

4-30-2007

## Supporting Information for: A highly efficient form of the selenocysteine insertion sequence element in protozoan parasites and its use in mammalian cells

Sergey V. Novoselov  
*University of Nebraska - Lincoln*

Alexey V. Lobanov  
*University of Nebraska - Lincoln*

Deame Hua  
*University of Nebraska - Lincoln*

Marina V. Kasaikina  
*University of Nebraska - Lincoln*

Dolph L. Hatfield  
*National Institutes of Health, Bethesda, MD*

*See next page for additional authors*

Follow this and additional works at: <http://digitalcommons.unl.edu/biochemgladyshev>

 Part of the [Biochemistry, Biophysics, and Structural Biology Commons](#)

---

Novoselov, Sergey V.; Lobanov, Alexey V.; Hua, Deame; Kasaikina, Marina V.; Hatfield, Dolph L.; and Gladyshev, Vadim, "Supporting Information for: A highly efficient form of the selenocysteine insertion sequence element in protozoan parasites and its use in mammalian cells" (2007). *Vadim Gladyshev Publications*. Paper 27.  
<http://digitalcommons.unl.edu/biochemgladyshev/27>

This Article is brought to you for free and open access by the Biochemistry, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Vadim Gladyshev Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

---

**Authors**

Sergey V. Novoselov, Alexey V. Lobanov, Deame Hua, Marina V. Kasaikina, Dolph L. Hatfield, and Vadim Gladyshev

## Supporting Information for:

# A highly efficient form of the selenocysteine insertion sequence element in protozoan parasites and its use in mammalian cells

Sergey V. Novoselov<sup>\*†</sup>, Alexey V. Lobanov<sup>\*</sup>, Deame Hua<sup>\*</sup>, Marina V. Kasaikina<sup>\*</sup>,  
Dolph L. Hatfield<sup>‡</sup>, and Vadim N. Gladyshev<sup>\*§</sup>

<sup>\*</sup>Department of Biochemistry, University of Nebraska, Lincoln, NE 68588;

<sup>‡</sup>Section on the Molecular Biology of Selenium, Laboratory of Cancer Prevention,  
National Cancer Institute, National Institutes of Health, Bethesda, MD 20892

PNAS article edited by W. Ford Doolittle, Dalhousie University, Halifax, NS, Canada

**Abstract:** Selenoproteins are an elite group of proteins containing a rare amino acid, selenocysteine (Sec), encoded by the codon, UGA. In eukaryotes, incorporation of Sec requires a Sec insertion sequence (SECIS) element, a stem-loop structure located in the 3'-untranslated regions of selenoprotein mRNAs. Here we report identification of a noncanonical form of SECIS element in *Toxoplasma gondii* and *Neospora canine*, single-celled apicomplexan parasites of humans and domestic animals. This SECIS has a GGA sequence in the SBP2-binding site in place of AUGA previously considered invariant. Using a combination of computational and molecular techniques, we show that *Toxoplasma* and *Neospora* possess both canonical and noncanonical SECIS elements. The GGA-type SECIS element supported Sec insertion in mammalian HEK 293 and NIH 3T3 cells and did so more efficiently than the natural mammalian SECIS elements tested. In addition, mammalian type I and type II SECIS elements mutated into the GGA forms were functional but manifested decreased Sec insertion efficiency. We carried out computational searches for both AUGA and GGA forms of SECIS elements in *Toxoplasma* and detected five selenoprotein genes, including one coding for a previously undescribed selenoprotein, designated SelQ, and two containing the GGA form of the SECIS element. In contrast, the GGA-type SECIS elements were not detected in mammals and nematodes. As a practical outcome of the study, we developed pSelExpress1, a vector for convenient expression of selenoproteins in mammalian cells. It contains an SBP2 gene and the most efficient tested SECIS element: an AUGA mutant of the GGA-type *Toxoplasma* SelT structure.

**Keywords:** genome, RNA structure, selenocysteine, selenoprotein

Author contributions: S.V.N. and V.N.G. designed research; S.V.N., A.V.L., D.H., and M.V.K. performed research; S.V.N., A.V.L., D.L.H., and V.N.G. analyzed data; and S.V.N., D.L.H., and V.N.G. wrote the paper.

<sup>§</sup> Corresponding author. E-mail: [vgladyshev1@unl.edu](mailto:vgladyshev1@unl.edu)

<sup>†</sup> Present address: Institute of Cell Biophysics, Russian Academy of Sciences, Pushchino, Moscow Region 142292, Russia.

Main article published in *Proceedings of the National Academy of Sciences of the United States* 104:19 (May 8, 2007), pp. 7857-7862. Published online April 30, 2007. doi 10.1073/pnas.0610683104

---

## SI Materials and Methods

**Materials.** Chemicals used in the study were purchased from Sigma, restriction enzymes from Amersham Pharmacia, DNA purification kits from Qiagen, mammalian cell culture reagents and the HEK 293 cell line from Invitrogen, and NIH 3T3 cells from ATCC. Sequences of primers used in the study are available upon request.

