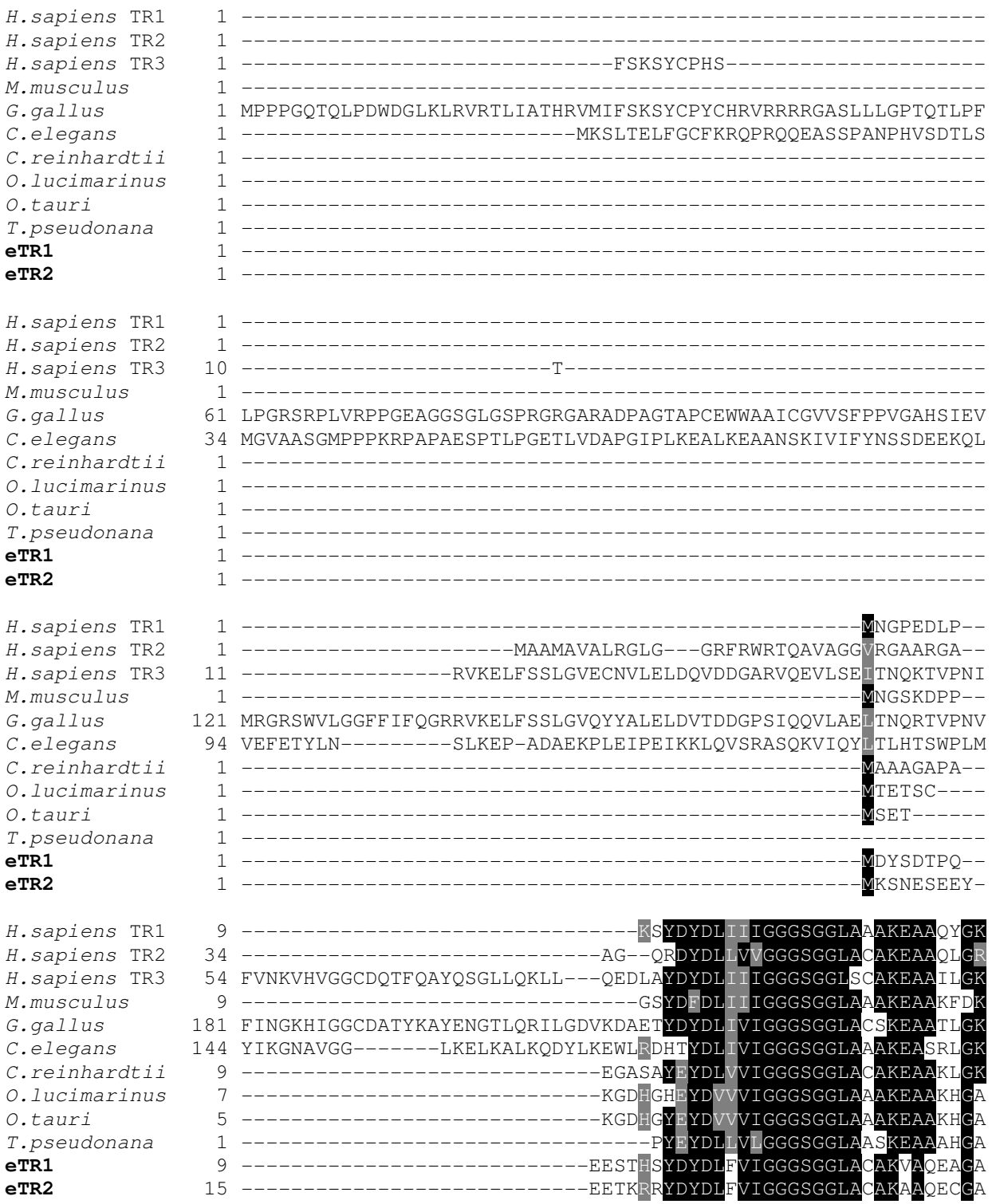
[Supplementary material is included following the article.]

**Supporting Figure S12. Nucleotide and amino acid sequences of glutathione peroxidase 2 (eGPx2).** Initiator AUG codon is shown in green; and Sec-encoding UGA and stop UAA codons are in bold and red. UGA codons that correspond to Cys are shown in red. The sequence corresponding to the SECIS element is underlined. In the protein sequence, initiator (M) and Sec (U) residues are highlighted. Cysteine (C) encoded by TGA codons are highlighted in red, by UGU in blue and by UGC in green.

