

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Vadim Gladyshev Publications

Biochemistry, Department of

---

November 1999

## New Mammalian Selenocysteine-containing Proteins Identified with an Algorithm That Searches for Selenocysteine Insertion Sequence Elements

Gregory V. Kryukov

*University of Nebraska-Lincoln*

Valentin M. Kryukov

*University of Nebraska-Lincoln*

Vadim N. Gladyshev

*University of Nebraska-Lincoln*, [vgladyshev@rics.bwh.harvard.edu](mailto:vgladyshev@rics.bwh.harvard.edu)

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



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

---

Kryukov, Gregory V.; Kryukov, Valentin M.; and Gladyshev, Vadim N., "New Mammalian Selenocysteine-containing Proteins Identified with an Algorithm That Searches for Selenocysteine Insertion Sequence Elements" (1999). *Vadim Gladyshev Publications*. 68.

<https://digitalcommons.unl.edu/biochemgladyshev/68>

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.

Nucleotide sequences for human SelT gene (accession number AF195141) and mouse SelR gene (accession number AF195142) have been submitted to GenBank.

# New Mammalian Selenocysteine-containing Proteins Identified with an Algorithm That Searches for Selenocysteine Insertion Sequence Elements

Gregory V. Kryukov<sup>a</sup>, Valentin M. Kryukov<sup>a,b</sup>, and Vadim N. Gladyshev<sup>a,c</sup>

<sup>a</sup> Department of Biochemistry, University of Nebraska–Lincoln, Lincoln, Nebraska 68588-0664

<sup>b</sup> Laboratory of Genetic Enzymology, Institute of Biochemistry and Physiology of Microorganisms, Pushchino, Moscow 142292, Russia

<sup>c</sup> Corresponding author. Fax: 402-472-7842; Email: vgladyshev1@unl.edu.

**Abstract:** Mammalian selenium-containing proteins identified thus far contain selenium in the form of a selenocysteine residue encoded by UGA. These proteins lack common amino acid sequence motifs, but 3'-untranslated regions of selenoprotein genes contain a common stem-loop structure, selenocysteine insertion sequence (SECIS) element, that is necessary for decoding UGA as selenocysteine rather than a stop signal. We describe here a computer program, SECISearch, that identifies mammalian selenoprotein genes by recognizing SECIS elements on the basis of their primary and secondary structures and free energy requirements. When SECISearch was applied to search human dbEST, two new mammalian selenoproteins, designated SelT and SelR, were identified. We determined their cDNA sequences and expressed them in a monkey cell line as fusion proteins with a green fluorescent protein. Incorporation of selenium into new proteins was confirmed by metabolic labeling with <sup>75</sup>Se, and expression of SelT was additionally documented in immunoblot assays. SelT and SelR did not have homology to previously characterized proteins, but their putative homologs were detected in various organisms. SelR homologs were present in every organism characterized by complete genome sequencing. The data suggest applicability of SECISearch for identification of new selenoprotein genes in nucleotide data bases.

**Abbreviations:** Sec, selenocysteine; SECIS, selenocysteine insertion sequence; ORF, open reading frame; GFP, green fluorescent protein; UTR, untranslated region; dbEST, data base of expressed sequence tags; NR, non-redundant data base; PAGE, polyacrylamide gel electrophoresis; MSGS, mammalian selenoprotein gene signature.

## INTRODUCTION

Selenocysteine (Sec)<sup>1</sup> is the recently discovered 21st amino acid in protein (1–3). The genetic code word for this rare amino acid, which occurs in representatives of all life kingdoms, is UGA (3). The number of Sec-containing proteins varies in different organisms. The *Escherichia coli* genome contains three genes for Sec-containing proteins (4), and all three are formate dehydrogenases containing a single Sec residue. Seven genes for Sec-containing proteins were found in the archaeon, *Methanococcus jannaschii* (5), and only one selenoprotein has been described in the nematode, *Caenorhabditis elegans* (6, 7). Interestingly, no selenoprotein genes are present in the yeast genome of *Saccharomyces cerevisiae*.

Sec-containing proteins are more common in mammals, in which 14 selenoproteins have been found to date (8). These include four types of glutathione peroxidase (9, 10), three types of thyroid hor-

none deiodinase (11, 12), three types of thioredoxin reductase (13), selenophosphate synthetase 2 (14), selenoprotein W (15), selenoprotein P (16), and the 15-kDa selenoprotein (17). Thirteen of these proteins contain a single Sec residue that is conserved among mammalian sequences and is often present at the enzyme redox active center. Selenoprotein P, which is the major selenium-containing protein in plasma, is an exception in that it contains 10–12 selenocysteines residues depending on the host species (16).

The 3'-untranslated regions (UTRs) of all mammalian selenoproteins contain a stem-loop structure, designated as the selenocysteine insertion sequence (SECIS) element (18). This element is essential for recognition of a UGA codon within the coding region of selenoprotein mRNA as a signal for Sec incorporation (19). SECIS elements have several conserved features depicted in Figure 1, A–D. The general structure of the SECIS elements, and in particular the Quartet (also called a SECIS core) of non-Watson-Crick interacting nucleotides and a double A motif in the apical loop are essential for SECIS element function (20–23).

The presence of a Sec residue in regions that are essential to function of selenoproteins explains many, if not all, of the biological effects of selenium when this micronutrient is present at suboptimal levels in the diet. Selenium deficiency results in decreased levels of selenium-containing proteins (24), and insufficient selenium levels were also associated with a decreased survival rate of HIV-infected patients (25), increased rate of cancer incidence (26), and several other health disorders (8, 27). Supplementation of the human diet with selenium offers a potentially effective means of preventing or diminishing human maladies. For example, dietary supplementation with selenium resulted in 48–63% reduction in the incidence of human prostate, lung, and colon cancers in a human clinical trial (26). The essential role of selenium for mammalian development is illustrated by the findings that disruption of the mouse Sec-tRNA gene results in early embryonic lethality (28). It is not known which selenoprotein(s) is (are) responsible for this effect.

Identification and characterization of new Sec-containing proteins is an important area of research that may take advantage of the available genome sequencing data and help to elucidate many biological effects of selenium. No algorithms are currently available that would correctly predict selenoprotein gene sequence in nucleotide data bases because selenoproteins have diverse functions (8) and lack a common amino acid consensus sequence.

In this report, we describe a computer program, SECISEARCH, that recognizes SECIS elements in nucleotide sequences. Using this program and additional criteria for recognition of mammalian selenoprotein, we found the genes for two new Sec-containing proteins, SelT and SelR, which do not have homology to previously characterized proteins genes (V. N. Gladyshev and G. V. Kryukov, submitted for publication). In addition, we provide experimental evidence for the expression of SelT and SelR in mammalian cells. It should also be noted that SelR is the first selenoprotein that has a direct homolog in a "minimal gene set for cellular life" (30).

## EXPERIMENTAL PROCEDURES

**SECISEARCH** — A program for SECIS element recognition in nucleotide sequences, SECISEARCH, is based on the algorithm that involves three steps of a search strategy: 1) primary sequence identification; 2) secondary structure prediction; and 3) minimum free energy estimation (Figure 2A). The program is composed of two modules. The first module is responsible for recognition of the SECIS element primary sequence and secondary structure consensus and is written on the basis of the PatScan program. Nucleotide sequences that satisfy both primary and secondary structure constraints are further analyzed in the second module that estimates the free energy of the stem-loop structure. This module is based on the Vienna RNA package RNAfold program for secondary structure prediction and free energy evaluation. It separately estimates the free energies for Helix I plus internal loop and Helix II plus apical loop regions of the putative SECIS element (Figure 2B). The free energy cut-off parameters are based on free energy minimization calculations of SECIS elements in previously characterized human selenoprotein mRNAs and were set as  $-7$  kcal/mol for Helix I and internal loop and  $-5$  kcal/mol for Helix II and apical loop. In the absence of a reliable algorithm that correctly predicts the three-dimensional structure of mRNA, these numerical values are relative but reflect the relationship between the values of the free energy of folding predicted for SECIS elements in previously characterized selenoproteins and stem-loop structures selected by SECISEARCH. The minimum free energy algorithm derived from the Vienna RNA package and employed by SECISEARCH is also present in the popular *mfold* program (31). The two modules of SECISEARCH were written in C and compiled by the GNU C compiler for Win32 platform. Perl scripts were used for module interaction and for presentation of the SECISEARCH data for further analyses.