**Databases and programs.** *Toxoplasma gondii*, *C. elegans*, human and mouse genome sequences and nonredundant protein sequences (<ftp://ftp.ncbi.nih.gov/genbank>) were obtained from NCBI. SECISearch

was used for identification of candidate SECIS elements (1). BLAST and FASTA programs were used for similarity searches (2).

*Identification of homologs of known selenoprotein genes.* A full set of known eukaryotic selenoproteins was used as query sequences and included all human selenoproteins (1), all *Plasmodium falciparum* selenoproteins (3), *Chlamydomonas* MsrA (4), *Gallus gallus* SelU (5), protein disulfide isomerase from *Emiliania huxleyi* (6) and *Danio rerio* Fep15 (7). A stand-alone version of TBLASTN and FASTA package were used for detection of nucleotide sequences corresponding to known selenoprotein families.

*A search for canonical Toxoplasma SECIS elements.* A stand-alone version of SECISearch with the default pattern was used. The search procedure included the following steps:

1. Analysis of primary nucleotide sequence and secondary structures. We used PatScan to search the target database for the candidates satisfying the NUGA\_\_AA\_\_GA pattern. This pattern represents almost all eukaryotic SECIS elements (8). The additional requirements were as follows: (i) distance between the quartet (NUGA) and the unpaired AA in the apical loop 10-13 nucleotides, (ii) length of the apical loop without the unpaired AA sequence 6-23 nucleotides, (iii) no more than one insertion, one deletion, and two mismatches in the stem preceding the unpaired AA, and (iv) presence of an additional stem upstream of the quartet. For each SECIS candidate found in the previous step, secondary structure was predicted and examined for consistency with the eukaryotic SECIS consensus model. Additional filters then excluded SECIS elements with more than two consecutive unpaired nucleotides and Y-shaped SECIS elements.

2. Estimation of the free energy. RNAfold from Vienna RNA package was used to calculate the free energies for whole structures and separately for their upper stem-loops. The threshold value was -12.6 kcal/mol for the whole structure and -3.7 kcal/mol for the upper stem-loop.

3. Protein identification. Analysis of location of SECIS elements and identification of ORFs were carried out. Candidate structures located on the complementary strand were filtered out.

4. ORF analysis. This final step consisted of sequence analyses of predicted ORFs and identification of candidate Sec-encoding UGA codons.

*A search for Toxoplasma noncanonical SECIS elements.* The search for noncanonical SECIS elements was carried out as described above for canonical SECIS elements, except that NTGA was replaced by NGGA in the primary sequence. Although no non-canonical SECIS elements other than the GGGA-type structures were discovered by homology searches involving known selenoproteins, the search settings were relaxed to allow any nucleotide preceding GGA (or UGA) for better sensitivity.

*Analysis of mammalian and nematode genomes, and EST sequences.* Analysis of human and mouse genomes was carried out as described in ref. 1, with search patterns modified to meet the modified SECIS consensus model (e.g., GGGA-and AUGA-type SECIS elements). Likewise, similar modifications were made in the nematode search procedure (10). In addition to completely sequenced genomes, the NCBI EST database was searched for the presence of NGGA-type SECIS elements.

*Cloning strategies.* GFP-fusion constructs developed and used in the study are shown in the scheme in Fig. 3A. Selenoprotein H (SelH) cDNA containing the in-frame TGA codon but lacking the entire 3'-UTR was amplified and cloned into pEFGP-C3 (Clontech), and all subsequent constructs containing *Toxoplasma* SECIS elements were developed using this GFP-SelHD3'UTR fusion construct (construct 2 in Fig. 3A). *Toxoplasma* SelT and SelS SECIS elements (130-bp region, constructs 3 and 5, respectively) or the sequences beginning with the corresponding stop codons and containing SECIS elements (>300-bp

region, constructs 4 and 6) were amplified and cloned immediately downstream of the SelH stop codon. The rationale was as follows: the SelH SECIS is located very close to the stop codon (construct 1). Therefore, the constructs having the 130-bp sequences of *Toxoplasma* SECIS elements were regarded as corresponding to substitution of the mammalian SECIS element with the *Toxoplasma* structures, whereas the constructs containing the 300-bp sequences of *Toxoplasma* SelT 3'-UTR or 350-bp sequence of *Toxoplasma* Sels 3'UTRs were substitutions that introduced the corresponding 3'-UTRs.

We further mutated the GG bases in the SECIS quartet in both *Toxoplasma* SelT and Sels to AT (i.e., *Toxoplasma* GG to AT mutants). Likewise, the AT bases in GFP-mSelHwt, GFP-mSelSw, and GFP-mSelMwt (Fig. 4A) fusions were mutated to the GG (i.e., mouse AT to GG mutants) using QuickChange mutagenesis kit (Stratagene).