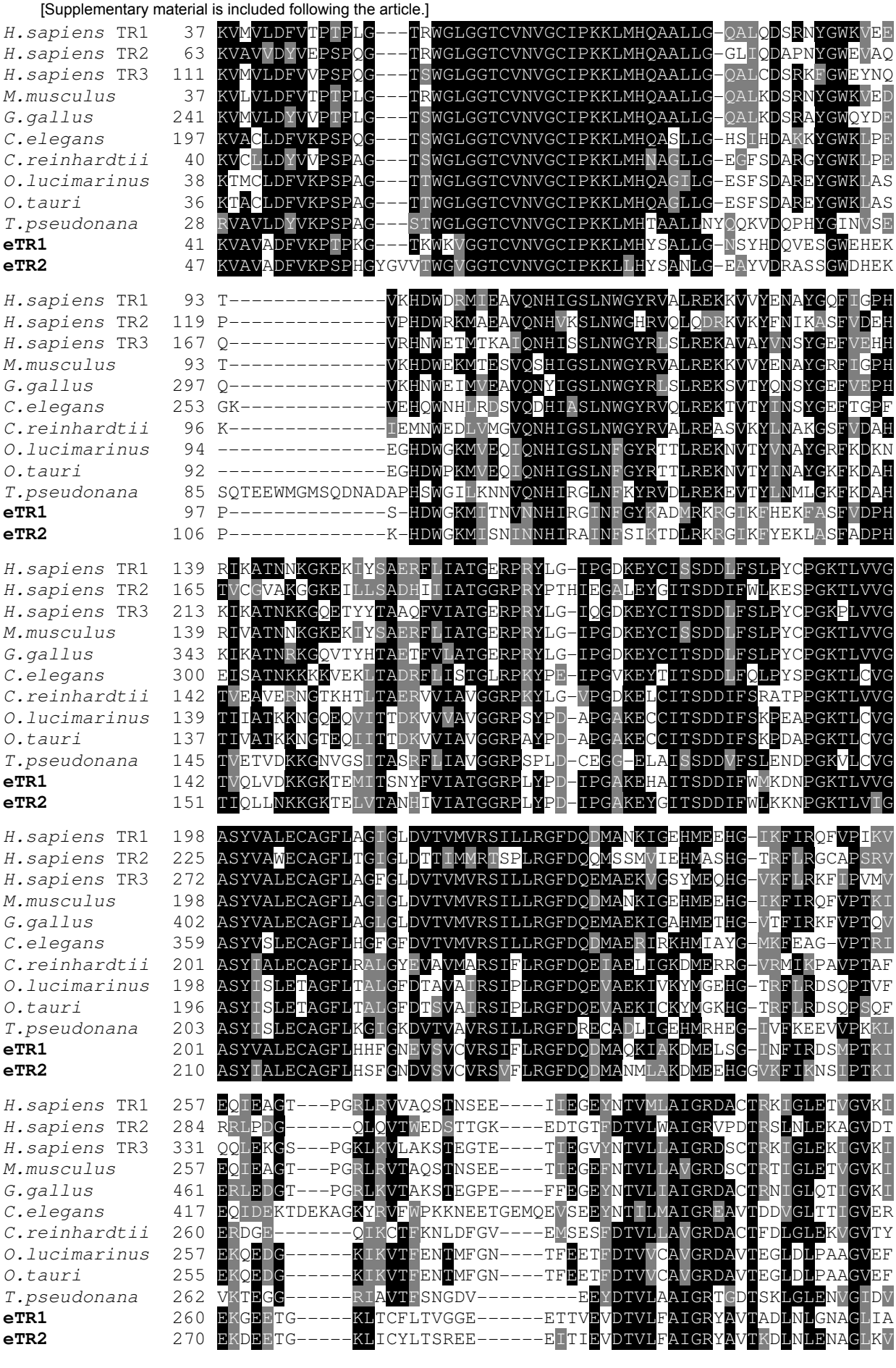
```
 CCCCAAAAACCCCCAAAAACCCCCAAAAACCCCGTGGTATTTATGAAATATTGATCAGTTTGAATATATAATTAATAAAAACT
TCTTGCATAATTATGGGAGCTGCTCTCTGCTTTAAAAAGCGGAAAGAGAAGCTAGAAACCACGGTGGAATCCCTGTTTGAGATA
              M  G  A  A  L  C  F  K  K  R  K  E  K  L  E  T  T  V  E  S  L  F  E  I

TCTGCAGAAGATATTGATGGGCAGGAACACCTTCTAGCGGATCTCGCTAAAGACAAGAAGTGTATAATGGTAGTTAATGTCGCC
 S  A  E  D  I  D  G  Q  E  H  L  L  A  D  L  A  K  D  K  K  C  I  M  V  V  N  V  A

TCGAAATGAGGATTGACCAAGACTCACTATACACAGATGGTCAAGATTCACAACAAATATAAAGACAAGGGATTTGAAATCTTT
 S  K  U  G  L  T  K  T  H  Y  T  Q  M  V  K  I  H  N  K  Y  K  D  K  G  F  E  I  F

GCTTTTCCTTGCAACCAGTTCTTGAGCCAAGAACCTGGATCAAATGAGGACATCAAGAAGTTTGCTAGAGAGAAATATGGAGCT
 A  F  P  C  N  Q  F  L  S  Q  E  P  G  S  N  E  D  I  K  K  F  A  R  E  K  Y  G  A

GAATTCCAATTATTTTCTAAGATCGATGTAAATGGGCCTAACACCCATGAAGTGTTCAGATTTTGTAGGAGACACTCTCCTCTC
 E  F  Q  L  F  S  K  I  D  V  N  G  P  N  T  H  E  V  F  R  F  C  R  R  H  S  P  L

TATGATGATGAGACAGACACTATCCAAAACATCCCATGGAACTTTGCTAAGTTCCTAATTGATGAGGAGGGAAATGTTGTAAAT
 Y  D  D  E  T  D  T  I  Q  N  I  P  W  N  F  A  K  F  L  I  D  E  E  G  N  V  V  N

TATTACTCTCCTAAATCAAATCCAGATGTTTGTGTTCCAATGATAGAGGAAATGCTTGGATTGTAATTCAGCCTGAATCAAGTC
 Y  Y  S  P  K  S  N  P  D  V  C  V  P  M  I  E  E  M  L  G  L

TAAAATGACTAGAGCTTGATAATGCAAATGACCTAAAATGATACCCCATTTAGCTAAACTACTCAAAAGCAATGGGGTTGAAGA
AAGGTAGTGGCATTTAAAAGATCGTCATTTTAATTAAAAATTTCCCC
```

[Supplementary material is included following the article.]