**Computer Searches** — Human expressed sequence tag data base (dbEST) (March 1999 release) was searched with SECISEARCH (Figure 2A) with parameters indicated in Table I. SECIS elements in the SelP gene were obvious out-rangers in terms of Helix II plus apical loop energy (Figure 2B). The use of energy cut-off parameters that excluded these SECIS elements allowed us to reduce the number of sequences selected by SECISEARCH over 6-fold and increase the proportion of sequences that corresponded to known selenoprotein genes from  $\sim 10$  to  $\sim 70\%$ . Nucleotide sequences that were selected by the program were manually analyzed against NCBI non-redundant (NR) and EST data bases and against the TIGR tentative human consensus data base. Sequences corresponding to the known selenoprotein genes were excluded, remaining sequences were further extended by computer search analyses of dbEST, and consensus sequences were obtained on the basis of multiple matches in all three data bases. Extended sequences were analyzed for the presence of open reading frames (ORFs) allowing in-frame TGA codons to be interpreted as either Sec residues or stop signals. ORFs containing in-frame TGA codons were identified and analyzed by computer search analyses for the presence of features characteristic of mammalian selenoprotein genes. Specifically, conservation of a TGA codon in nucleotide sequences, putative Sec-flanking areas in amino acid sequences and SECIS elements in 3'-UTR were tested in mammalian sequences. ORFs were further searched for homologies to NR and EST sequences using BLAST programs, and sequences homologous to putative selenoprotein sequences and containing a cysteine codon in place of a Sec-encoding TGA codon in a putative selenoprotein were identified.

**cDNA Sequencing** — EST clones (number 590409 from "Stratagene endothelial cell 937223" library, accession number AA156969, and number 736225 from "Soares mouse 3NME12 5'" library, accession number AA270410) were obtained from Research Genetics, Inc. Plasmids were isolated using a Nucleobond AX100 Kit (CLONTECH). The nucleotide sequences were determined using the Dye Terminator Cycle Sequencing method.

**Constructs with Green Fluorescein Protein (GFP)** — GFP-SelT and GFP-SelR constructs were made on the basis of pEGFP-C3 expression vector (CLONTECH). In each of these constructs, GFP was located upstream of a selenoprotein gene. Human SelT cDNA was amplified with primers W5U119, 5'-GTGCCAGCAAGAGATTAAG-3', and W5L791, 5'-GCCAGCACT-GAAACACTATC-3', and cloned into the *SmaI* site of pEGFP-C3. The SelR cDNA was directly cloned into the *XhoI/BamHI* sites of pEGFP-C3. Both plasmids were transformed into *E. coli* strain NovaBlue (Novagen), and the plasmids were isolated using Plasmid Maxi Kit (Qiagen).

**Cell Growth, Transfection, and Metabolic Labeling with  $^{75}\text{Se}$**  — A monkey CV-1 cell line was grown on Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum to  $\sim 80\%$  confluence and transfection was carried out using LipofectAMINE reagent (Invitrogen) according to the manufacturer's protocol for attached cells. 3  $\mu\text{g}$  of DNA and 15  $\mu\text{l}$  of LipofectAMINE were used for each 60-mm plate. Cells were labeled between 6 and 36 h after transfection with 100  $\mu\text{Ci}$  of freshly neutralized [ $^{75}\text{Se}$ ]selenious acid (1000 Ci/mmol, Research Reactor Facility, University of Missouri, Columbia, MO) as described (17). Cells were washed four times from the remaining  $^{75}\text{Se}$  with phosphate-buffered saline, harvested, and the samples analyzed on 10% SDS-NuPAGE gels (Novex).  $^{75}\text{Se}$ -Labeled proteins were visualized on SDS-PAGE gels with a Storm PhosphorImager system (Molecular Dynamics).

**Immunoblot Detection** — Rabbit polyclonal antibodies were raised against a synthetic polypeptide corresponding to residues 148–163 of the human SelT protein. This peptide was conjugated to a keyhole limpet hemocyanin before injection into a rabbit. Western blot analyses were performed with an ECL system (Amersham Pharmacia Biotech).

## RESULTS

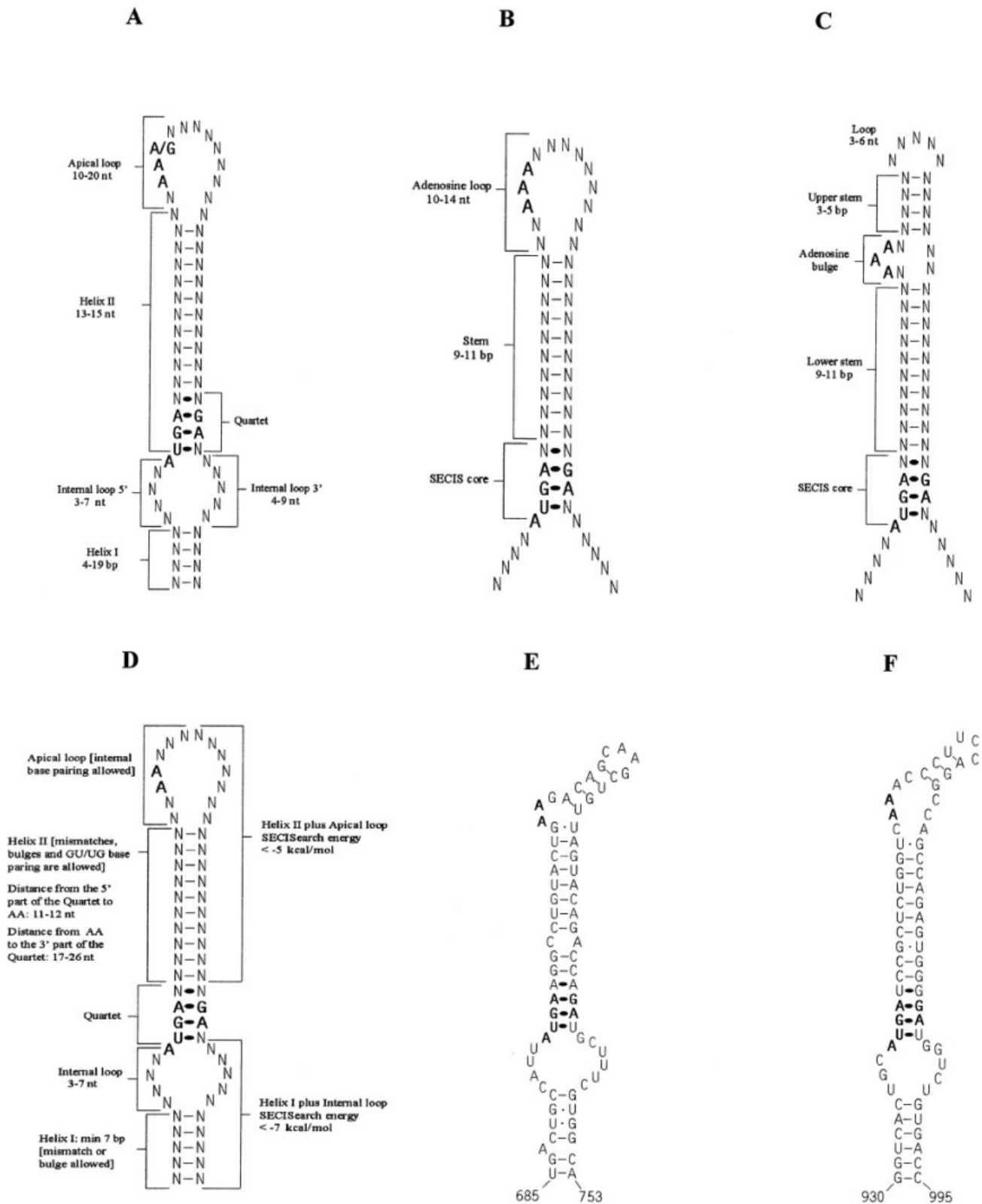
**SECIS Element Consensus Structure** — Mammalian SECIS elements (18–22) are composed of Helices I and II, internal and apical loops, and a non-Watson-Crick base paired SECIS core, Quartet (Figure 1, A–D). Conserved nucleotides in the SECIS element sequence are an A directly preceding the Quartet, TGA in the 5' segment and GA in the 3' segment of the Quartet, and an unpaired AA in the apical loop. Another characteristic feature of a mammalian SECIS element is the length of Helix II, which separates the non-Watson-Crick Quartet and AA in the apical loop by 11–12 nucleotides.