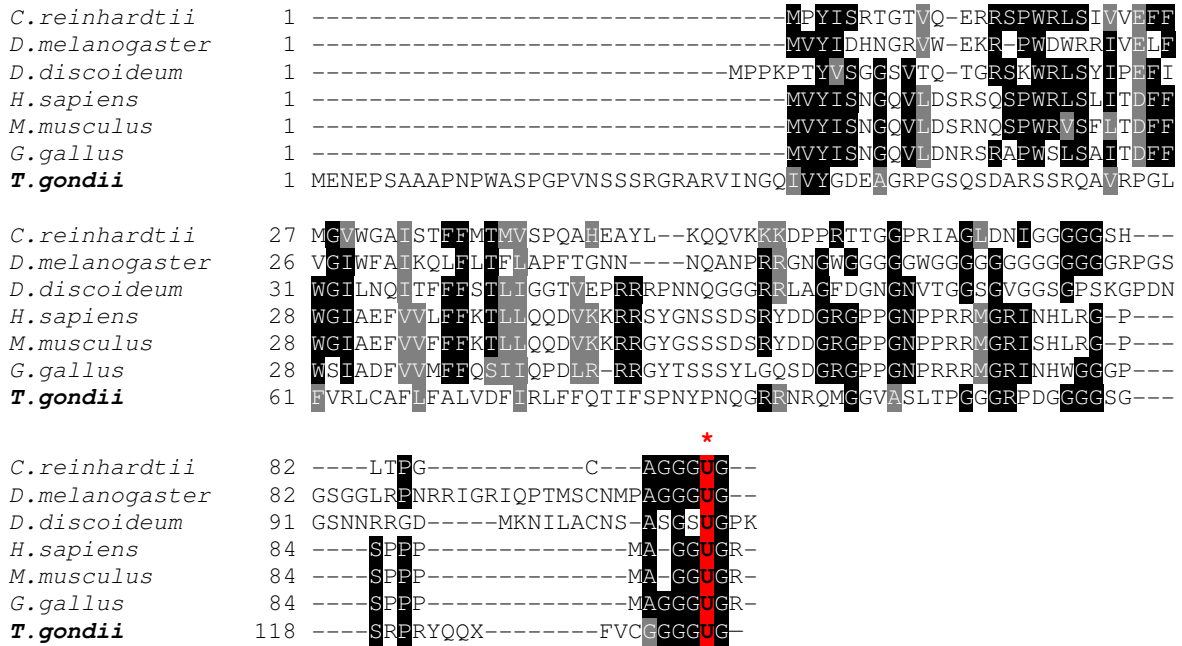
The vector for expression of selenoproteins in mammalian cells was developed on the basis of pBudCE4.1 (Invitrogen). First, the C-terminal domain of rat SBP2 was cloned into the first cloning site for expression under the EF 1apromoter. Subsequently, the *Toxoplasma* SelT AT>GG SECIS was cloned into the second multiple cloning site. Finally, mouse GPx1 sequence containing an in-frame TGA codon, but lacking a 3'-UTR, was amplified and cloned into the vector. As a control, we used the construct wherein mGPx1-*Toxoplasma* SelT SECIS AT>GG mutant was cloned into pBudCE4.1 that did not have the rat SBP2 sequence.

*Cell culture, transfection, and metabolic labeling.* Mouse NIH 3T3 and human HEK 293 cells were cultured in DMEM supplemented with 10% FBS, 100 units/ml penicillin, and 100 units/ml streptomycin. Cells were seeded in six-well plates and transfected as follows: NIH 3T3 cells using Lipofectamin and Plus reagent (Invitrogen) according to the manufacturer's protocol, and HEK 293 using the calcium-phosphate method (9) in OPTI-MEM (Invitrogen), or cotransfected in a ratio of 2:1 with the rat SBP2 expression construct (20) that was the generous gift of Drs. Paul Copeland and Donna Driscoll, Cleveland Clinic Foundation. In 12-24 h after transfection, the medium was replaced with DMEM supplemented with Se (specific activity 1,000 Ci/mmol), and the cells were incubated for an additional 12-24 h.

*SDS/PAGE and Western blot analysis.* After transfection, cells were washed with PBS, harvested, lysed in 200  $\mu$ l of lysis buffer (Sigma), electrophoresed using NuPAGE system (Invitrogen), and transferred onto PVDF membranes. The membranes were exposed to a PhosphorImager screen and metabolically labeled proteins were visualized using a PhosphorImager system (GE Healthcare). The membranes were then probed with anti-GFP rabbit antiserum (Invitrogen) as primary and anti-rabbit HRP-conjugated antibodies as secondary antibodies. The Western blot signals were then detected with an ECL system (GE Healthcare).