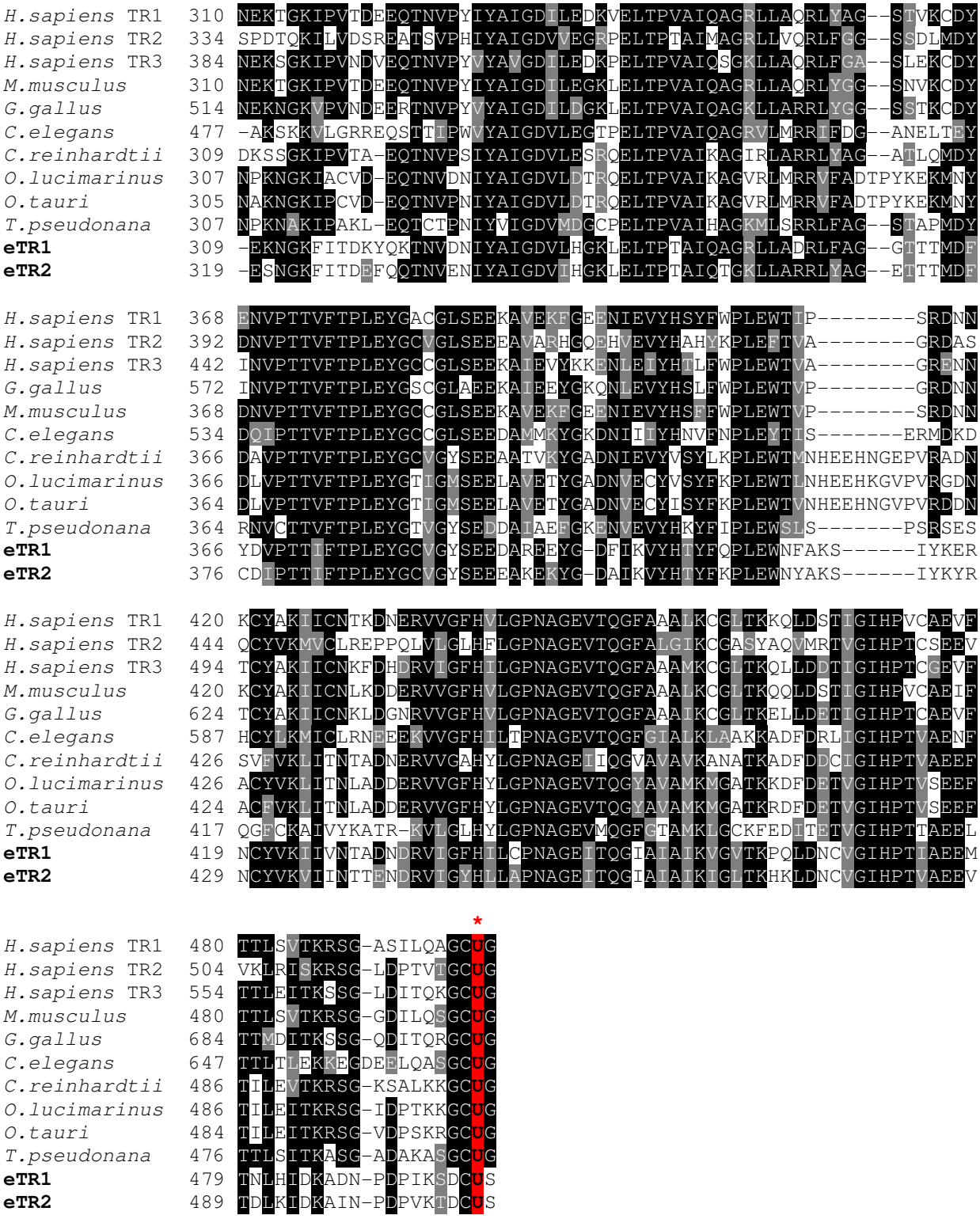
**Supporting Figure S13. Multiple sequence alignment of thioredoxin reductase (TR) sequences.** Sec residues are highlighted in red and indicated with an asterisk. Sequences with the following accession numbers were used to generate the alignment: AAN32903.1 (*C. reinhardtii*), NP_056577.2 (*M. musculus*), XP_414371.1 (*G. gallus*), newV2.0.genewise.318.11.1 (*T. pseudonana*), NP_877419.1 (*H. sapiens*), AAD39929.1 (*H. sapiens*), BAA77601.2 (*H. sapiens*), NP_501085.2 (*C. elegans*), XM_001415547 (*O. lucimarinus*), CR954199.1 (*O. tauri*).



[Supplementary material is included following the article.]

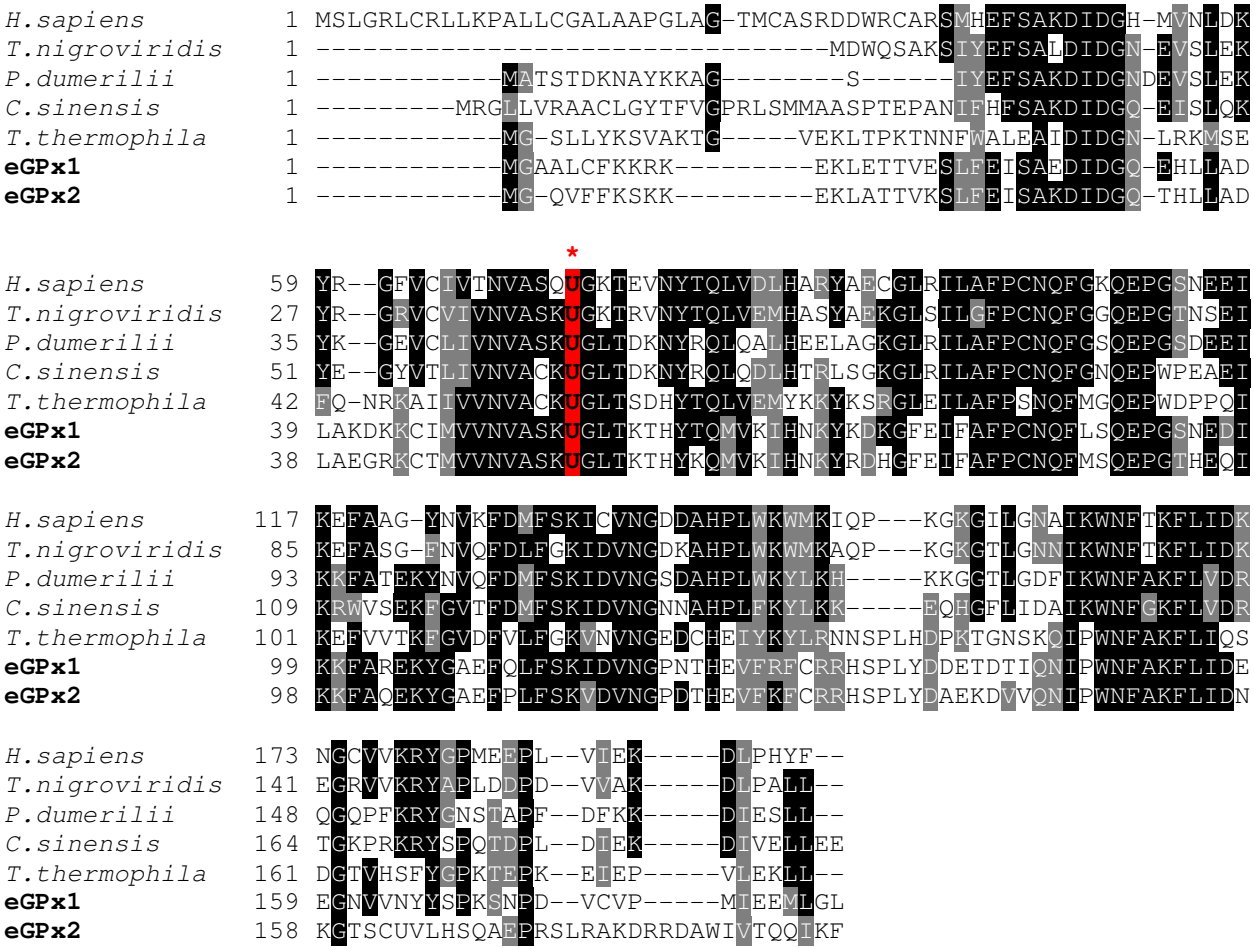


[Supplementary material is included following the article.]



