THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 274, No. 53, Issue of December 31, pp. 38147–38154, 1999
Printed in U.S.A.

Novel Selenoproteins Identified *in Silico* and *in Vivo* by Using a Conserved RNA Structural Motif*

(Received for publication, September 23, 1999)

Alain Lescure‡, Daniel Gautheret§, Philippe Carbon‡, and Alain Krol‡¶

From ‡Unité Propre de Recherche CNRS 9002, Structure des Macromolécules Biologiques et Mécanismes de Reconnaissance, Institut de Biologie Moleculaire et Cellulaire, 15, Rue René Descartes, 67084 Strasbourg Cedex, France and §Unité Mixte de Recherche CNRS 1889, Information Génétique et Structurale, 31, Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France

Selenocysteine is incorporated into selenoproteins by an in-frame UGA codon whose readthrough requires the selenocysteine insertion sequence (SECIS), a conserved hairpin in the 3′-untranslated region of eukaryotic selenoprotein mRNAs. To identify new selenoproteins, we developed a strategy that obviates the need for prior amino acid sequence information. A computational screen was used to scan nucleotide sequence data bases for sequences presenting a potential SECIS secondary structure. The computer-selected hairpins were then assayed *in vivo* for their functional capacities, and the cDNAs corresponding to the SECIS winners were identified. Four of them encoded novel selenoproteins as confirmed by *in vivo* experiments. Among these, SelZf1 and SelZf2 share a common domain with mitochondrial thioredoxin reductase-2. The three proteins, however, possess distinct N-terminal domains. We found that another protein, SelX, displays sequence similarity to a protein involved in bacterial pilus formation. For the first time, four novel selenoproteins were discovered based on a computational screen for the RNA hairpin directing selenocysteine incorporation.

Selenium is an essential trace element whose deficiency can interfere with normal embryonic development and fertility or favor the appearance of certain cancers and viral diseases such as human immunodeficiency virus and coxsackievirus (1). The amino acid selenocysteine is the major biological form of selenium in bacteria and animals. It is found in the active site of selenoproteins and is directly involved in the catalytic reaction. In this regard, the capacity of the selenocysteine selenol group to become ionized at physiological pH, the cysteine thiol group requiring a higher pH, accounts for the higher rate of catalysis of selenoenzymes (2). Seven selenoprotein families have been characterized so far in mammals (3): the glutathione peroxidase and thioredoxin reductase families, involved in scavenging reactive oxygen species and maintaining the redox status of the cell; three iodothyronine deiodinases participating in the thyroid hormone metabolism; and last, SelW and SelP, which have not been attributed a function yet. More recently, a 15-kDa selenoprotein of unknown function has been purified (4). Selenophosphate synthetase-2, the seventh selenoprotein, is remarkable in that it contains selenocysteine, but is also a key actor in the biosynthesis of this amino acid (5).

Selenocysteine is encoded by an in-frame UGA codon, implying the existence of a mechanism capable of distinguishing the UGA selenocysteine codon from a translational stop. This process requires, in eukaryotes, the presence of the selenocysteine insertion sequence (SECIS),[1] a hairpin residing in the 3′-untranslated region of selenoprotein mRNAs that is essential for readthrough of the UGA selenocysteine codon (6). Sequence comparisons and structure-function experiments generated a consensus secondary structure model for the SECIS element in which a functional motif could be identified (7, 8).

Compelling evidence for the existence of molecular links between selenium deficiencies and biological disorders came from molecular genetics experiments. Targeted disruption of the mouse selenocysteine tRNA gene led to early embryonic lethality, implying that selenoprotein synthesis is essential to mammals (9). Studies carried out on knockout mice lacking the glutathione peroxidase underlined the protective role of selenium against free radicals (10) or coxsackievirus-induced myocarditis in Keshan disease (1). Further supporting the biological importance of this trace element, selenium labeling experiments in rats determined the existence of more selenoproteins to be identified and characterized (11). To undertake this task, we intended here to exploit the mine of information stored in EST data bases. The central question in such a project is how the relevant cDNAs can be retrieved without the knowledge of even a partial protein sequence. To circumvent the obstacle, a strategy was developed based on the absolute requirement of a SECIS element for selenoprotein translation. The finding of such a hairpin in a cDNA should therefore signal the presence of an attached coding sequence. Two assets were exploited to extract new SECIS elements from EST data bases. The first one was the detailed knowledge of the secondary structure of the SECIS element, which is conserved in all known selenoprotein mRNAs. The second one was the utilization of a program capable of detecting potential RNA secondary structures in nucleotide sequence data bases. Combined with molecular biology and *in vivo* experiments, this approach led to the discovery of four novel selenoproteins using a single RNA element as a structural tag.

* This work was supported by grants from the Association pour la Recherche sur le Cancer, the Ligue Nationale contre le Cancer, the Ligue contre le Cancer du Haut-Rhin, and the Centre Volvic pour la Recherche sur les Oligo-Eléments. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF166124, AF166125, AF166126, and AF166127.