*Enrichment of recombinant His-tagged GPx1 protein on metal-affinity resin.* Forty-eight hours after transfection of mammalian cells with various His tag-GPx1 expression constructs, the cells were harvested, lysed in PBS containing protease inhibitors (complete protease mixture, Roche) by brief sonication and centrifuged for 5 min. Supernatants were collected, normalized with respect to protein concentration using Bradford method (Bio-Rad), and mixed with TALON affinity resin (Clontech). Total protein (0.75 mg; 1 mg/ml, 750  $\mu$ l) per 40-50  $\mu$ l of the resin was used. The samples were incubated under delicate rotation for 1 h at 4°C. After incubation, the resins were washed extensively, and the bound proteins were eluted by heating in an SDS/PAGE loading buffer and analyzed by gel electrophoresis and immunoblotting. After analysis of Se-labeled proteins as described above, the membranes were probed in Western blots with anti-GPx1 antibodies (GeneTex) according to the manufacturer's protocol.

## Supporting Figure 1



**Fig. 1. Multiple sequence alignments of apicomplexan selenoprotein SelK.** Sequences with the following accession numbers were used in the alignment: TgEST\_95058496 (*T. gondii*), AAH13162.2 (*H. sapiens*), Q9JLJ1 (*M. musculus*), NP\_001020612.1 (*G. gallus*), AAN32902.1 (*C. reinhardtii*), XP\_646897.1 (*D. discoideum*), and NP\_572763.3 (*D. melanogaster*). Selenocysteine residues (U) are highlighted in red and indicated by asterisks.

## Supporting Figure 2

```

H. sapiens          1 MA---LAVRVVYCGAUGYKSKYLQKKKLEDEF-----GRLDICGEGTPQATGFFEVMVA
M. musculus       1 MA---LAVRVVYCGAUGYKPKYLQKKELEHEFP-----GCLDTCGEGTPQVTGFFEVTVA
D. rerio          1 MT---VKVHVYVYCGGUGYRPFKIKLKTLLLEDEF-----NELEITGEGTPSITGMLEVEVN
C. reinhardtii W1 1 MAP---VQVHVLYCGGUGYGSRYRSLNENAI RMKFPN---ADIKFSFEATPQATGFFEVEVN
C. reinhardtii W2 1 MAKTSIAAQVVMCGGUGYRGRYSLVEAYRRRFP-----LWVPTSPTTQRC SLEAFEISVN
T. gondii        1 MEQT-VEITIQFCGGUGYRPFYFDRAEALIRSWLSDAELRRVSTEGHEDEFGITGNFEIRIN
N. caninum       1 MART-VEITIQFCGGUGYRPFYFDRAEALIRSWFTDVYFRHVSTEGHEDEFGITGNFEIRID

H. sapiens          54 GKL I - -KGDGY TESKFLKLV-AAIKAALAQG-
M. musculus       54 GKL VHSKK-RGDG V DTESKFRKLV-TAIKAALAQ CQ
D. rerio          54 GKL VHSKK-NGDGFVDSDSKMQKIV-TAIEQAMGK--
C. reinhardtii W1 56 GELVHSKK-NGGGHVDNQE KVERI IGEALAK--
C. reinhardtii W2 57 GGLVHSKE-KGMQFPYAPESWSGCT-----
T. gondii        60 GKL VHSKKT KKQGF L HANKEQQEV RQKLKEALGN--
N. caninum       60 GVLVHSKKT RRQGF L HANKEQQEV V RQKIREALDN-

```

**Fig. 2. Multiple sequence alignments of apicomplexan selenoprotein SelW.** The alignment is based on the following sequences: NP\_003000.1 (*H. sapiens*), NP\_033182.1 (*M. musculus*), AAO86696.1 (*D. rerio*), BU654801.1 and BP092691.1 (*C. reinhardtii*), TgEST\_95057361 (*T. gondii*), and TC2958 (*N. caninum*). Selenocysteine residues (U) are highlighted in red and indicated by asterisks.