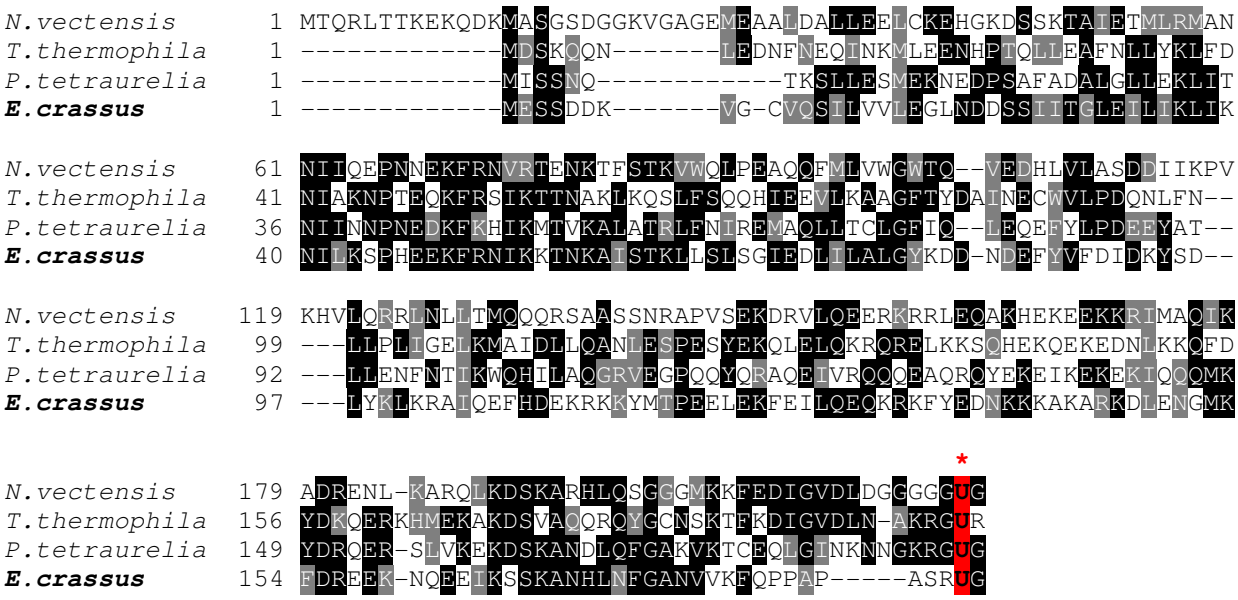
[Supplementary material is included following the article.]

**Supporting Figure S14. Multiple sequence alignment of glutathione peroxidase (GPx) sequences.** Sec is shown in red and indicated with an asterisk. The following sequences were used to generate the alignment: NP_002076.2 (*H. sapiens*), XM_001011053.2 (*T. thermophila*), CR709346.2 (*T. nigroviridis*), CT032433.1 (*P. dumerilii*), and ABK58679.1 (*C. sinensis*).



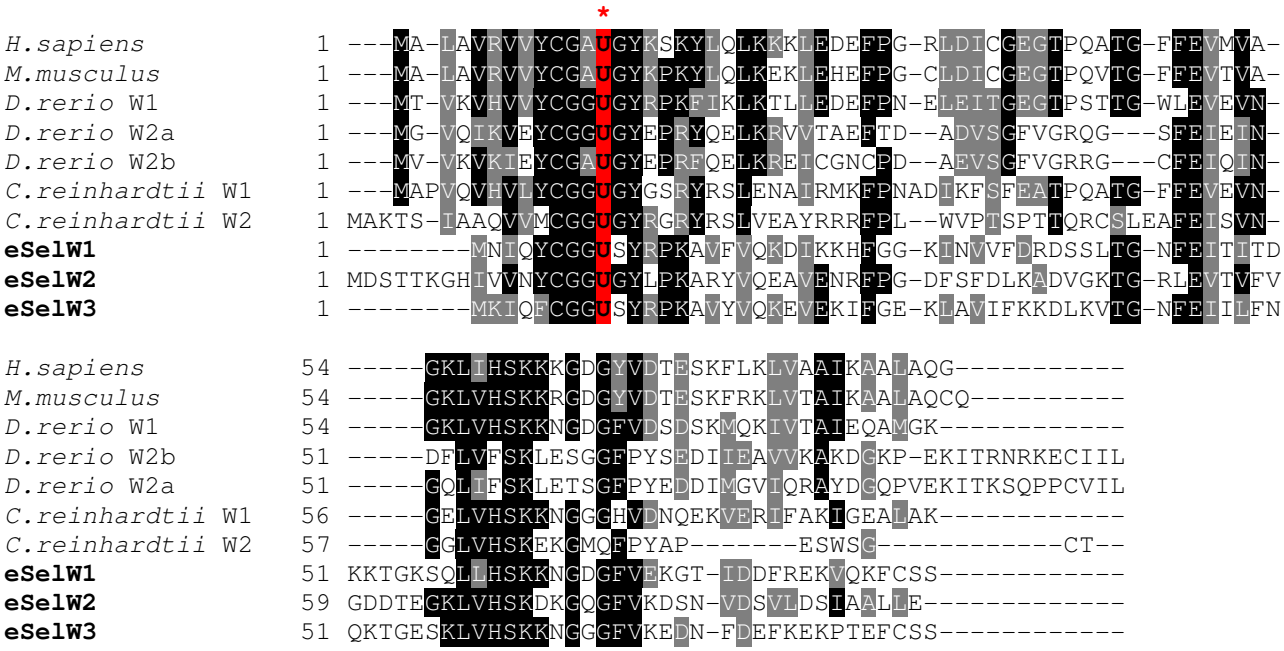
[Supplementary material is included following the article.]

**Supporting Figure S15. Multiple sequence alignment of selenoprotein ep22 sequences.** Sec residues are highlighted in red and indicated with an asterisk. Alignment was created using the following sequences: XM_001455925.1 (*P. tetraurelia*), CT822387.1 (*P. tetraurelia*), XM_001633753.1 (*N. vectensis*), and XM_001023471.2 (*T. thermophila*).