Two distinct but related models were recently proposed for the SECIS element consensus sequence and structure. Krol and collaborators (20) suggested a single consensus for all mammalian SECIS elements (Figure 1A), while Berry and collaborators (22) divided SECIS elements into two distinct subfamilies, type I and type II (Figure 1, B and C). These subfamilies were different in the area of the apical loop; that is, type II SECIS elements had an additional mini-stem that placed AA in the bulge. These two types of SECIS elements were interconvertible by mutations that remove/create the mini-stem suggesting a similar structure and function of both SECIS element types (22).

Even though the consensus sequences shown in Figure 1, A–C, indeed represent SECIS elements in selenoprotein mRNA, we analyzed known SECIS elements from a different perspective and propose a consensus that utilizes the free energy of a stem-loop structure in addition to primary and secondary consensus structures (Figure 1D). When free energies predicted from a computer analysis of SECIS element folding were analyzed for known SECIS elements, all Sec-inserting stem-loop structures exhibited similar free energy parameters. The free energy values for Helix I plus internal loop and for Helix II plus apical loop were determined separately and plotted against each other (Figure 2B). No differences were observed in the free energy values for type I and type II SECIS elements shown in Figure 1, B and C, suggesting that the mini-stem serves to stabilize the stem-loop structure. The mini-stem is formed when an apical loop is large enough to destabilize the structure, but it may not be required for the SECIS elements with smaller apical loops. These considerations are similar to the suggestion that the mini-stem may serve to maintain

thermodynamic stability or to nucleate SECIS element folding (22). The commonality in the energetic criteria for SECIS elements allowed us to maintain a single model, depicted in Figure 1D, for designing an algorithm that searches for novel SECIS elements. Additional features

that distinguish the SECIS element shown in Figure 1D, from previously proposed models, Figure 1, A–C, are: (a) assigning a distance of 11–12 nucleotides between the Quartet and AA in the apical loop instead of a 9–11-base pair stem (20, 22); (b) identifying a distance be-



**Figure 1. Mammalian SECIS elements.** A, SECIS element consensus proposed by Krol and collaborators (20, 21). B, type I SECIS element consensus proposed by Berry and collaborators (19, 22). C, type II SECIS element consensus proposed by Berry and collaborators (19, 22). D, SECIS element consensus used in SECISearch. E, SECIS element in human SelT mRNA. F, SECIS element in human SelR mRNA.



tween the unpaired AA and the 3' segment of the Quartet; and (c) the use of the unpaired AA instead of AA(A/G) in the apical loop.

**SECISearch** — A computer program, SECISearch, was developed for identification of SECIS elements in nucleic acid sequences. The algorithm includes three major steps: 1) primary consensus sequence search; 2) analysis of mRNA secondary structure; and 3) estimation of the free energy for the predicted secondary structure. The descriptor was developed on the basis of the SECIS element consensus (Figure 1D) and adjusted to recognize SECIS elements in known selenoproteins. The free energy values were estimated for known SECIS elements and the lowest negative free energy cutoff parameters were determined (Figure 2B). The online version of SECISearch was also developed and designed to provide researchers with a tool to test nucleotide sequences for the presence of potential SECIS elements. It should be noted, however, that recognition of a SECIS element in a nucleotide sequence by SECISearch does not identify the sequence as a portion of a selenoprotein gene containing a functional SECIS element. Instead, this analysis is a first step in testing nucleotide sequences for the presence of a SECIS element, and further experimental and/or computer analyses are required (see below).

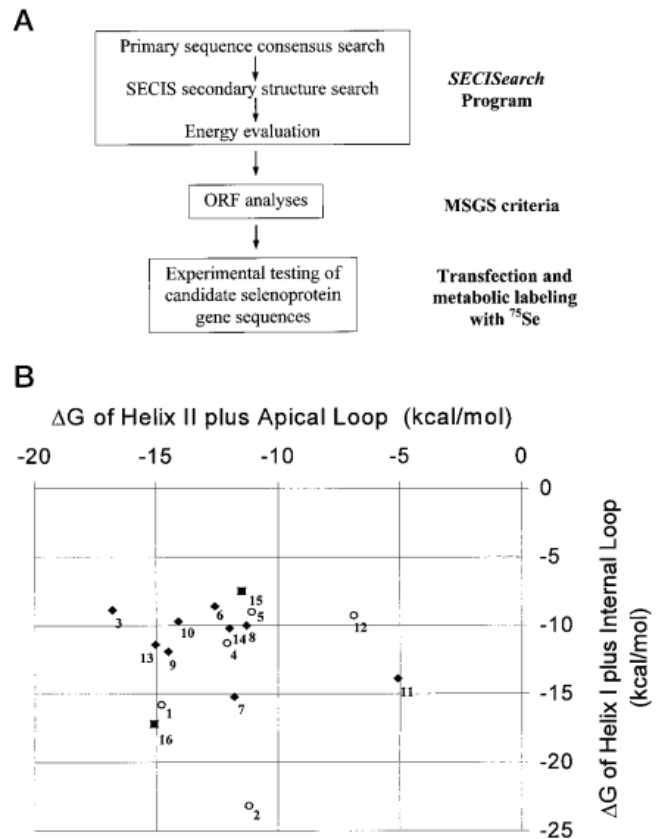
**Computer Search** — SECISearch was applied to search dbEST and NR for nucleotide sequences containing SECIS elements. The search strategy involving this program and further analysis of ESTs selected by SECISearch is summarized in Figure 2A. SECISearch was highly selective for authentic SECIS elements. Depending on the parameters applied for a particular search (free energy of Helix I/internal loop, free energy of Helix II/apical loop, length of Helices I and II, etc.), ~10–70% of the sequences selected by the program in dbEST corresponded to SECIS elements in previously characterized selenoproteins genes. The example of SECISearch analysis of human dbEST is shown in Table I. The primary sequence consensus step reduced the number of tested ESTs to 14.3%, and only 2.6% (32,652 ESTs) of the initial 1,253,123 sequences satisfied the secondary structure consensus. The criteria used in primary sequence and secondary structure searches were sufficient to detect all known selenoproteins in NR. Calculations of the free energy for the predicted stem-loop structures with thermodynamic parameters that satisfy 12 out of 13 selenoproteins with known SECIS elements reduced the number of ESTs to 0.078% (974 individual ESTs). Eleven known selenoproteins were represented in this subset of human ESTs by 678 sequences. The remaining 296 ESTs were further grouped by multiple sequence alignments and each group was manually analyzed with the help of BLAST programs resulting in detection of two new selenoproteins (Figure 1, E and F) that were represented by 35 ESTs.

However, not all of the SECISearch-selected sequences were derived from selenoprotein genes. This is likely due to a low degree of sequence and structure conservation in mammalian SECIS elements and perhaps due to a limited knowledge of actual SECIS element structure and of additional cis-acting elements involved in Sec insertion into mammalian proteins.

Thus, in addition to functional SECIS elements, non-functional, “pseudo” SECIS elements were found by the current version of SECISearch during analyses of nucleotide data bases. A number of such SECIS-like non-functional structures were immediately recognized due to their presence in 5'-UTRs and coding regions (as well as in 3'-UTRs if translation was terminated at TAA or TAG) of known proteins. However, pseudo SECIS elements may also be present in the 3'-UTR of proteins in which TGA signals terminate translation, but certain properties of mRNA, such as the distance between TGA and the SECIS element, or mRNA tertiary structure, prevent SECIS element from signaling Sec incorporation. It has been established, for instance, that the TGA-SECIS element distance of <51 nucleotides was insufficient in decoding of TGA as Sec (23), while the distance

of >204 nucleotides was sufficient (32). Analyses of SECISearch-selected ESTs were also occasionally complicated by the uncertainty of the correspondence of an entry to a coding or complementary strand.