¶ To whom correspondence should be addressed. Tel.: 33-3-88-41-70-50; Fax: 33-3-88-60-22-18; E-mail: krol@ibmc.u-strasbg.fr.

[1] The abbreviations used are: SECIS, selenocysteine insertion sequence; EST, expressed sequence tag; ORF, open reading frame; PCR, polymerase chain reaction; bp, base pair(s); HA, hemagglutinin; RACE, rapid amplification of cDNA ends; GPx, glutathione peroxidase; UTR, untranslated region; TrxR2, mammalian mitochondrial thioredoxin reductase-2; contig, group of overlapping clones.

EXPERIMENTAL PROCEDURES

Computational Screen and Sequence Comparisons—The search for new SECIS elements was conducted in GenBank™, sequence-tagged site, and EST data bases with the RNAMOT pattern search program (12, 13) with the descriptor shown in Fig. 1*A*. 600,300 3′- and 5′-ESTs were scanned, representing a total of ~222 × 10^6 nucleotides.[2] Positive hits were aligned with ClustalW (14). The same descriptor run against a randomized sequence of 10^7 nucleotides (A, T, G, and C frequencies, 25% each) yielded three hits. ORFs and ESTs were identified by BLAST searches (15) in the GenBank™ and EST data bases and aligned with ClustalW.

Cloning of the New SECIS Elements—The new SECIS elements were obtained by standard PCR amplification of a human B cell library or of human or mouse genomic DNAs (gifts of S. Elledge, J. L. Mandel, and F. Guillemot, respectively) with oligonucleotides GGGTGATCAGGGGT$(N)_{24}$ and CGGGGTACCTGGAT$(N)_{24}$ as the 5′- and 3′-primers, respectively. $(N)_{24}$ corresponds to 24 nucleotides complementary to the SECIS sequence, including the top 4 base pairs of helix I (see Fig. 1*A*). SECIS AA109465 was constructed by nested PCR. The PCR primers introduced a *Bcl*I site at the 5′-end and a *Kpn*I at the 3′-end of the SECIS elements in addition to a 4-bp stem below helix I (see Fig. 1*B*). To replace the naturally occurring SECIS element in the glutathione peroxidase reporter, the SECIS candidates were introduced in pGHA-BcK at the *Bcl*I-*Kpn*I sites (8). This plasmid encodes a triple-HA tag fused in-frame to the N terminus of the glutathione peroxidase coding sequence (8).

Identification and Cloning of the cDNAs Encoding the Novel Selenoproteins—ESTs corresponding to the functional SECIS elements were identified by querying EST data bases with BLASTN at NCBI. Sequences were aligned with the CAP program (16), producing a contiguous sequence. GenBank™ accession numbers AA180412, AA057045, H44779, and R44842, corresponding to the longest cDNA clones identified for SelN, SelX, SelY, and SelZ, respectively, were purchased from Genome Systems. Longer cDNAs, AF007144 for SelY and R47273 for SelN, were kindly provided by W. Yu and M. M. Y. Waye.

A 1333-bp cDNA fragment corresponding to SelX was identified by screening a HeLa oligo(dT) library (a gift of P. Chambon) with a probe spanning positions 1–197 of AA057045. This *Eco*RI-*Xho*I fragment in pBluescript KS was called pSelX. For SelN, the sequence alignment showed that cDNA R47273 overlapped the 5′-most 578 bp of AA180412. R47273 and AA180412 were entirely sequenced and fused by ligation of the 702-bp *Xba*I fragment of R47273 to *Xba*I-digested AA180412, yielding pSelN2, a 2742-bp *Eco*RI-*Xho*I fragment in pBluescript SK. Another fragment of 2066 bp, overlapping the 1544 bp 5′ to pSelN2, was obtained by screening a HeLa random-primed library (a gift of P. Chambon) with a probe complementary to positions 1–702 of the R47273 *Xba*I fragment, giving rise to plasmid pSelN3. The 1543-bp *Xba*I fragment of pSelN3 was inserted into the *Xba*I-digested plasmid containing AA180412, generating pSelN4. Additional 5′-sequences of pSelN4 were obtained by 5′-Marathon RACE using the human prostate Marathon-Ready cDNA and the Advantage cDNA PCR kit (CLONTECH). The PCR fragment obtained was digested by *Not*I-*Ehe*I, and the resulting 999-bp fragment was ligated to the *Not*I-*Ehe*I-digested pSelN4 plasmid, yielding pSelN, a 3955-bp *Not*I-*Xho*I fragment in pBluescript SK. The cDNA R44842 containing the 1505-bp *Hin*dIII-*Not*I fragment in pLafmid BA was entirely sequenced and named pSelZ. Similarly to SelN, additional 5′-sequences were obtained by 5′-Marathon RACE, giving rise to the 1170-bp (M15) and 1150-bp (M19) PCR fragments, different in sequence. Into the blunt-ended *Hin*dIII-*Sma*I-digested pSelZ plasmid was inserted either the 1121-bp *Sma*I fragment from M19 or the 1141-bp *Sma*I fragment from M15 to generate pSelZf1 (2021 bp) and pSelZf2 (2041 bp) cDNAs, respectively.