### Supporting Figure 3

```

T. gondii      1 MEEALREM HSRLPKADQIQALNLLIKIVNNV GSANPEELERFRCINSGSTALQQR
N. caninum    1 MEEALQEV QSRLPKADQIQALNLLIKIVTNI PAATPEEVERFRCINSGSTALQQR

T. gondii     61 LLRHGPVYENLLLALGFYRTTEPPVSRPLPQPNQYFFLPEHADRAQLLADLELLRATVA
N. caninum    61 LLRHGPVYENLLLALGFYRTADPPLSCPLTQANQYFFLPDHADGGRLLADLELLRATVA

T. gondii    121 SLETEGD---DRMPAAERLTSG TGAPRKVTTTSRAIRDSS AAHARNQEELRQLREEQ
N. caninum   121 SLEAEGGNAIESSEPTAERLNSA QGAQRKVTTTSRAIRDSS SMHARNQEELRRLREEQ

T. gondii    178 RARFEQRSE QATGGITGWLSASLAPSAS SAAQPAQPRHPEPADVPTPGGSRREGSGGN
N. caninum   181 RLRFQRSE EPAGGIAGWFSSSLAP PSAQPAG---P-----

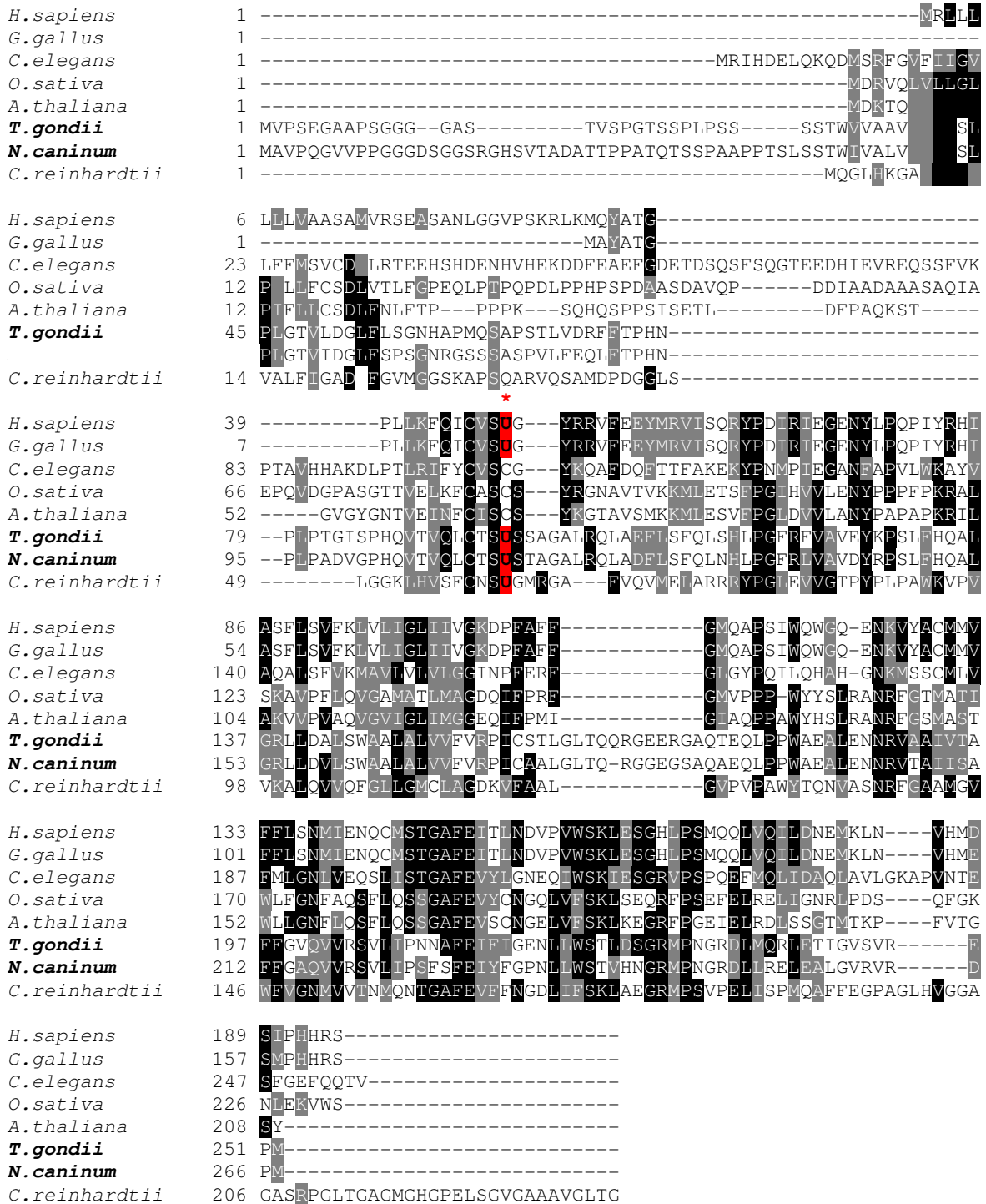
T. gondii    238 AASRFFKSLFGGRSGSRSEEGH GAANRRDRDSRGPRMKTIKDLPPAPQRRGUG
N. caninum   219 --S-----LFGSRSGSRSEEGR DCTSQRGGDSRGPRMKTIKDLPPAPRRRGUG

```

**Fig. 3. Multiple sequence alignments of apicomplexan selenoprotein Sels-like.** The following sequences were used in the alignment: TgTwinScan\_4798 (*T. gondii*), and TC3699 and TC3703 (*N. caninum*). Selenocysteine residues (U) are highlighted in red and indicated by asterisks.