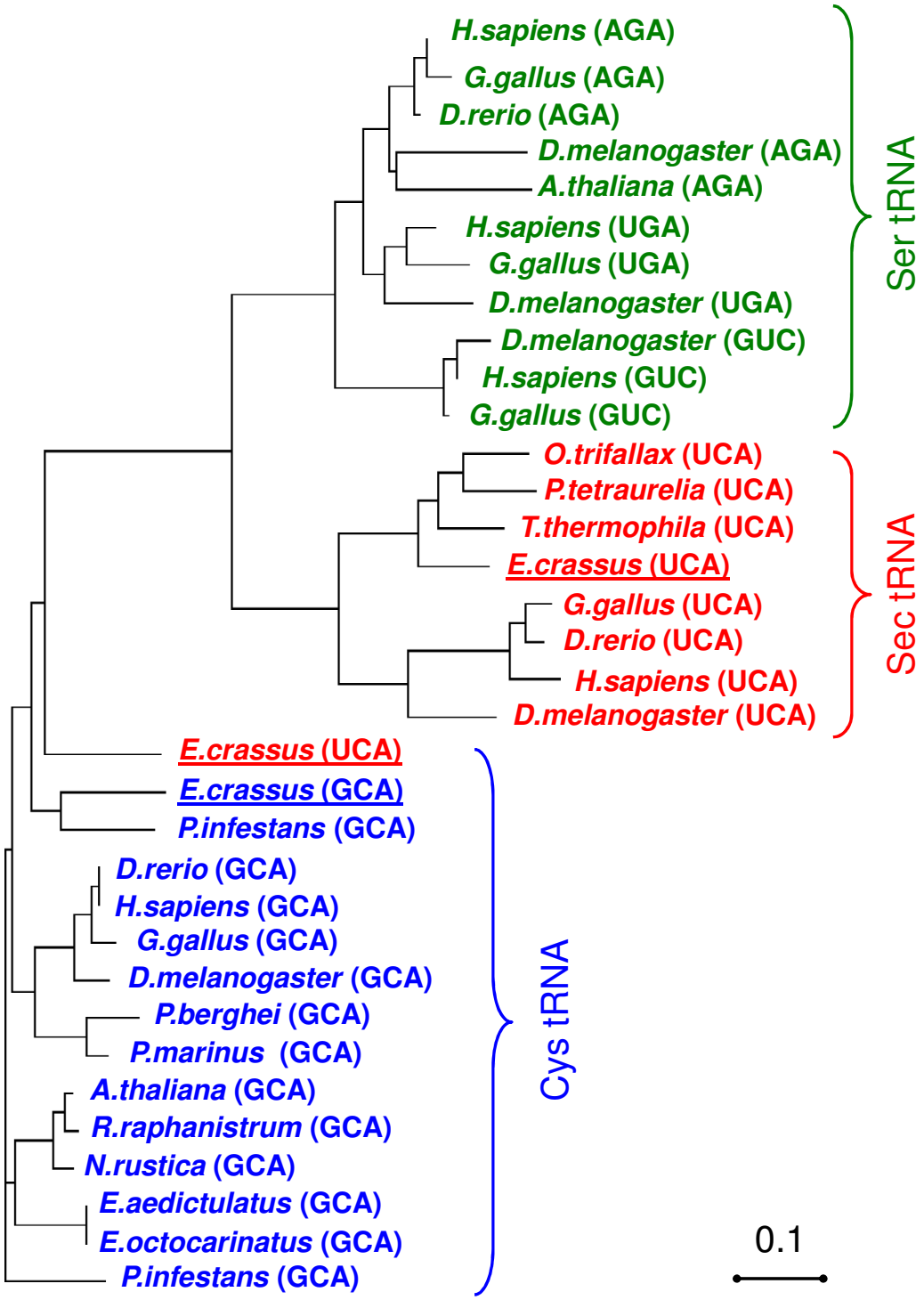


[Supplementary material is included following the article.]

**Supporting Figure S16. Multiple sequence alignment of SelW sequences.** Sec residues are highlighted in red and indicated with an asterisk. Sequences with the following accession numbers were used to generate the alignment: NP_003000.1 (*H. sapiens*), NP_033182.1 (*M. musculus*), AAO86696.1 (*D. rerio*), NP_919398.1 (*D. rerio*), NP_919399.3 (*D. rerio*), BU654801.1 (*C. reinhardtii*), BP092691.1 (*C. reinhardtii*).

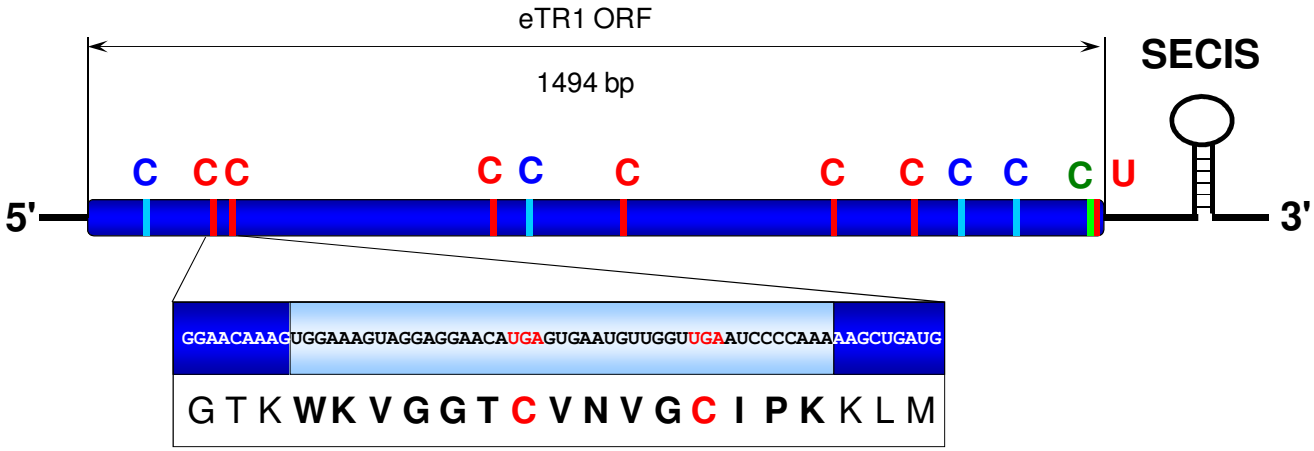


**Supporting Figure S17. Phylogenetic tree of Sec, Cys and Ser tRNA sequences.** Ser tRNAs are shown in green, Cys tRNAs in blue, Sec tRNAs and *Euplotes* Cys tRNA containing UCA anticodon in red. *Euplotes* tRNAs are underlined. Representative tRNA sequences were used to evaluate the origin of the *Euplotes* Cys tRNA containing the UCA anticodon.



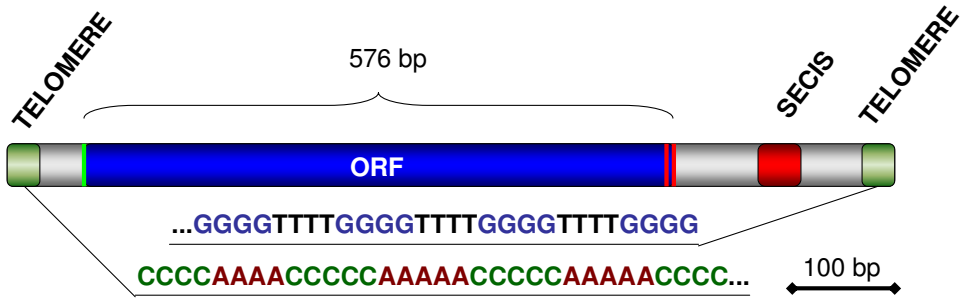
[Supplementary material is included following the article.]