However, we found that, even if the majority of SECISearch-selected ESTs were classified as SECIS elements in known selenoproteins or as pseudo-SECIS structures, the number of remaining nucleotide sequences was typically too large to test in a cell line system. Indeed, experimental screening of the subset of nucleotide sequences that are selected by SECISearch but cannot be judged for authenticity (functionality) would require enormous experimental effort. This suggests that the current version of SECISearch is not efficient in



**Figure 2. SECIS element and selenoprotein searches.** *A*, strategy used to search for new selenoprotein genes in nucleotide data bases. Nucleotide sequences in dbEST were analyzed with the SECISearch program that subsequently searched for the primary consensus sequence, secondary consensus structure, and the free energy parameters characteristic of SECIS elements in known selenoprotein genes. Selected EST sequences were manually analyzed for the presence of open reading frames that satisfy MSGS criteria. Candidate selenoproteins were experimentally characterized by transfection of a mammalian cell line with a GFP-selenoprotein construct and detecting the fusion protein by  $^{75}\text{Se}$  labeling. *B*, free energy plot for SECIS elements found in human selenoprotein mRNAs. Free energy values,  $\Delta G$  (in kcal/mol), for Helix I plus internal loop and for Helix II plus apical loop were calculated separately for each SECIS element and plotted against each other. *Open circles* indicate Type I, and *closed diamonds*, Type II SECIS elements proposed by Berry and collaborators (22). *Large squares* indicate new selenoproteins, SelT and SelR, identified with SECISearch. 1, cytosolic glutathione peroxidase; 2, gastrointestinal glutathione peroxidase; 3, plasma glutathione peroxidase; 4, phospholipid hydroperoxide glutathione peroxidase; 5, thyroid hormone deiodinase 1; 6, thyroid hormone deiodinase 2 (DI2); 6, thyroid hormone deiodinase 3 (DI3); 8, thioredoxin reductase 1; 9, thioredoxin reductase 3 (TR3); 10, selenophosphate synthetase 2; 11, Selenoprotein P1 (SelP1) first SECIS element; 12, SelP1s SECIS element; 13, selenoprotein W1; 14, the 15-kDa selenoprotein; 15, SelT; 16, SelR.

**Table I. An example of SECISearch analysis of human dbEST**

Primary sequence and secondary structure parameters in this search detected all selenoproteins with known SECIS elements in NR and dbEST. The free energy parameters in this search detected 12 out of 13 selenoproteins. SECIS elements in SelP did not satisfy the energy threshold for Helix I used in this search. The number of human ESTs that was selected by SECISearch with parameters sufficient to find 13 human selenoproteins was 6,055.

Human ESTs in dbEST	1,253,123
Satisfied primary sequence consensus ATGAN/11–12 nt/AA/18–27 nt/GA	179,614
Satisfied secondary structure consensus Helix I, 7 nt, 1 mismatch and 1 bulge on each side are allowed; internal loop, 3–7 nt, 59 branch ends with A; Quartet, TGAN. . . NGAN; Helix II, 11–12 nt, 3 mismatches and 1 bulge on each side are allowed; apical loop, 7–17 nt, starts with AA	32,652
Satisfied free energy threshold $\Delta G$ for Helix Ia plus internal loop, ,7.4 kcal/mol; $\Delta G$ for Helix II plus apical loop, ,11.0 kcal/mol	974
Corresponded to known selenoprotein genes	678
Analyzed with MSGS	296
Corresponded to <i>SelR</i> gene	33
Corresponded to <i>SelT</i> gene	2

*a* Ten additional nucleotides upstream of the 5' branch and 10 nucleotides downstream of the 3' branch of predicted Helix I were considered in free energy estimation for Helix I plus the internal loop. This reflects the fact that, in most SECIS elements, Helix I is longer than the 7 nt assigned in the consensus secondary structure, and that Helix I could not be accurately predicted from that consensus.

finding new selenoprotein mRNAs exclusively and that identification of additional cis-acting elements may be necessary for selenoprotein gene sequence recognition.

Analysis of putative Sec-flanking areas in cDNAs selected by SECISearch provides such additional criteria for selection of selenoprotein genes. We recently proposed a set of criteria, designated as mammalian selenoprotein gene signature (MSGS),<sup>2</sup> that are helpful for identifying selenoprotein mRNA sequences. According to MSGS, a typical new selenoprotein will not only contain a SECIS element in the 3'-UTR of its mRNA, but this SECIS element will also be conserved among mammalian mRNAs for this protein. In addition, Sec and Sec-flanking regions for this protein should be conserved in mammalian amino acid sequences for this protein, and homologous sequences for this protein (most often in lower eukaryotes) should contain a cysteine residue in place of Sec (or these homologous sequences should conserve Sec). These criteria are consistent with the sequences of all known mammalian selenoproteins. It should be noted, however, that selenoprotein P contains both conserved and non-conserved Sec residues, and this protein only satisfies MSGS criteria based on its conserved selenocysteines.

Application of MSGS criteria to the remaining unclassified nucleotide sequences selected by SECISearch allowed us to classify a significant proportion of these sequences as those that are not derived from selenoprotein cDNAs. Still, a number of sequences in dbEST could not be classified, because these sequences were represented by an insufficient number of ESTs, were incomplete or erroneous, or their homologous sequences were not detected in other organisms. Accumulation of sequence data in nucleotide data bases will help us to classify these EST sequences.

*Identification of SelT and SelR* — Two sequences that were initially selected by SECISearch (Figure 1, *E* and *F*) were represented by a large number of EST sequences, which allowed us to obtain complete human cDNA sequences for these proteins from multiple sequence alignments of ESTs. Analysis of open reading frames present in cDNAs revealed the presence of in-frame TGA codons in the coding regions and the presence of SECIS elements in 3'-UTRs. Further protein homology analyses found no homology to previously characterized proteins, but revealed characteristics that satisfied both SECISearch and MSGS criteria, suggesting that these sequences encoded selenoproteins. The new proteins were designated as SelT (Figure 3*A*) and SelR (Figure 3*B*).

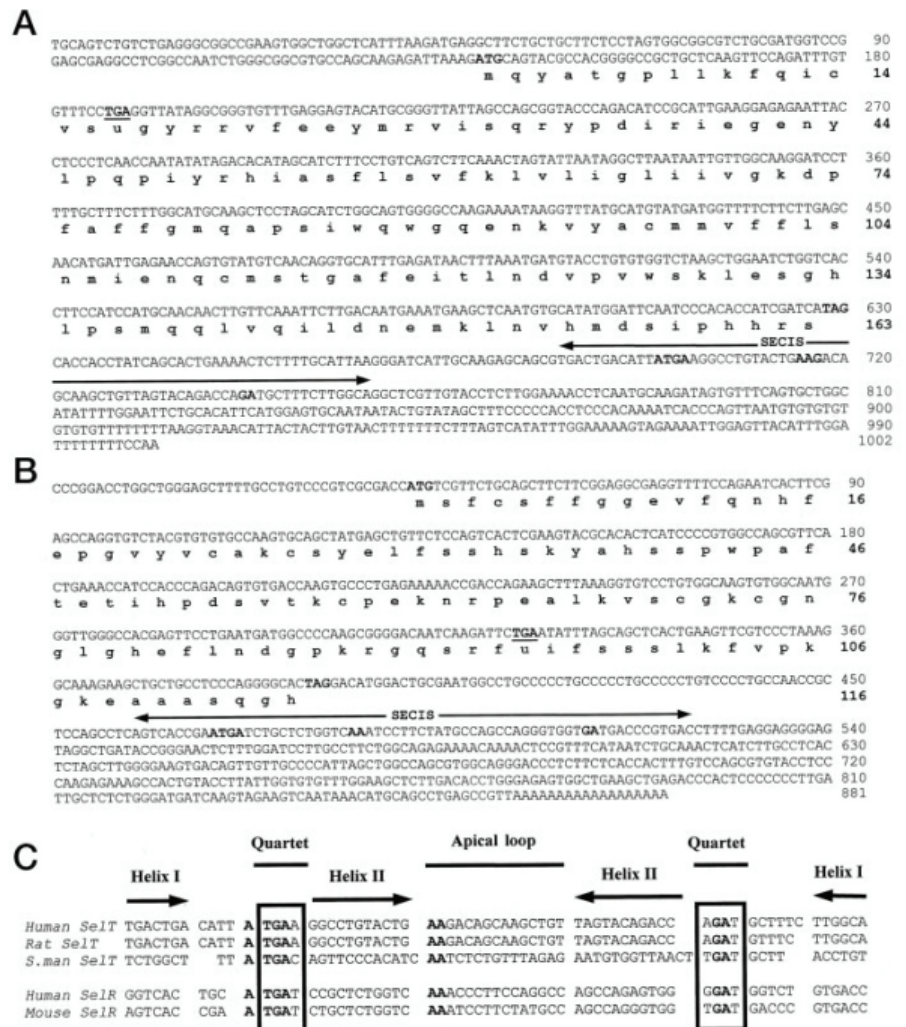
*SelT* — The human EST clone containing SelT cDNA was obtained. Its sequence was experimentally determined and was in agreement with the cDNA consensus sequence obtained by multiple sequence alignments of ESTs. The cDNA sequence of SelT (Figure 3*A*) was 1002 nucleotides long and contained an ORF of 163 amino acid residues with a calculated mass of 18.8 kDa. The Sec residue, Sec17, encoded by TGA, was located in the N-terminal portion of the protein. The SECIS element (Figure 1*E*) was located 509 nucleotides downstream of the TGA codon for Sec determined as the distance between a SECIS core, the Quartet, and a Sec codon.