cDNA Constructs for in Vivo Expression of SelX, SelN, and SelZ—The cDNAs coding for the different proteins, either with or lacking the SECIS elements, were inserted into the eukaryotic expression vector pXJ41 (a gift of P. Chambon) under the transcriptional control of the cytomegalovirus promoter. A triple-HA tag was fused in-frame to the N termini of SelX, SelN, and SelZ by incorporating, by site-directed mutagenesis, *Pst*I or *Hin*dIII sites into pGHA-BcK, downstream of the HA tag sequence, with oligonucleotide GCTCAGTGCGGCCGCTCGTTCTGCAGTCTGCTGCTCGGCTC or GCTCAGTGCGGCCGCGAAGCTTCTGCTGCTCGGCTC (restriction sites underlined), generating constructs pGHA-BcK+*Pst*I and pGHA-BcK+*Hin*dIII, respectively. *Pst*I-*Stu*I digestions of pSelX generated the 1040-bp *Pst*I-*Stu*I fragment that was ligated to the blunt-ended *Pst*I-*Bgl*II-digested pGHA-BcK+*Pst*I plasmid to produce pHASelX. Ligation of the 739-bp blunt-ended *Pst*I-*Hin*dIII fragment from pSelX to the blunt-ended *Pst*I-*Bgl*II-digested pGHA-BcK+*Pst*I plasmid generated pHASelXΔSECIS. The 139-bp *Eco*RI-*Hin*dIII fragment from pGHA-BcK+*Hin*dIII was ligated to *Eco*RI-*Hin*dIII-digested pXJ41, resulting in construct $pXJ(HA)_3$. A *Hin*dIII restriction site was introduced into pSelN by site-directed mutagenesis with oligonucleotide CGGCCGCCCGGGCAAGCTTACATCAGCCC (*Hin*dIII site underlined), with the last T of the site corresponding to the first base of the first codon identified in SelN (position 2 in SelN), yielding pSelN+*Hin*dIII. *Hin*dIII-*Kpn*I digestion of pSelN+*Hin*dIII generated a 3973-bp fragment that was inserted into *Hin*dIII-*Kpn*I-cleaved $pXJ(HA)_3$ to generate pHASelN. The 2314-bp blunt-ended *Hin*dIII-*Nhe*I fragment from pSelN+*Hin*dIII was inserted into the blunt-ended *Hin*dIII-*Kpn*I-cleaved $pXJ(HA)_3$ vector to generate pHASelNΔSECIS. A *Bam*HI site was introduced by site-directed mutagenesis into pSelZ with oligonucleotide GGCCTGCAGGGATCCCGCTTACCCTC or GCGGCCGCAGGAATGGATCCTCTTTATTTGCATTGC (*Bam*HI sites underlined) at either position 1137 (3′ adjacent to the TAA stop codon) or 1469 (13 bp upstream of the poly(A) tail), respectively. This gave rise to constructs pSelZ-Bamsh and pSelZ-Bamlg, respectively. The 1140- and 1471-bp *Hin*dIII-*Bam*HI fragments, arising from *Hin*dIII-*Bam*HI digestions of pSelZ-Bamsh and pSelZ-Bamlg, were subcloned into the *Hin*dIII-*Bgl*II-digested $pXJ(HA)_3$ vector, giving rise to pHASelZ and pHASelZΔSECIS, respectively.

Transfection of COS-7 Cells, ^{75}Se Labeling, and Glutathione Peroxidase Assays—COS-7 cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, 2 mM L-glutamine, and 0.1 mg/ml gentamycin according to standard cell culture procedures. Transient transfections were carried out by calcium phosphate precipitation as described (8), with 5 μg of test DNA, 4 μg of selenocysteine tRNA expression vector, and 1 μg of plasmid LacZ-cytomegalovirus as the transfection standard. Sodium selenite (10 nM) was added to the culture medium. Cells were washed after 16 h and harvested 24 h later by scraping. Lysis was carried out by the freeze-thaw procedure in 50 μl of 100 mM Tris-HCl (pH 8). For protein analysis, the lysis buffer was adjusted to 20 mM HEPES-NaOH (pH 7.9), 12.5 mM $MgCl_2$, 150 mM KCl, 0.1 mM EDTA, 10% glycerol, and 0.5% Tween and further incubated on ice for 20 min. The crude cell extract was then centrifuged at 4 °C for 3 min at 13,000 × *g* to remove cell debris. The supernatant was used for subsequent analysis. For ^{75}Se labeling, 6 μCi of $Na_2{}^{75}SeO_3$ (2.5 μCi/μg selenium; University of Missouri Research Reactor) were added to each 100-mm plate 24 h after transfection of the plasmids. Cells were further incubated for 20 h before harvesting. Lysis was as described above.