**Supporting Figure S18. Schematic representation of eTR1 mRNA.** The positions of Sec and Cys UGA codons are shown in red, Cys UGU codons in blue and a Cys UGC codon in green. One tryptic peptide containing two UGA-encoded Cys codons and identified by MS/MS is shown below in bold (the corresponding nucleotide sequence is shown on light blue background). Also, see Fig. S20-S22.



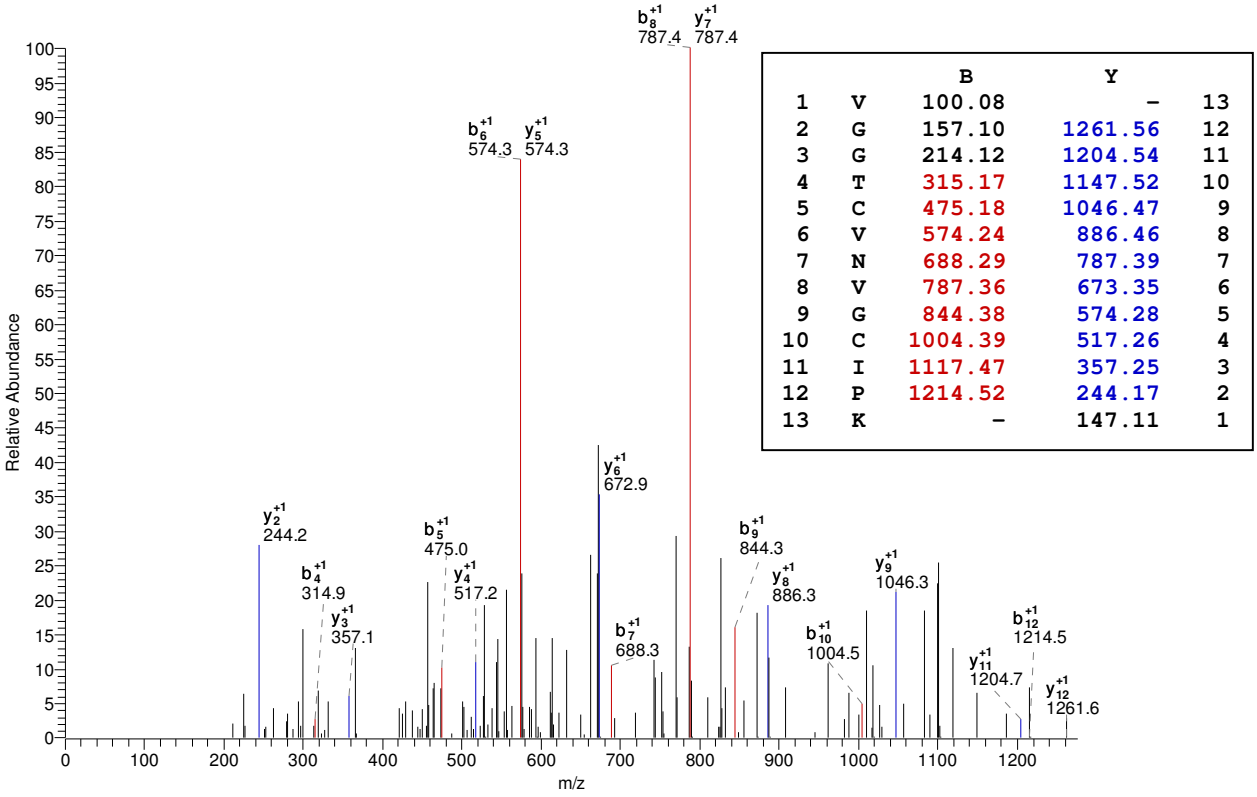
[Supplementary material is included following the article.]

**Supporting Figure S19. Organization of the chromosome containing the ep22 selenoprotein gene.** Telomeres are indicated in green with their sequences shown below the chromosome, SECIS element in red and the ORF in blue. Location of the initiator (Met) codon is shown by a green line, and Sec and stop codons by red lines.



[Supplementary material is included following the article.]

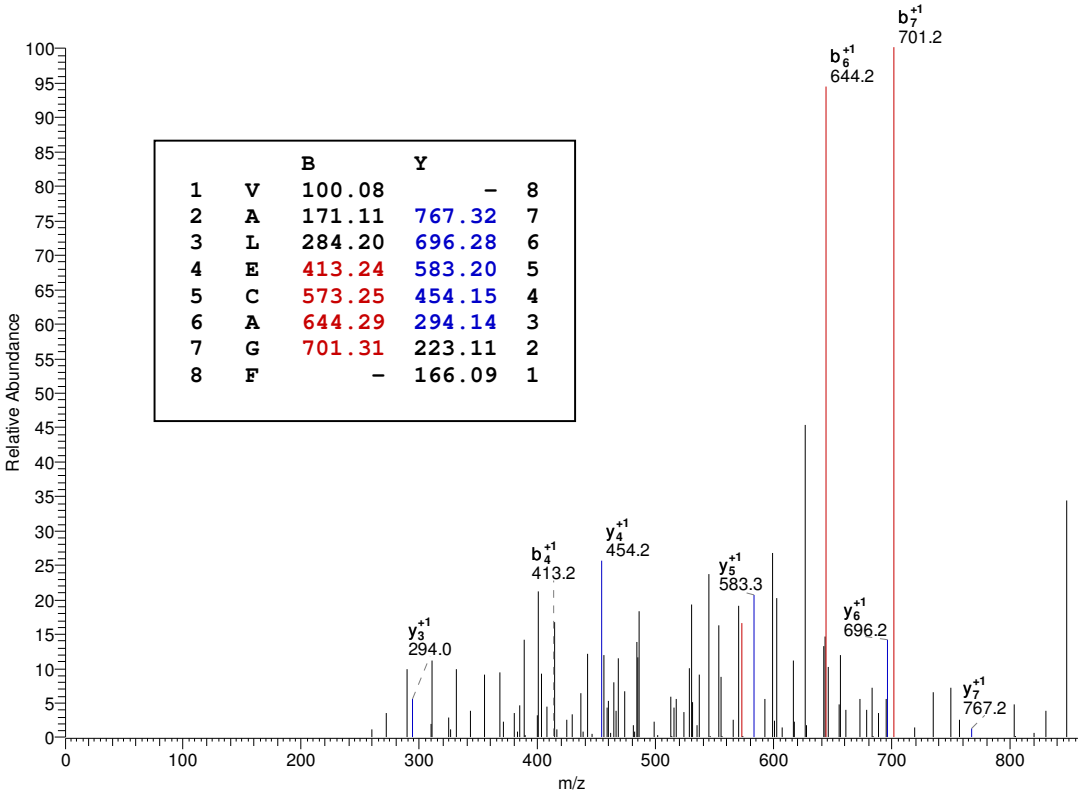
**Supporting Figure S20. LC-MS/MS data and the spectrum of the "VGGTCVNVGCIPK" peptide from eTR1.** The protein was alkylated with iodoacetamide to protect Cys residues. Detected b and y ions are shown in the table.