Further computer analyses of the SelT amino acid sequence revealed the presence of homologous sequences in other animals (Figure 4*A*) as well as in plants. Interestingly, Sec was present in SelT homologs in *Schistosoma mansoni* and zebrafish, while Cys was present in place of Sec in SelT homologs in *Drosophila melanogaster*, *C. elegans* (Figure 4*A*), and *Arabidopsis thaliana*. The SelT region containing Sec had a high degree of homology, and one of the conserved residues, Cys-14, was separated from Sec by two other amino acid residues. This putative redox center, CXXU, was similar to that found in thioredoxins and glutaredoxins (CXXC) (33), selenoprotein W (CXXU) (15), and several other redox active proteins.

Table II, A, shows the incidence of SelT EST clones in human tissues and organs for which at least one library has three or more independent SelT cDNAs in human dbEST. In addition, 18 other cDNA libraries from a variety of tissues and organs were represented by ESTs containing the *SelT* nucleotide sequence. The data suggest that SelT mRNA is expressed at low levels in a broad range of tissues and organs.

*SelR* — Nucleotide sequences corresponding to human SelR cDNA were abundant in dbEST and the complete human SelR cDNA sequence was obtained by multiple sequence alignments. Human SelR had a calculated mass of 12.6 kDa. The nucleotide sequence of a mouse EST clone containing a full-length cDNA sequence for SelR was experimentally determined revealing an ORF that was highly homologous to the human SelR (Figs. 3*B* and 4*B*). The SECIS element (Figs. 1*F* and 3*C*) was located in the 3'-UTR, 142 and 554 nucleotides downstream of the Sec codon in mouse and human sequences, respectively. Sequences homologous to *SelR* were detected in other genomes. Mammalian SelR contained Sec, while homologs in non-mammalian eukaryotes (animals, plants, and yeast) and in prokaryotic organisms contained Cys in place of Sec (Figure 4*B*). We have

**Figure 3. cDNA sequences encoding SelT and SelR.** *A*, human SelT cDNA and protein sequences. *B*, mouse SelR cDNA and protein sequences. TGA codons that encode Sec (shown as U) in coding regions are *underlined*. Location of SECIS elements in 3'-UTR are shown by *arrows* and nucleotides that represent the primary sequence SECIS element consensus are shown in *bold*. *C*, alignment of the predicted SECIS elements of the mRNAs encoding SelTs (human, rat, and *S. mansoni* sequences) and SelRs (human and mouse sequences). Accession numbers for ESTs corresponding to rat SelT, *S. mansoni* SelT, and human SelR SECIS element sequences are AI231051, AI395-351, and AA896979, respectively.



also detected the sequences of two mammalian proteins, SelR-c1 and SelR-c2, that were homologous to SelR (Figure 4B). The ORFs for these proteins were assembled on the basis of multiple sequence alignments of EST sequences. These proteins contained Cys in place of Sec in SelR. Human SelR-c1 cDNA sequence has recently appeared in GenBank (accession number AA038899).

Interestingly, SelR genes had direct homologs in all completely sequenced genomes, including several bacterial and archaeal genomes as well as yeast *S. cerevisiae* and nematode *C. elegans* genomes. A direct homolog of the SelR gene was present in a minimal gene set for cellular life. This set of 256 protein-coding genes was initially obtained by direct comparison of the smallest (468 protein genes, *SelR* homolog is the MG446 gene) known genome, Gram-negative bacterium *Mycoplasma genitalium*, with that of Gram-positive bacterium *Hemophilus influenzae* (30). The minimal gene set contained the protein genes that are thought to be necessary and sufficient to sustain cellular life (30). Most of proteins encoded in a "minimal gene set" have previously been characterized, but the function of the SelR homolog is not known.

The incidence of human SelR mRNA expression that was calculated as the number of SelR ESTs per 10,000 ESTs in a particular cDNA library is shown in Table II, B. SelR mRNA exhibits moderate levels of expression and is present in a variety of adult and fetal tissues. Estimation of mRNA expression levels as EST incidence is only semi-quantitative since many cDNA libraries represented in dbEST are normalized.

Genomic sequence for human SelR was obtained by searching NR with a human SelR cDNA sequence as template. One genomic clone, AC005363, contained the complete 5-kb SelR genomic DNA sequence that was organized in 4 exons and 3 introns. The 5'-UTR and the initiation codon were located in the first exon, while Sec-encoding TGA and the entire 3'-UTR were present in the last exon. The genomic clone was derived from human chromosome 16p13.3.

*Detection of SelT and SelR as Fusion Proteins with GFP* — To demonstrate the occurrence of SelT and SelR in mammals, we initially expressed the proteins as fusion proteins with GFP. Coding regions of human SelT and mouse SelR cDNAs were cloned in a pEGFP-C3 vector that had a GFP gene upstream of the cloned protein. This allowed us to express the proteins of a higher molecular weight than the predicted naturally occurring SelT and SelR. The expected masses of the GFP-SelT and GFP-SelR fusion proteins were ~48.9 and ~42.6 kDa, respectively. No selenoproteins of similar masses were previously detected in mammalian cell lines, a feature that helped in easy detection of SelT and SelR fusion proteins by metabolic labeling of cells with <sup>75</sup>Se. Hence, monkey CV-1 cells were transfected with *GFP-SelT*, *GFP-SelR*, and control plasmids, incubated in the presence of <sup>75</sup>Se, and <sup>75</sup>Se-labeled proteins were detected by SDS-PAGE gels and PhosphorImager analyses. Distribution of selenoproteins in CV-1 cells transfected with a control plasmid (Figure 5A, lanes 2 and 3) was similar to other mammalian cell types where <sup>75</sup>Se-labeled proteins of 57 kDa (thioredoxin reductase) and 25 kDa (glutathione peroxidase) were among the most abundant



**Figure 4. Multiple alignments of mammalian SelT and SelR with their homologs.** *A*, multiple alignment of human and mouse SelTs with homologs from other species. The mouse SelT sequence was assembled from 8 and *D. melanogaster* sequence from 12 independent ESTs. The accession number for *C. elegans* putative protein is CAB01692. *S. mansoni* sequence was obtained by conceptual translation of a single EST (accession number AI067883). Zebrafish sequence was assembled from 3 ESTs (accession numbers AI497309, AI477145, and AI617064). *B*, multiple alignment of human and mouse SelR with eukaryotic, bacterial and achaeal homologs. GenBank accession numbers for *M. pneumoniae*, *E. coli*, *M. thermoautotrophicum*, *S. cerevisiae*, *A. thaliana*, and *C. elegans* SelR homologs, and for human SelR-c1 protein are P75129, P39903, AAB85216, P25566, CAA17151, P34436, and AA038899, respectively. Human SelR-c2 sequence was assembled from five independent ESTs.