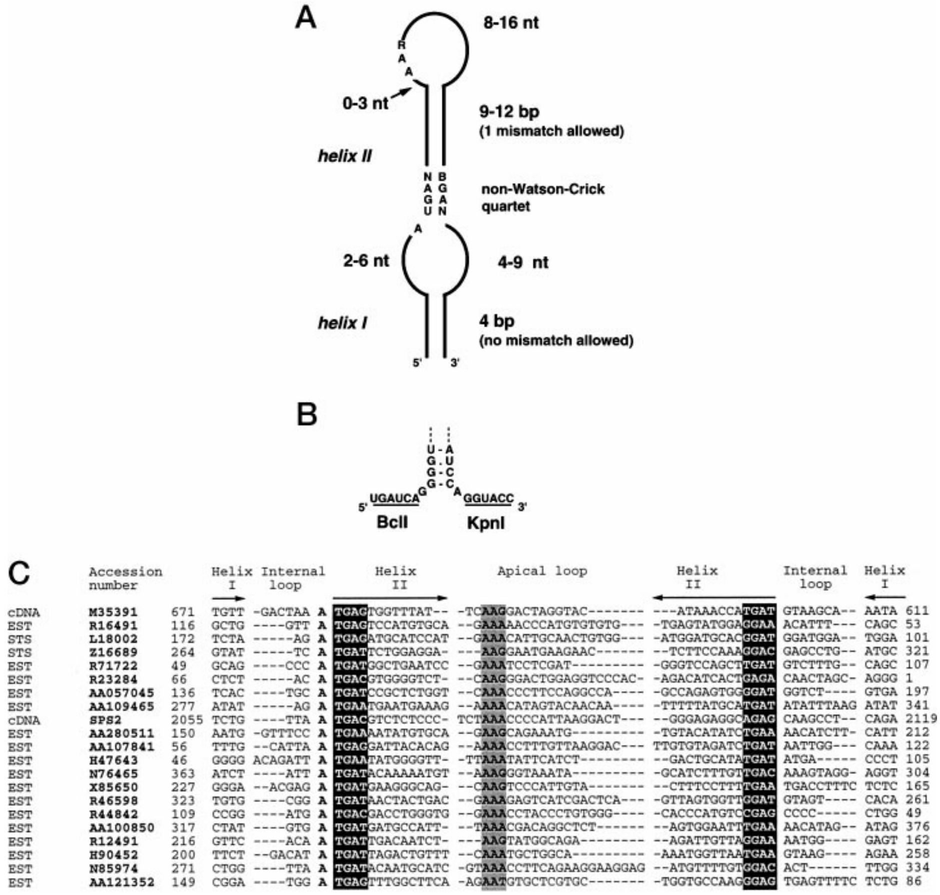
Western blot analysis, normalized to β-galactosidase activities, was performed as described (8). For the glutathione peroxidase (GPx) activity assays (8), the HA tag was removed by *Not*I digestion followed by self-ligation. Prior to GPx activity measurements, β-galactosidase activities were assayed with 5 μl of crude cell extract to normalize the results. Assays were performed in triplicate.

Immunoprecipitations—The HA-tagged proteins were immunoprecipitated by incubating 25 μl of lysis supernatant with 30 μl of anti-HA antibody 12CA5 linked to protein A-Sepharose beads in a total volume of 250 μl of lysis buffer for 1 h at room temperature. The beads were spun down, washed four times in 200 μl of lysis buffer for 15 min, mixed with 20 μl of loading buffer (100 mM Tris-HCl (pH 6.8), 150 mM dithioerythritol, 4% SDS, 20% glycerol, and 0.2% bromphenol blue), heated in boiling water for 3 min, and centrifuged.

RESULTS

A Computational Screen for New SECIS Elements—To scan for sequences that could adopt secondary structures similar to the SECIS element, we developed a computational screen based on the pattern search program RNAMOT (12, 13). An input primary/secondary structure descriptor (Fig. 1*A*) for RNAMOT was inferred from sequence comparisons and the SECIS consensus structure experimentally determined at the time of the search (7, 8). To test the validity of the descriptor, RNAMOT was run against the GenBank™ non-redundant data base (10^9 nucleotides at the time), generating 34 different SECIS elements belonging to the then known selenoprotein mRNAs. An additional hit (M35391 in Fig. 1*C*) was found in an intron of the human procollagen α2 chain gene. Given its localization, it is not likely to represent a *bona fide* SECIS element. However, it was retained because it contained all the