| | | B | Y | |
|---|---|---|---|---|
| 1 | V | 100.08 | – | 13 |
| 2 | G | 157.10 | 1261.56 | 12 |
| 3 | G | 214.12 | 1204.54 | 11 |
| 4 | T | 315.17 | 1147.52 | 10 |
| 5 | C | 475.18 | 1046.47 | 9 |
| 6 | V | 574.24 | 886.46 | 8 |
| 7 | N | 688.29 | 787.39 | 7 |
| 8 | V | 787.36 | 673.35 | 6 |
| 9 | G | 844.38 | 574.28 | 5 |
| 10 | C | 1004.39 | 517.26 | 4 |
| 11 | I | 1117.47 | 357.25 | 3 |
| 12 | P | 1214.52 | 244.17 | 2 |
| 13 | K | – | 147.11 | 1 |

[Supplementary material is included following the article.]

**Supporting Figure S21. LC-MS/MS data and spectrum of the "VALECAGF" peptide from eTR1.** The protein was alkylated with iodoacetamide to protect Cys residues. Detected b and y ions are shown in the table.



[Supplementary material is included following the article.]

**Supporting Figure S22.Peptide sequences identified in *E. crassus* selenoproteins by LC-MS/MS.**
The peptides detected by mass-spectrometry are shown in red. In the protein sequence, the initiator (M) and Sec (U) amino acid residues are highlighted in green and blue, respectively. Cysteines (C) encoded by UGA and detected by LC-MS/MS are underlined.

eTR1

MDYSDTPQEESTHSYDYDLFVIGGGSGGLACAKVAQEAGAKVAVADFVKPTPKGTKWKVGGTCVNVGCIPKK
LMHYSALLGNSYHDQVESGWEHEKPSHDWGKMITNVNNHIRGINFGYKADMRKRGIKFHEKFASFVDPHTVQ
LVDKKGKTEMITSNYFVIATGGRPLYPDIPGAKEHAITSDDIFWMKDNPGKTLVVGASYVALECAGFLHHFG
NEVSVCVRSIFLRGFDQDMAQKIAKDMELSGINFIRDSMPTKIEKGEETGKLTCFLTVGGEETTVEVDTVLF
AIGRYAVTADLNLGNAGLIAEKNGKFITDKYQKTNVDNIYAIGDVLHGKLELTPTAIQAGRLLADRLFAGGT
TTMDFYDVPTTIFTPLEYGCVGYSEEDAREEYGDFIKVYHTYFQPLEWNFAKSIYKERNCYVKIIVNTADND
RVIGFHILCPNAGEITQGIAIAIKVGVTKPQLDNCVGIHPTIAEEMTNLHIDKADNPDPIKSDCUS

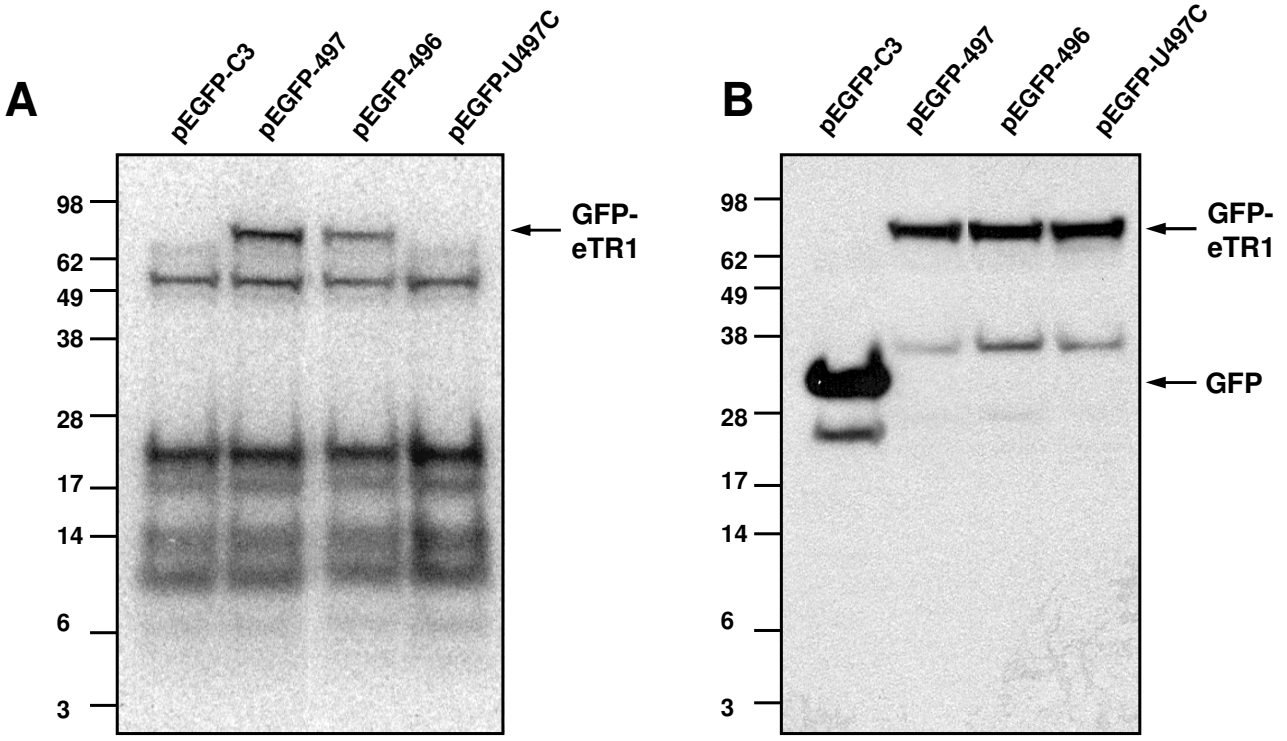
eGR

MKKKIFDYLVIGGGSGGIASANRAAKHGTKVGIIECQALGGTCVNVGCVPKKIVMFNAASFLENRELYEGYG
LSAASELKLDFPTLKKNRDAYIEWLNGIYGGMIANNSMTLIKGWAKFVDNHTVEVDGKDRYTADHILIAVGS
RPDKGGFEGDEYCIDSDGFFALEDLPKKVIVYGGGYIGTELGQILHALGTKVIQVVRSEILRIADDDIREQL
YKLMDLSGYEYRKKTTIVKVEKTDSGLRVHLSDGTIEEDVDECIVATGRPANVEGLGLENTDIELTKKGHIK
VDEFNVTTVPNVYAVGDVTGAAELTPVAIKTGRTLVERLFNNRPDLRMDHDLIPTVIFSHPPIGMIG