**A**

```

Human selT ~~~~~
Mouse selT ~~~~~
Zebrafish ~~~~~
S.mansoni ~~~~~
C.elegans MRIHDELQKQ DMSRFGVFII GVLFFMSVCD VLRTVSAAEH SHDENHVHEK 50
D.melanogaster ~~~~~

Human selT ~~~~~M QYATGELLKF 11
Mouse selT ~~~~~M QYATGELLKF 11
Zebrafish ~~~~~M QFATGELLKF 11
S.mansoni MANIGMNAVV VFGCAILTWF DFWGLKGSNS SEKTIPIKK PINPILQF 50
C.elegans DDFEAEFGDE TDSQSFSQGT EEDHIEVREQ SSEVKPTAVH HAKDILTRI 100
D.melanogaster ~~~~MERLTG RNVALLVLCL CAGYALVFAE GEKEI PVTKF GQNIATHTTF 46

Human selT QIVVUGVRR VEEYMRVIS QRYEDIRIEG ENYLEQPIYR HIASFLSVFK 61
Mouse selT QIVVUGVRR VEEYMRVIS QRYEDIRIEG ENYLEQPIYR HIASFLSVFK 61
Zebrafish QIVVUGVRR VEEYTQALY QRYEDIRIEG ENYLEPLLYR HIASFLSMFK 61
S.mansoni MGVVUGVRR VEEEMSRIVL QRYVSVFIEG DTYPFAWRRA HLAKSLQLFL 100
C.elegans FTVVSCGKQ AADQFTTFAK EKYNMFIEG ANFAPLWKA YVAQALSFKV 150
D.melanogaster LVVYSCGKQK AEDYVGLLG EKYEIQVNS GNYDFGLNY YLSKMIFFAL 96

Human selT LVLIGLIIVG KDPFAFFMQ APSIQWQGE NKVYAMMVV FLSNMIENQC 111
Mouse selT LVLIGLIIVG KDPFAFFMQ APSIQWQGE NKVYAMMVV FLSNMIENQC 111
Zebrafish LLLIGVILG KDPFALCMQ AFGIWWVSCQ NKVYAMMVV FFSNMIENQC 111
S.mansoni YSIIIMLVFG FDLGYLEYP TPNFVSYAMQ NKVSLMFTF LLQWVIEGQL 150
C.elegans MAVLVLVGG INPFRFSLG YPQILQAHG NKMSLMLVF MLGNLWQSL 200
D.melanogaster IIIIVSVVSA VSPFTFLGLN TFSNWSHQQA NKVYAMMVV FLGNMLVQL 146

Human selT MTFGAFEITL NDVPVWSKLE SGHLFSMQQL VQILDNEMKL NVHMDSIPHH 161
Mouse selT MTFGAFEITL NDVPVWSKLE SGHLFSMQQL VQILDNEMKL NVHMDSIPHH 161
Zebrafish MTFGAFEITL NDVPVWSKLE SGHLFSMQQL VQIL----- 145
S.mansoni LDTGAFEITL DDIPVWSKLY TNRIQPEEL LHIIDNHVKL RDSTKTIQSV 200
C.elegans ISTGAFEVYL GNEQIWSNIE SGRVSPQEF MQLIDAQALV LGKAPVNTES 250
D.melanogaster IESTGAFEITL NDVPVWSKLY TGRFSEPEVL FQIIDNHLQF TEKVQENPDF 196

Human selT RS----- 163
Mouse selT RS----- 163
Zebrafish ----- 163
S.mansoni GPPL----- 204
C.elegans FGEFQQTV ----- 258
D.melanogaster VK----- 196

B
Human SelR ~~~~~
Mouse SelR ~~~~~
Human SelR-c1 MARLLWLLRG LTLGTAPRRA VRQAGGGGP GTGPGLGEG SLATCELPLA 50
Human SelR-c2 ~~~~~VFS 3
M.pneumoniae ~~~~~MSKYQKKS 8
E.coli ~~~~~MANKPS 6
M.therm. ~~~~~MKDR IPFVSVAKRN VEMVERIELS 24
S.cerevisiae ~~~~~MKNWSRLYV ITVVRTFPGR RNIVLTQYWN KSKKMSDESN 39
A.thaliana ~~~~~MAMTAAA VPSGGSFQKQ 17
C.elegans ~~~~~MTTKKFR MEDVGLSKLK VEKNPK-DVK 26

Human SelR ~~~~~MSF CSFFGGEVFQ NHFEPVAVC AKCGYEFSS 33
Mouse SelR ~~~~~MSF CSFFGGEVFQ NHFEPVAVC AKCSYELFSS 33
Human SelR-c1 KSEWQKLTLP EQFYVTREKQ TEPPTFGIYL NNKEAAMVHC VCCDSPFSS 100
Human SelR-c2 QEELRKRLLTQ LQYDVTQNAH TESAFEGEYTT HHKDEPILRC VVCGTPEFKS 53
M.pneumoniae DGELKRTLTK LQYDVTQNAH TEPPTYTNEYN RHYKELIYVD ITSGEPEFIS 58
E.coli AEELKKNLSE MQFYVTQNHG TEPPTFGRL LLNRKDSVHC LICDAPVFS 56
M.therm. DDEWREILDV EAFRVARKAG TEPPTFGKYH DLHDDIIRRC ICCTDPEFS 74
S.cerevisiae DVKWDALTP LQLMVLKDKA TERPNTGAYL HTNESVAVHC ANCDRPLVSS 89
A.thaliana DEWNRVLSV EQFRVLRKLG TDKRGKGEFT KKFEPVAVC AGCGTALVKS 67
C.elegans QTEWKSVLNP EVYRVARESG TETPHTGGFN DHFEPVAVC LCCGSELVNS 76

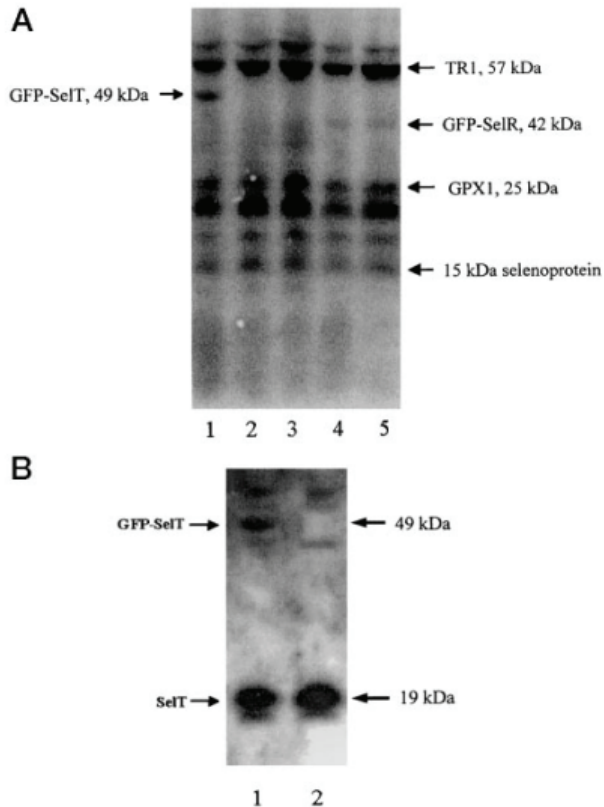
Human SelR RSKYAHSSPQ EATETIHAD SVAKRPE--- ----HNRS ALKVSCKGCG 75
Mouse SelR HSKYAHSSPQ EATETIHPD SVTKCPE--- ----KNRPE ALKVSCKGCG 75
Human SelR-c1 EKRYCSGTG ESESEAHGTS GSDESHTGIL RRLDTSLSGA RTEVVKQCE 150
Human SelR-c2 ETKFDSGCG ESEHDV---- --INSEAIT FTDDFSYGMH RVETSQSCQG 96
M.pneumoniae TPKFSGCG ESETKP---- --ISQDLIA NYRDESHGMI RTEVRAKNSN 101
E.coli QTRVDSGCG ESEYEP---- --VSEESIR YIKDLSHGMQ RIEIRCGNCD 99
M.therm. ETKFDSGCG ESEYDV---- --VSEHNK LREDRSLGMV RCEVLCARCD 117
S.cerevisiae KAKFDARCG ESEYE---- --EVSPGAI YHRDNLMPA RVEICCARCG 132
A.thaliana TTKFDSGCG EAFD----- --AI-PGAIK QTPEAG--GR RMEITCAVCD 107
C.elegans DANFWAGCG EASESVG-- --QDANIV RIVDRSHGMH RTEVRCKTCD 120

Human SelR NGLGHEE--- -LNDGPK-PG QSRBUIFSSS LKFVPKGEKA AASQGH---- 116
Mouse SelR NGLGHEE--- -LNDGPK-RG QSRBUIFSSS LKFVPKGEKA AASQGH---- 116
Human SelR-c1 AHGLGHEE--- -DGPG-PN QSRBUIFSSS LKFVPKGEKA AASQGH---- 182
Human SelR-c2 AHGLGHEE--- -DGPR-PT GKRVCINSAA LSFTP----- 125
M.pneumoniae SHLGHVSR--- -DGPEEHG GLRVCINSAA LKFPFAEME SAGYGEYLKL 146
E.coli AHGLGHEE--- -DGPR-PT GKRVCINSAA LKFPFAEME SAGYGEYLKL 137
M.therm. AHGLGHEE--- -DGPR-PT GKRVCINSAA LKFPFAEME SAGYGEYLKL 151
S.cerevisiae GHLGHVPEGE GWKQLNLKPK DTRVCINSAA LNLKDD----- 168
A.thaliana GHLGHVPEGE GY----STPT QSRBUIFSSS LKFASADSSK ----- 143
C.elegans AHGLGHEE--- -DGPKETT GKRVCINSVC MAFKDKD--- ----- 152

Human SelR ~~~~~
Mouse SelR ~~~~~
Human SelR-c1 ~~~~~
Human SelR-c2 ~~~~~
M.pneumoniae FKESN 151
E.coli ~~~~~
M.therm. ~~~~~
S.cerevisiae ~~~~~
A.thaliana ~~~~~
C.elegans ~~~~~