[2] May 1997 release.



```
     Accession         Helix Internal      Helix           Apical loop             Helix      Internal  Helix
     number              I     loop          II                                      II         loop      I
cDNA M35391     671  TGTT -GACTAA A TGAGTGGTTTAT-- -TCAAGGACTAGGTAC------- ---ATAAACCATGAT GTAAGCA-- AATA 611
EST  R16491     116  GCTG ----GTT A TGAGTCCATGTGCA --GAAAAACCCATGTGTGTG--- -TGAGTATGGAGGAA ACATTT--- CAGC 53
STS  L18002     172  TCTA -----AG A TGAGATGCATCCAT --GAAACATTGCAACTGTGG--- -ATGGATGCACGGAT GGATGGA-- TGGA 101
STS  Z16689     264  GTAT -----TC A TGATTCTGGAGGA- ---AAGGAATGAAGAAC------ --TCTTCCAAAGGAC GAGCCTG-- ATGC 321
EST  R71722     49   GCAG ----CCC A TGATGGCTGAATCC --GAAATCCTCGAT--------- -GGGTCCAGCTTGAT GTCTTTG-- CAGC 107
EST  R23284     66   CTCT -----AC A TGACGTGGGGTCT- --CAAGGGACTGGAGGTCCCAC- -AGACATCACTGAGA CAACTAGC- AGGG 1
EST  AA057045   136  TCAC ----TGC A TGATCCGCTCTGGT --CAAACCCTTCCAGGCCA---- -GCCAGAGTGGGGAT GGTCT---- GTGA 197
EST  AA109465   277  ATAT -----AG A TGAATGAATGAAAG --AAAACATAGTACAACAA---- -TTTTTATGCATGAT ATATTTAAG ATAT 341
cDNA SPS2       2055 TCTG ----TTA A TGACGTCTCTCCC- TCTAAACCCCATTAAGGACT--- --GGGAGAGGCAGAG CAAGCCT-- CAGA 2119
EST  AA280511   150  AATG -GTTTCC A TGAAAATATGTGCA --GAAGCAGAAATG--------- -TGTACATATCTGAA AACATCTT- CATT 212
EST  AA107841   56   TTTG --CATTA A TGAGGATTACACAG --AAAACCTTTGTTAAGGAC--- TTGTGTAGATCTGAT AATTGG--- CAAA 122
EST  H47643     46   GGGG ACAGATT A TGAATATGGGGTT- -TTAAATATTCATCT-------- --GACTGCATATGAT ATGA----- CCCT 105
EST  N76465     363  ATCT ----ATT A TGATACAAAAATGT --AAAGGGTAAATA--------- -GCATCTTTGTTGAC AAAGTAGG- AGGT 304
EST  X85650     227  GGGA --ACGAG A TGATGAAGGGCAG- --CAAGTCCCATTGTA------- --CTTTCCTTTTGAA TGACCTTTC TCTC 165
EST  R46598     323  TGTG ----CGG A TGATAACTACTGAC --GAAAGAGTCATCGACTCA--- -GTTAGTGGTTGGAT GTAGT---- CACA 261
EST  R44842     109  CCGG ----ATG A TGACGACCTGGGTG --GAAACCTACCCTGTGGG---- -CACCCATGTCCGAG CCCC----- CTGG 49
EST  AA100850   317  CTAT ----GTG A TGATGATGCCATT- --TAAACGACAGGCTCT------ --AGTGGAATTTGAA AACATAG-- ATAG 376
EST  R12491     216  GTTC ----ACA A TGATTGACAATCT- ---AAGTATGGCAGA-------- --AGATTGTTAGGAA AATGG---- GAGT 162
EST  H90452     200  TTCT --GACAT A TGATTAGACTGTTT --CAAATGCTGGCA--------- -AAATGGTTAATGAA GTAAG---- AGAA 258
EST  N85974     271  CTGG ----TTA A TGATACAATGCATC -GTAAACCTTCAGAAGGAAGGAG --ATGTTTTGTGGAC ACT------ TTGG 334
EST  AA121352   149  CGGA ----TGG A TGAGTTTGGCTTCA -AGAATGTGCTCGTGC------- -TGGTGCCAAGGGAG TGAGTTTTC TCTG 86
```

FIG. 1. **The 21 SECIS candidates arising from the computer screen.** *A*, the SECIS descriptor used. Sequence (*R*, purine; *B*, any nucleotide except A; *N*, any nucleotide) and structure constraints were derived from the SECIS consensus secondary structure (7). Shown are the lengths allotted to single strands (nucleotides (*nt*)) and helices (bp). *B*, sequence of the 4-bp stem linked to all SECIS elements in our experiments. *C*, the 21 SECIS candidates, displayed as a structural alignment highlighting the SECIS features described for *A*. The invariant *A* is in *boldface*; conserved apical loop residues are *shaded*; and the sequences of the non-Watson-Crick quartet are in *black boxes*. The relevant accession numbers and the positions of the SECIS elements in the sequences are indicated. *STS*, sequence-tagged site.

features of the SECIS consensus structure. Also, a search with an alternative descriptor carrying N instead of B at the top base pair of the non-Watson-Crick quartet led to the discovery of a SECIS element in the 3′-UTR of the selenophosphate synthetase-2 cDNA. This cDNA was characterized earlier, but no SECIS element could be found by the authors (5).