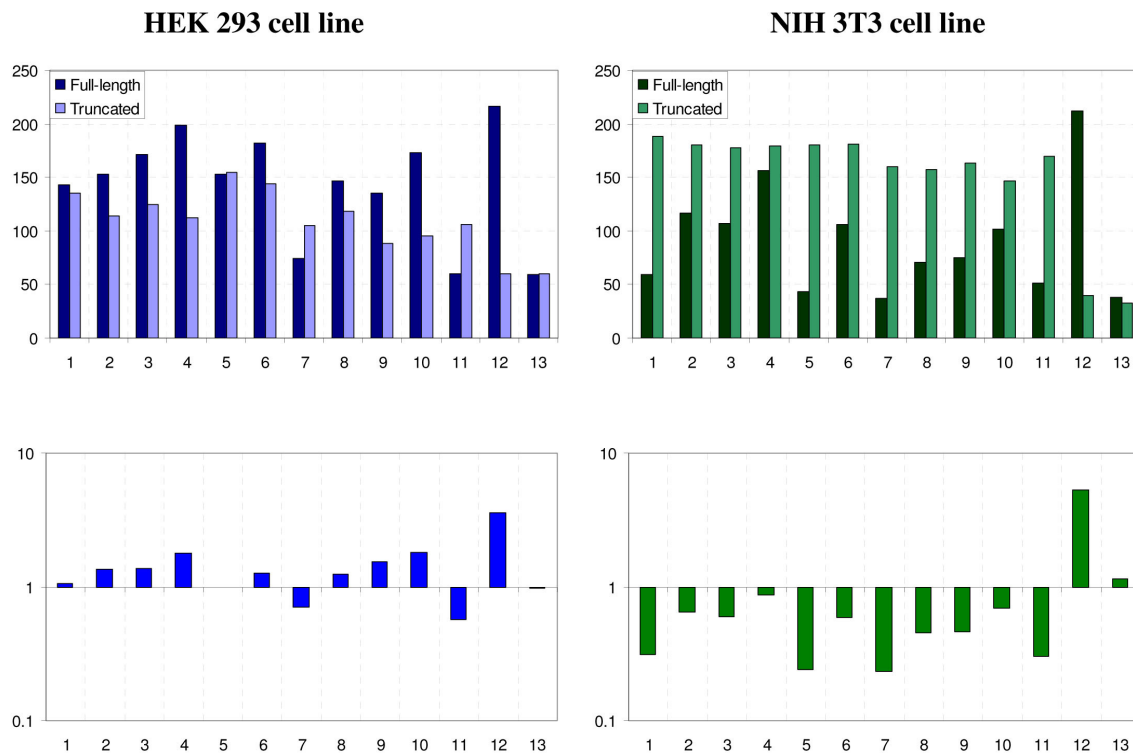


## Supporting Figure 4



**Fig. 4. Multiple sequence alignments of apicomplexan selenoprotein SelT.** Accession numbers of the sequences are as follows: AAH26350.2 (*H. sapiens*), NP\_001006557.2 (*G. gallus*), CAB01684.1 (*C. elegans*), NP\_915340.1 (*O. sativa*), BAD43801.1 (*A. thaliana*), BQ818029.1 (*C. reinhardtii*), TgESTzyi41b04.y1 and TgESTzyd07e11.y1 (*T. gondii*), and TC2223 and TC1872 (*N. caninum*). Selenocysteine residues (U) are highlighted in red and indicated by asterisks.