```





**Figure 5. Characterization of SelT and SelR.** *A*, detection of  $^{75}\text{Se}$ -labeled SelT and SelR. CV-1 cells were transfected with the plasmids encoding GFP-SelT (*lane 1*) or GFP-SelR (two independent transfection experiments are shown in *lanes 4* and *5*), or with a control plasmid (*lanes 2* and *3*). Transfected cells were grown in the presence of [ $^{75}\text{Se}$ ]selenite and  $^{75}\text{Se}$ -labeled proteins were resolved by SDS-PAGE and visualized with a PhosphorImager. Location and molecular weights of major selenoproteins thioredoxin reductase 1 (*TR1*), glutathione peroxidase 1 (*GPXI*), and the 15-kDa selenoprotein are indicated. *B*, immunoblot detection of SelT. Samples shown in *A* (*lanes 1* and *2*) were probed with anti-SelT antibodies. *Lane 1*, cells transfected with the GFP-SelT plasmid; *lane 2*, cells transfected with a control plasmid.

**Table II. Incidence of human SelT (in A) and SelR (in B) gene expression**

Incidence of human SelT and SelR gene expression is based on the occurrence of three or more independent cDNA clones in a particular dbEST library. In addition, EST cDNA clones with the incidence of one or two clones per dbEST library were found in libraries representing pineal gland, testis tumor, prostate, eye, heart, fetal liver spleen, white blood cells, bone, uterus, and endothelial cells (for SelT), and parathyroid tumor, spleen, brain, endothelial cells, heart, lung, testis, fibroblasts, pancreas tumor, retina, aorta, and gall bladder (for SelR).

Library	Incidence per 10,000 ESTs <i>numbers/library size</i>
<b>A</b>	
Infant brain (1NIB)	1.5 (6/40704)
Melanocyte (2NbHM)	1.6 (3/18816)
Placenta (Nb2HP)	1.4 (3/21888)
<b>B</b>	
Colon (2 libraries)	4.2 (8/19200)
Pregnant uterus (NbHPU)	3.0 (8/26880)
Prostate (NCI CGAP Pr22)	2.7 (3/11136)
Total fetus (Nb2HF8_9w)	2.1 (7/33024)
Placenta (2 libraries)	1.8 (7/39552)
Melanocyte (2NbHM)	1.6 (3/18816)
Kidney (2 libraries)	1.4 (6/41856)
Fetal liver spleen (2 libraries)	1.2 (12/98400)

selenoproteins (34). The cells transfected with the *SelT* fusion construct exhibited an additional band at ~49 kDa (Figure 5*A*, *lane 1*), and the cells transfected with the SelR fusion construct exhibited the band at ~42 kDa (Figure 5*A*, *lanes 4* and *5*), in complete agreement with expected masses for fusion proteins. These data established the expression of mammalian SelT and SelR, the presence of selenium in the proteins and the presence of functional SECIS elements in the genes for new selenoproteins.

**Immunoblot Detection of SelT** — Rabbit polyclonal antibodies were developed against the synthetic peptide corresponding to the C-terminal portion of SelT. These antibodies specifically recognized selenoprotein in CV-1 cells (Figure 5*B*, *lane 2*). The antibodies were also sufficient to detect the GFP-SelT fusion protein (Figure 5*B*, *lane 1*). Since antibodies to SelT recognized the protein region encoded by gene sequences located downstream of UGA, the readthrough of this codon must have occurred, which is consistent with UGA decoded as Sec. Overall, the immunoblot detection of SelT provided additional evidence for the presence of this selenoprotein in mammalian cells.

## DISCUSSION

We described herein identification of two new Sec-containing proteins through the algorithm that searches nucleotide sequence data bases for mammalian SECIS elements. Fourteen mammalian selenoproteins that were previously characterized had distinct functions and lacked a common amino acid consensus sequence. Therefore, the standard computation biology tools, such as BLAST or FASTA, that are based on the homology/consensus searches in protein and nucleotide sequences, could not be applied for identification of new selenoproteins that do not belong to one of the previously characterized subfamilies of selenoproteins. However, the 3'-UTR regions in selenoprotein mRNAs contained a common cis-acting sequence, SECIS element, that is necessary for Sec incorporation (18, 19). Analyses of previously identified SECIS elements revealed only several conserved nucleotides, but these were located in the conserved positions within the mRNA stem-loop structure (Figure 1, *A-D*) (20–23). We therefore developed a computer program, SECISearch, that recognized previously characterized as well as new SECIS elements in nucleotide data bases.

The SECIS element consensus sequence, ATGAN/11–12 nucleotides/AA/18–27 nucleotides/GA, was used in the first step of SECISearch to reduce the number of sequences for future analyses. The second step involved the search for the SECIS element secondary structure that is composed of Helixes I and II, and the internal and apical loops (Figure 1). However, the SECIS elements in known selenoprotein genes often contain certain “imperfections,” such as bulges, mismatches, and GU/UG base pairing within helices that complicate the description of the SECIS element secondary structure consensus. Such imperfections should be included in the algorithm for accurate description of the SECIS element, but they significantly decrease the specificity of the search criteria. This may result in an overwhelming number of false positives that contain multiple imperfections and that are unlikely to be functional SECIS elements. Experimental testing of these sequences would require enormous effort that is not justifiable.

To avoid this problem, an additional step, the calculation of the minimum free energy of mRNA secondary structure, was incorporated into SECISearch. Free energies of Helix I plus internal loop and Helix II plus apical loop were determined for SECIS elements present in known selenoprotein genes, and the free energy cut-off parameters were established that allowed a great reduction in the number of pseudo-SECIS elements, while keeping the option for imperfections. Free energies were evaluated separately for Helix I plus internal loop

and for Helix II plus apical loop rather than for the entire structure because the SECIS core feature, Quartet, involves non-Watson-Crick base pair interactions that could not be described by the available software. Separate free energy calculations for the upper and lower portions of SECIS element provided two independent parameters that increased the selectivity of SECISearch analyses.

Two versions of SECISearch, the data base search version and the online version, used somewhat different parameters and were designed for different purposes. The former may be used to search large nucleotide sequence data bases for a smaller number of sequences containing SECIS elements that are most similar to the descriptor SECIS element. It was designed to minimize the probability of finding pseudo-SECISes. The latter is best suited for testing a query nucleotide sequence for the presence of a potential SECIS element and was designed to minimize the probability of missing the actual SECIS element.

Analyses of the sequences selected by SECISearch from dbEST revealed a high selectivity of this program for SECIS elements as evidenced by the large proportion of SECIS elements from known selenoprotein genes (Table I). On the other hand, pseudo-SECISes constituted a significant proportion of the SECISearch selection, likely due to a low degree of structure and sequence conservation observed in SECIS elements and, perhaps, due to incomplete understanding of the SECIS element structure and other features of selenoprotein genes.

To aid in recognition of SECIS elements in the set of *SECISearch*-selected EST sequences, we used a set of criteria, designated as MSGS,<sup>2</sup> that describes the common features found in selenoprotein nucleotide and amino acid sequences: 1) conservation of Sec and Sec-flanking areas; 2) conservation of SECIS element in the 3'-UTR; and 3) the presence of homologous protein sequences that contain Cys in place of Sec (or the presence of distinct homologs in which Sec is conserved).<sup>2</sup> Although MSGS increased the efficiency of searches, a number of the SECISearch-selected sequences could not be tested by this approach because of the lack of homologous sequences available in the data bases. This resulted in difficulties in estimating the number of false positives. Further analyses will be required to assign these sequences as true or false SECIS elements. We hope that the accumulation of nucleotide sequences in dbEST and NR as well as further adjustments in the SECISearch algorithm will result in detection of new selenoprotein sequences in the SECISearch-selected sequences. In the present searches of SECIS elements, two unique nucleotide sequences found by SECISearch satisfied all MSGS criteria. These were experimentally verified and found to encode selenoproteins.