In a second step, the search was conducted in the GenBank™ EST data base (222 × 10^6 nucleotides). After discarding ambiguous hits containing one or more undefined nucleotides, RNAMOT found 376 sequences, including 153 mouse, 101 human, 92 *Brugia malayi*, and 30 other animal and plant ESTs. A sequence alignment was performed with ClustalW (14), and we plotted the derived neighbor-joining tree to obtain a clustered representation of the matches. This identified 62 individual sequences that could be classified into three families. One family comprised sequences corresponding to the known SECIS elements; another contained groups of unknown SECIS; and the last one contained orphans represented by one or two ESTs only. RNA sequences carrying several AU or GC repeats, prone to adopt alternative secondary structures by changes in the base pairing register, were rejected because of their low biological significance. The remaining elements were assessed in terms of stabilities. Some sequences were discarded based on their low thermodynamic stabilities due to too many consecutive G·U base pairs, either 5′ to the non-Watson-Crick quartet or below the apical loop. Seventeen SECIS candidates were eventually obtained that met the requirements imposed by the different subscreens and presented the features of the SECIS consensus element. For easy correspondence with the EST sequences, the SECIS elements were called by the accession number of one of their parental ESTs (Fig. 1*C*). All the sequences belonged to human or mouse ESTs, except AA109465, which was a member of a family of 92 *B. malayi* ESTs. Running the program against the GenBank™ sequence-tagged site data base generated nine sequences. Only four of them, accession numbers L18002, Z16689, Z74617, and Z75892 (Fig. 1*C*), were successful in the subsequent screens. The last two corresponded to the R16491 and R23284 ESTs characterized in the GenBank™ EST data base search. After this first round of selection, 21 SECIS candidates were obtained, comprising 2 cDNAs, 17 ESTs, and 2 sequence-tagged sites.

Functional Assays of the Selected SECIS Candidates—The 21 SECIS candidates were then tested for *in vivo* function. The SECIS DNAs were obtained by PCR amplification of genomic DNA or cDNA libraries. Concomitant with the PCR amplification and due to the uneven stability of helix I in the different SECIS elements, an identical 4-bp stem was added below helix I in all SECIS elements (Fig. 1*B*) in order for the SECIS RNAs to exhibit similar stabilities. GPx being a selenoprotein, its translation requires a functional SECIS element in the 3′-UTR of its mRNA. The SECIS DNA candidates were then introduced separately into the 3′-UTR of a GPx cDNA reporter to replace the residing SECIS element. In this construct, the GPx coding sequence carries an HA tag fused in frame at the N terminus to allow detection of the translated proteins with the anti-HA antibody. Whether or not the SECIS candidates were active could be apprehended by a rapid assay involving COS-7 transfections of the constructs, followed by Western blotting experiments. A functional SECIS candidate should lead to translation of a full-length GPx. In contrast, with an inactive SECIS element, the UGA selenocysteine codon will be recognized as a stop codon, leading to translation of a shortened 9.5-kDa polypeptide. Translation of the mRNA coding for the HA-tagged GPx, carrying its own SECIS, generated a product of ~27 kDa (Fig. 2*A*, *lane 2*). Construct GPx-mutSECIS had the G·A/A·G to A·G/G·A substitution in the non-Watson-Crick quar-

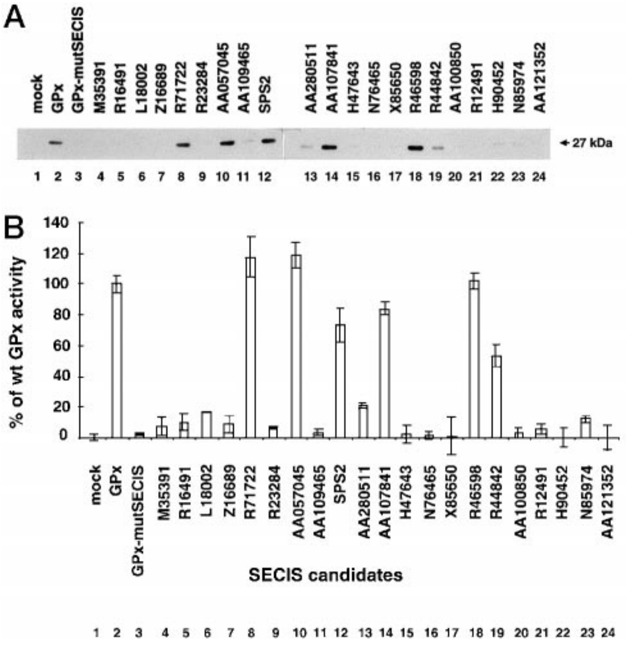


FIG. 2. ***In vivo* functional assays of the SECIS candidates.** *A*, capacities to direct readthrough of the glutathione peroxidase UGA selenocysteine codon. Each SECIS sequence (accession numbers from Fig. 1*C*) was introduced into the 3′-UTR of the HA-tagged GPx cDNA reporter, transfected into COS-7 cells. The lengths of the proteins were evaluated by Western blot analysis with the anti-HA antibody. *Lane 1*, mock-transfected COS-7 cells; *lane 2*, transfection of the wild-type GPx cDNA; *lane 3*, transfection of the GPx-mutSECIS construct harboring a debilitated SECIS element. The position of the wild-type GPx in *lane 2* is a size marker for the expected translation products. *Lanes 1–12* and *13–24* are from different gels. *B*, GPx activities arising from the transfected GPx cDNAs carrying the SECIS candidates. Average values (from three independent transfections carried out in triplicate), subtracted from the background level of the endogenous GPx, are given with respect to the transfected wild-type (*wt*) GPx taken as 100%. Controls (*bars 1–3*) are as described for *lanes 1–3* in *A*. *SPS2*, selenophosphate synthetase-2.