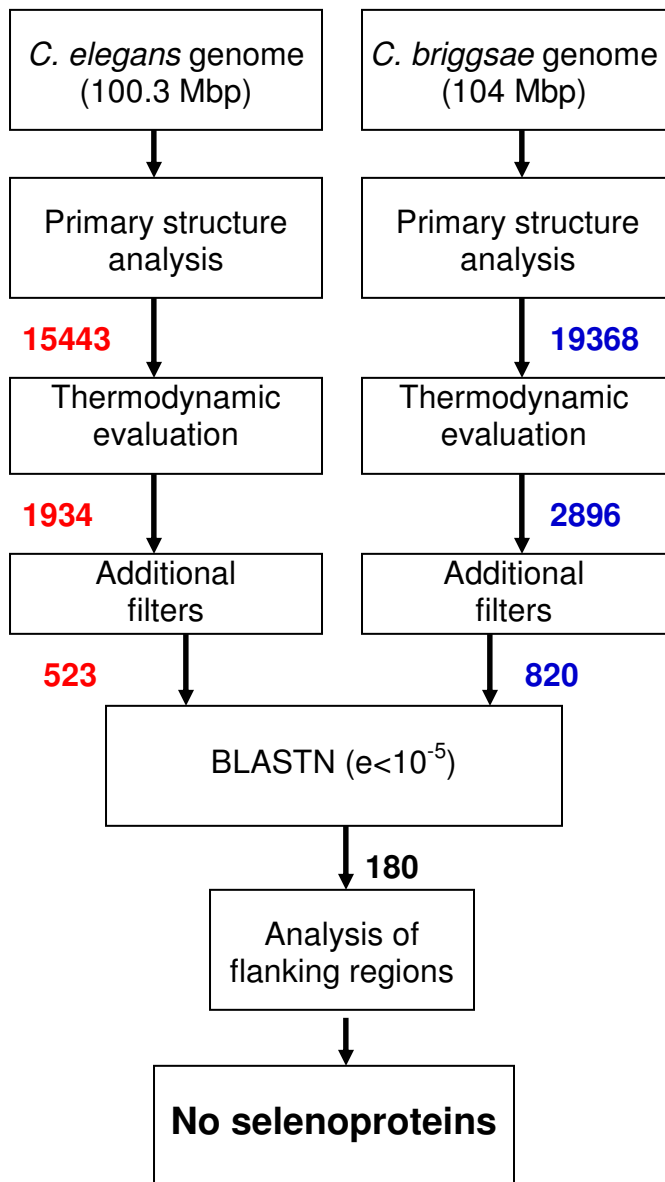
## Supporting Figure 5



**Fig. 5. Evaluation of band intensities in the western blots in Figure 4.**

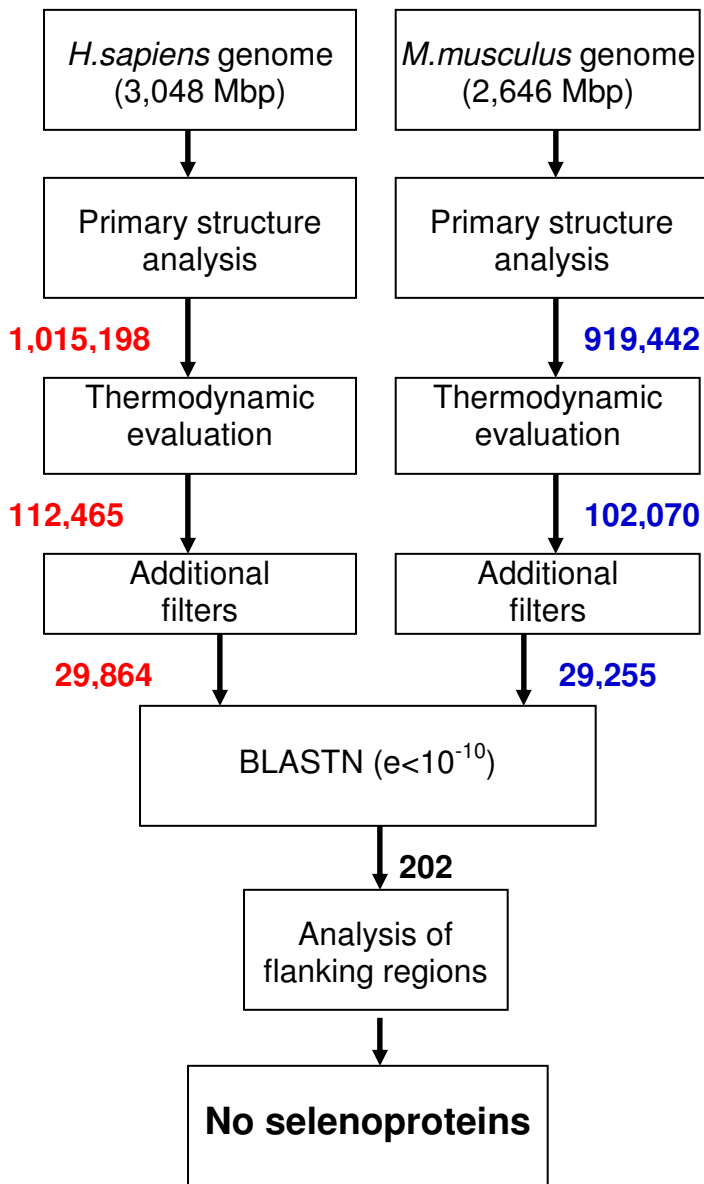
Quantification of bands for HEK 293 (blue) and NIH 3T3 (green) cells is shown in absolute values for each lane. Logarithmic scale is used for representation of intensity ratio of full-length and truncated forms of proteins (lower panels). Numbering is the same as in Figure 2. Scion Image 4.0 software (Scion Corporation) was used for image processing and analysis.