Nucleotide sequences for two new selenoproteins, SelT and SelR, were represented by a number of ESTs in the human dbEST (Table II) and these new selenoproteins had numerous homologous sequences in lower eukaryotes. SelT and SelR did not have homology to previously characterized proteins, but had all the features that established them as selenoproteins; *i.e.* their nucleotide sequences contained conserved in-frame UGA codons in the coding region and conserved SECIS elements in the 3'-UTRs. In addition, several non-mammalian organisms contained sequences that were homologous to SelT and SelR genes and contained a cysteine codon in place of UGA, the feature characteristic of mammalian selenoproteins. This feature may provide, in the future, a basis for the alternative approach to search for selenoprotein genes.

Experimental evidence for the presence of SelT and SelR in mammals included expression of selenoproteins in a monkey CV-1 cell line as fusion proteins with GFP and detection of selenium in the expressed proteins by metabolic labeling of the proteins with <sup>75</sup>Se. Im-

munoblot detection of SelT confirmed the occurrence of this selenoprotein in mammalian cells. Additional evidence for the presence of Sec in SelT is the selenium-dependent regulation of SelT biosynthesis. Selenium is often a limiting factor in selenoprotein synthesis, and selenium supplementation results in elevation of selenoprotein levels in mammalian tissues and cell cultures (29). Using immunoblot analyses, we recently found that the addition of selenium to a mammalian cell culture results in increased expression of SelT (G. V. Kryukov and V. N. Gladyshev, unpublished data).

An interesting feature observed for SelR was the presence of its homologs encoded in all genomes for which the complete nucleotide sequence is available. This suggests that the SelR homologs are present in all living organisms. Although the function of SelR is not known, the presence of a SelR homolog in a minimal gene set for cellular life suggests the importance of SelR for one of the basic processes in cellular metabolism. It should be noted that disruption of the mouse Sec-tRNA gene resulted in early embryonic lethality (28), and this effect is likely due to a null expression of one or more selenoproteins. Further gene knockout studies may be necessary to determine if SelR is one of such selenoproteins essential for development. Two Cys-containing homologs of SelR (Figure 4B) were also detected in mammals and it is not known if SelR deficiency may be compensated for in mammalian cells.

In conclusion, we describe a method for identifying new selenoprotein genes in nucleotide data bases. These genes could be found with the help of a computer program, SECISearch, that recognized SECIS elements in selenoprotein genes. We applied this program to search dbEST and identified two new selenoprotein genes with no homology to known proteins. In addition, we provided experimental evidence for the natural occurrence of these proteins. These data demonstrate the applicability of identification of selenoprotein genes through recognition of SECIS elements, and we suggest that this new method will be useful for identification of selenoprotein sequences in current and future large-scale sequencing projects. Identification of new selenoproteins may help in explaining many biological effects of selenium.

**Acknowledgements** — We are grateful to Dolph Hatfield, Ruma Banerjee, Audrey Aikin, Boris Trubitsin, and the anonymous reviewer for helpful suggestions. This work was supported by a NATO Linkage grant and a Layman Award (to V. N. G.).

## REFERENCES

1. Hatfield, D. L., Gladyshev, V. N., Park, S. I., Chittum, H. S., Carlson, B. A., Moustafa, M. E., Park, J. M., Huh, J. R., Kim, M., and Lee, B. J. (1999) *Comp. Nat. Prod. Chem.* **4**, 353–380
2. Stadtman, T. C. (1996) *Annu. Rev. Biochem.* **65**, 83–100
3. Bock, A., Forchhammer, K., Heider, J., and Baron, C. (1991) *Trends Biochem. Sci.* **16**, 463–467
4. Sawers, G., Heider, J., Zehelein, E., and Bock, A. (1991) *J. Bacteriol.* **173**, 4983–4993
5. Wilting, R., Schorling, S., Persson, B. C., and Bock, A. (1997) *J. Mol. Biol.* **266**, 637–641
6. Gladyshev, V. N., Krause, M., Xu, X. M., Korotkov, K. V., Kryukov, G. V., Sun, Q.-A., Lee, B. J., Wootton, J. C., and Hatfield, D. L. (1999) *Biochem. Biophys. Res. Commun.* **259**, 244–249
7. Buettner, C., Harney, J. W., and Berry, M. J. (1999) *J. Biol. Chem.* **274**, 21598–21602

8. Gladyshev, V. N., and Hatfield, D. L. (1999) *J. Biomed. Sci.* **6**, 151–160
9. Sunde, R. A. (1994) in *Selenium in Biology and Human Health* (Burk, R. F., ed), pp. 146–177, Springer-Verlag New York Inc., New York
10. Ursini, F., Maiorino, M., and Roveri, A. (1997) *Biomed. Environ. Sci.* **10**, 327–332
11. Larsen, P. R. (1997) *Biochem. Soc. Trans.* **25**, 588–592
12. St. Germain, D. L., and Galton, V. A. (1997) *Thyroid* **7**, 655–68
13. Sun, Q.-A., Wu, Y., Zappacosta, F., Jeang, K.-T., Lee, B. J., Hatfield, D. L., and Gladyshev, V. N. (1999) *J. Biol. Chem.* **274**, 24522–24530
14. Guimaraes, M. J., Peterson, D., Vicari, A., Cocks, B. G., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Ferrick, D. A., Kastelein, R. A., Bazan, J. F., and Zlotnik, A. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 15086–15091
15. Vendeland, S. C., Beilstein, M. A., Chen, C. L., Jensen, O. N., Barofsky, E., and Whanger, P. D. (1993) *J. Biol. Chem.* **268**, 17103–17107
16. Burk, R. F., and Hill, K. E. (1999) *Bioessays* **21**, 231–237
17. Gladyshev, V. N., Jeang, K. T., Wootton, J. C., and Hatfield, D. L. (1998) *J. Biol. Chem.* **273**, 8910–8915
18. Berry, M. J., Banu, L., Chen, Y. Y., Mandel, S. J., Kieffer, J. D., Harney, J. W., and Larsen, P. R. (1991) *Nature* **353**, 273–276
19. Low, S. C., and Berry, M. J. (1996) *Trends Biochem. Sci.* **21**, 203–208
20. Walczak, R., Westhof, E., Carbon, P., and Krol, A. (1996) *RNA* **2**, 367–379
21. Walczak, R., Carbon, P., and Krol, A. (1998) *RNA* **4**, 74–84
22. Grundner-Culemann, E., Martin, G. W., 3rd, Harney, J. W., and Berry, M. J. (1999) *RNA* **5**, 625–635
23. Martin, G. W., 3rd, Harney, J. W., and Berry, M. J. (1996) *RNA* **2**, 171–182
24. Burk, R. F., and Hill, K. E. (1993) *Annu. Rev. Nutr.* **13**, 65–81
25. Baum, M. K., and Shor-Posner, G. (1998) *Nutr. Rev.* **56**, S135–139
26. Clark, L. C., Combs, G. F., Jr., Turnbull, B. W., Slate, E. H., Chalker, D. K., Chow, J., Davis, L. S., Glover, R. A., Graham, G. F., Gross, E. G., Krongrad, A., Leshner, J. L., Jr., Park, H. K., Sanders, B. B., Jr., Smith, C. L., and Taylor, J. R. (1996) *J. Am. Med. Assoc.* **276**, 1957–1963
27. Lee, B. J., Park, S. I., Park, J. M., Chittum, H. S., and Hatfield, D. L. (1996) *Mol. Cells* **6**, 509
28. Bosl, M. R., Takaku, K., Oshima, M., Nishimura, S., and Taketo, M. M. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 5531–5534
29. Hill, K. E., Chittum, H. S., Lyons, P. R., Boeglin, M. E., and Burk, R. F. (1996) *Biochim. Biophys. Acta* **1313**, 29–34
30. Mushegian, A. R., and Koonin, E. V. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10268–10273
31. Zuker, M., and Stiegler, P. (1981) *Nucleic Acids Res.* **9**, 133–148
32. Wen, W., Weiss, S. L., and Sunde, R. A. (1998) *J. Biol. Chem.* **273**, 28533–28541
33. Siedler, F., Rudolph-Bohner, S., Doi, M., Musiol, H. J., and Moroder, L. (1993) *Biochemistry* **32**, 7488–7495
34. Gladyshev, V. N., Stadtman, T. C., Hatfield, D. L., and Jeang, K. T. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 835–839