tet of the SECIS element that impaired its function (8), providing here also minute amounts of GPx (compare *lanes 2* and *3*). This construct provided the background level. Consistent with earlier observations (8), no 9.5-kDa protein appeared with GPx-mutSECIS, presumably due to the instability of such an unnatural short polypeptide *in vivo*. Fig. 2*A* shows that, among the 21 SECIS tested, only R71722, AA057045, selenophosphate synthetase-2, AA107841, R46598, and R44842 could mediate production of a full-length GPx with an efficiency comparable to that of the authentic GPx SECIS (compare with *lanes 2* and *3*). A seventh element, AA280511 (*lane 13*), also produced full-length GPx, but with a lower efficiency.

Since the active site of GPx contains an essential selenocysteine, measuring the enzymatic activity will attest that this amino acid was effectively incorporated into the protein. After transfection into COS-7 cells of the cDNA constructs carrying the SECIS candidates, GPx activities were assayed from crude cell extracts and compared with that of wild-type GPx (Fig. 2*B*, *bar 2*). As anticipated, no significant activity emanated from GPx-mutSECIS (*bar 3*). Wild-type or slightly higher than wild-type activities were observed with R71722 (105%), AA057045 (110%), and R46598 (~100%). AA107841, selenophosphate synthetase-2, and R44842 retained 80, 73.5, and 54% of the wild-type activity, respectively. The activity dropped to 20% with AA280511 (*bar 13*). A correlation between both approaches could be thus established, showing that those SECIS candidates producing full-length GPx also conferred wild-type or significant GPx activity.

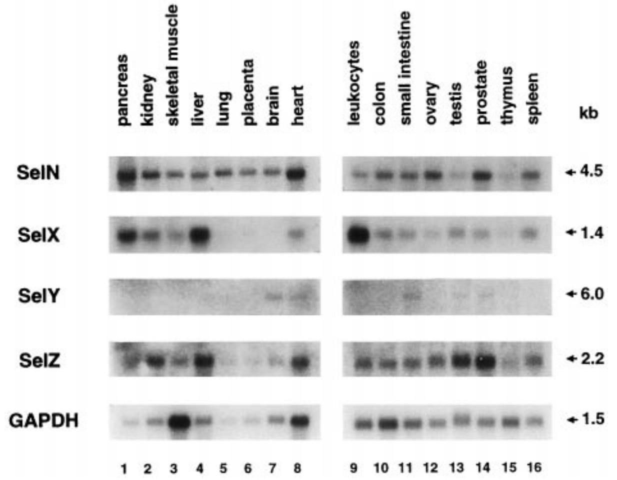


FIG. 3. **Tissue-specific expression patterns of the SelN, SelX, SelY, and SelZ mRNAs.** Shown are the results from Northern blot hybridization of human multiple tissues. Poly(A)$^{+}$ RNAs (CLONTECH) were hybridized sequentially with ^{32}P-labeled probes derived from positions 739–990 in SelX, positions 1218–1889 in SelN, positions 1227–1762 in AF007144 (SelY), and positions 952–1663 in SelZf1 DNAs. *Lanes 1–8* and *9–16* were on separate filters. The glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) mRNA was the internal standard. *Arrows* point to the estimated sizes of the mRNAs in kilobases (*kb*).

Synthesis of a full-length, enzymatically active GPx could be obtained with seven of the selected SECIS elements, indicating that they were capable of promoting selenocysteine insertion. Possible explanations for the inactivity of the other candidates will be discussed.

Identification of the cDNAs Harboring the New Functional SECIS Element—In the previous assay, we functionally characterized the selenophosphate synthetase-2 SECIS element of the selenophosphate synthetase mRNA. Next, we sought the open reading frames lying upstream of the remaining new SECIS elements. ESTs physically linked to each SECIS element were searched in the GenBank™ EST data base with BLASTN. The EST sequences collected after an iterative BLASTN search were processed with the CAP program (16) to assemble one contiguous cDNA sequence. The longest cDNAs were obtained and sequenced. The sequence of the cDNA that we found linked to SECIS AA280511 revealed that the SECIS element resides in fact on the opposite strand relative to the putative ORF. Yet constituting a potential *bona fide* SECIS element, we could not identify an ORF in the proper orientation. We found that SECIS elements AA107841 and R46598 corresponded to the SECIS elements of selenoprotein mRNAs characterized while our study was underway. Indeed, the sequence of the cDNA linked to SECIS AA107841 was found to be identical to that of the 15-kDa selenoprotein (4). The length of the mRNA bearing SECIS R46598, which we call SelY, was estimated to be 6 kilobases by Northern blot analysis (Fig. 3, *lanes 7* and *8*). This size suggested that it could correspond to the mRNA of type 2 iodothyronine deiodinase, whose coding frame, deprived of the 3′-UTR, was isolated earlier (17). Our cloning and sequencing of SelY cDNA showed that it was identical to the 3′-UTR of type 2 iodothyronine deiodinase (18).