## Supporting Figure 6



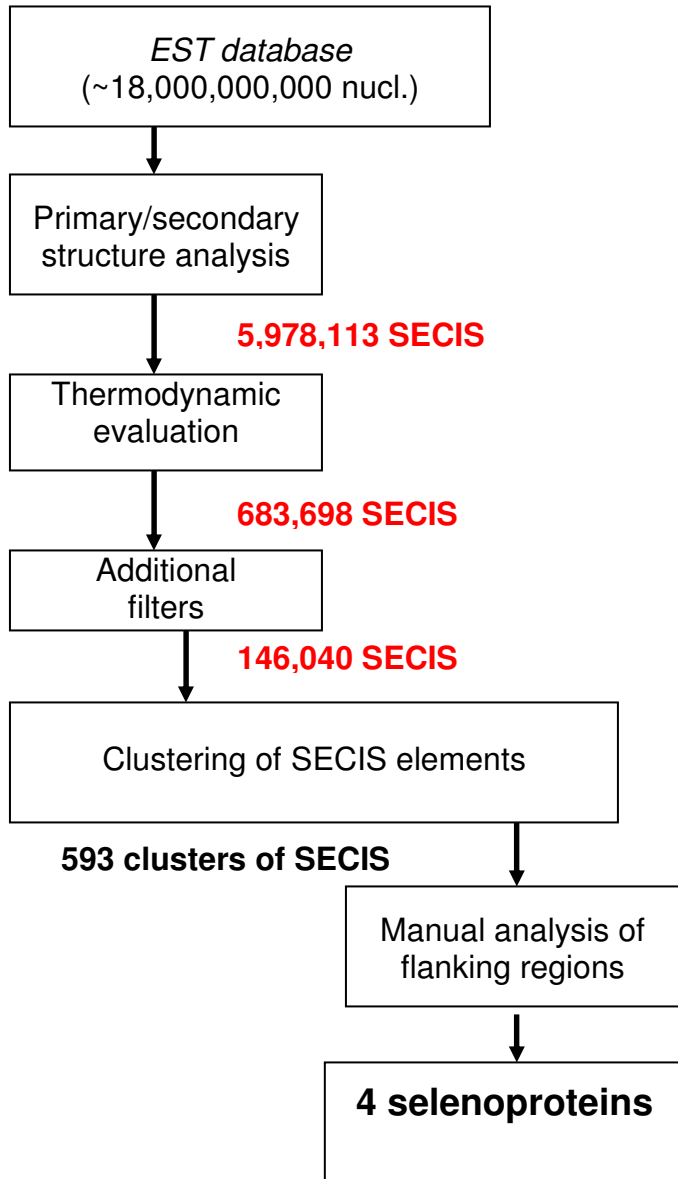
**Fig. 6. Analysis of nematode genomes with a modified version of SECISearch.** Each step in the search procedure is shown as a separate box, with the numbers of SECIS candidates indicated in red for *C. elegans* and in blue for *C. briggsae*.

## Supporting Figure 7



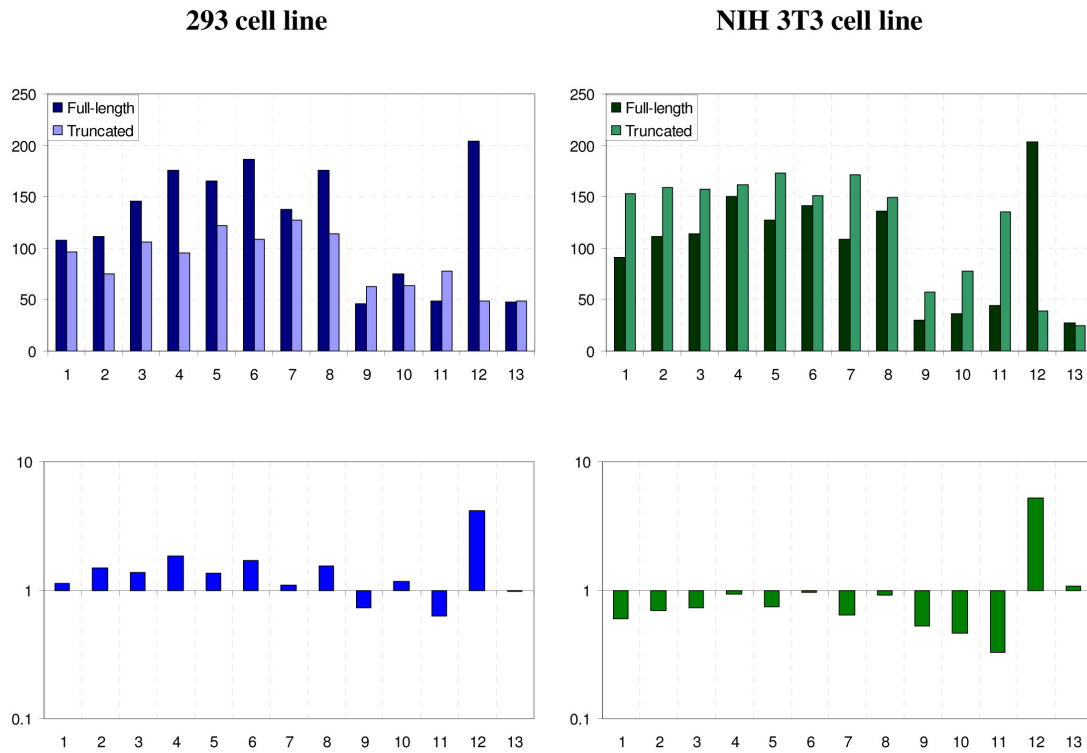
**Fig. 7. Analysis of human and mouse genomes.** Numbers shown in red correspond to SECIS candidates in *H. sapiens*, and those in blue to SECIS candidates in *M. musculus*.

## Supporting Figure 8



**Fig. 8. Analysis of NCBI EST database.** SECIS candidates identified in each step are indicated. Only SelT and SelS from *T. gondii* and *N. caninum* were identified in this search.

## Supporting Figure 9



**Fig. 9. Evaluation of band intensities in the western blots in Figure 4.**

Quantification of bands for HEK 293 (blue) and NIH 3T3 (green) cells is shown in absolute values for each lane. Logarithmic scale is used for representation of intensity ratio of full-length and truncated forms of proteins (lower panels). Numbering is the same as in Figure 4. Scion Image 4.0 software (Scion Corporation) was used for image processing and analysis.