eSelW2

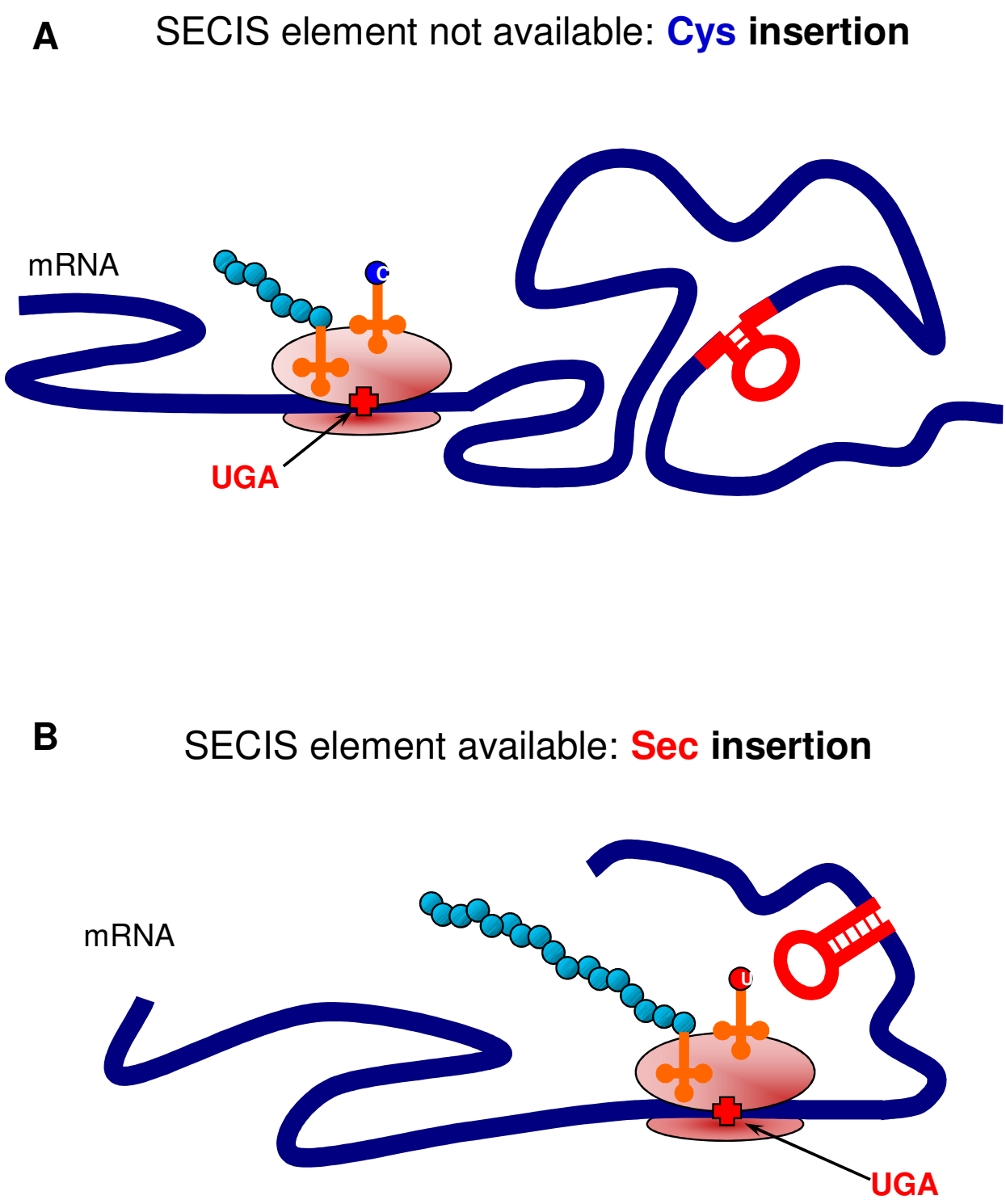
MDSTTKGHIVVNYCGGUGYLPKARYVQEAVENRFPGDFSFDLKADVGKTGRLEVTVFVGDDTEGKLVHSKDK
GQGFVKDSNVDSVLDSIAALLE

**Supporting Figure S23. Sec and Cys insertion in eTR1. (A)** Expression of GFP-eTR1 in HEK 293 cells. Cells were transfected with pEGFP-C3 (control), a pEGFP-eTR1 construct containing a single UGA codon at the natural Sec position 497 (pEGFP-497) or constructs that had a single UGA at unnatural position 496 (pEGFP-496) or no UGA at all wherein the UGA Sec codon was replaced with UGC at position 497 (pEGFP-U497C). **(B)** Western blotting of samples shown in **A** with anti-GFP antibodies. *Arrows* show the positions of GFP and full-size GFP-eTR1. Molecular masses of protein standards (in kDa) are shown on the *left*.



[Supplementary material is included following the article.]

**Supporting Figure S24. Model of Cys or Sec insertion at defined positions in eTR1.** SECIS element (shown in red) is unavailable for Sec insertion upstream of position 470 leading to Cys insertion. Translating ribosome changes mRNA structure such that SECIS can now support Sec insertion.



**Supporting references**


S1. G. V. Kryukov *et al.*, *Science* **300**, 1439 (2003)
S2. A. V. Lobanov *et al.*, *Genome Biol* **8**, R198 (2007)
S3. M. Margulies *et al.*, *Nature* 437, 376 (2005)
S4. C. L. Jahn, S. Z. Doktor, J. S. Frels, J. W. Jaraczewski, M. F. Krikau, *Gene* 133, 71 (1993)
S5. M. Grimm, C. Brunen-Nieweler, V. Junker, K. Heckmann, H. Beier, *Nucleic Acids Res* **26**, 4557 (1998)
S6. Q. A. Sun, V. N. Gladyshev, *Methods Enzymol* **347**, 451 (2002)