Since translation of the cDNA sequences linked to the remaining three SECIS elements, R71722, AA057045, and R44842, showed no homology to known selenoproteins, the cDNAs were termed SelN, SelX, and SelZ, respectively. The sizes of the SelN, SelX, and SelZ mRNAs were estimated by Northern blot analysis to be 4.5, 1.4, and 2.2 kilobases, respectively (Fig. 3). By screening a HeLa oligo(dT) library with a probe complementary to the SelX SECIS DNA, we identified a 1333-bp fragment presumably corresponding to the full-size

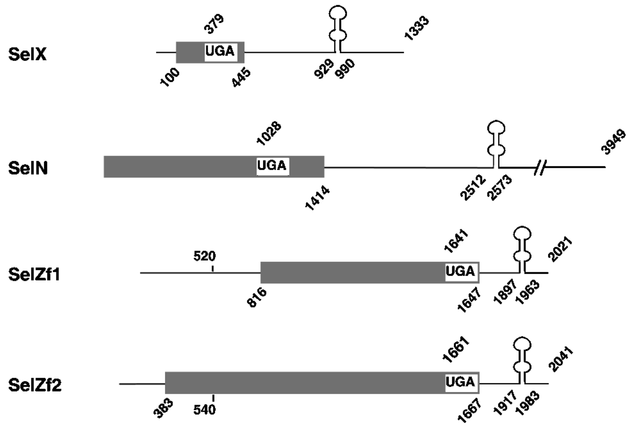


FIG. 4. **Diagrammed representations of the SelX, SelN, SelZf1, and SelZf2 cDNAs.** The coding and untranslated regions are represented by *gray boxes* and *single lines*, respectively. UGA selenocysteine codons are *boxed*; SECIS elements are depicted by *stem-loop structures*.

SelX cDNA. The sequence analysis revealed the existence of a 345-bp-long ORF with an in-frame TGA codon at position 379 (Fig. 4). As expected for a selenoprotein mRNA, its SECIS element effectively resides within the 3′-UTR. Querying EST data bases with BLAST identified a 2231-bp cDNA that was incomplete since the corresponding mRNA was 4.5 kilobases long (Fig. 3). Upstream sequences were thus obtained by screening a HeLa random-primed cDNA library and 5′-Marathon RACE, extending them by 1718 bp. Assembled together, the fragments gave rise to a 3949-bp SelN cDNA, the sequence of which indicated that the reading frame was still open. However, as the 3949-bp SelN cDNA contained a 1414-bp ORF with a characteristic in-frame TGA codon at position 1028, it was used for subsequent analysis. Here also, the SECIS element occurred within the 3′-UTR of the SelN cDNA (Fig. 4).

The sequencing of the EST corresponding to SECIS R44842 determined the presence of a 1505-bp cDNA that contained an ORF that obviously extended upstream of the characterized sequence. This cDNA was called SelZ. Additional 5′-sequences were searched by 5′-Marathon RACE. Surprisingly, we obtained two different PCR fragments with different 5′-sequences. Each fragment obtained, added separately to the SelZ cDNA, generated the 2021-bp SelZf1 and 2041-bp SelZf2 cDNAs. The 5′-sequences of these cDNAs differ upstream of positions 520 in SelZf1 and 540 in SelZf2 and are followed by the common SelZ region (Fig. 4). Since the corresponding transcripts are approximately the same size, they could not be distinguished by Northern blot analysis with a probe complementary to the common SelZ sequence (Fig. 3). Putative ATG initiation codons were identified by the presence of upstream sequences homologous to the Kozak consensus sequence (19) at positions 816 in SelZf1 and 383 in SelZf2. A TGA codon was found in the common region, potentially encoding a selenocysteine at the C-terminal penultimate position in both proteins. For SelZf1 and SelZf2, the SECIS element was localized 250 bp downstream of the putative TAA stop codon (Fig. 4).

Can the New SECIS Elements Mediate Readthrough of the Selenocysteine Codon in Their Own mRNA Contexts?—SelX and SelN were fused at the N terminus to an HA tag, generating constructs HASelX and HASelN, respectively. In SelZf1 and SelZf2, the putative selenocysteine codon resides at the penultimate C-terminal position in a domain common to both proteins. Therefore, only the SelZ common region was epitope-tagged at the N terminus, giving rise to HASelZ. After transfection of the constructs into COS-7 cells, the tag allowed immunodetection by the anti-HA antibody of the proteins contained in the cell extracts, hence evaluation of their sizes. In ΔSECIS constructs, the absence of the SECIS element should convert the UGA selenocysteine to a stop codon, thus producing a shortened polypeptide. Based on the cDNA sequence, HASelX should generate either 17.2- or 15-kDa proteins, according to selenocysteine codon readthrough. Transfection of HASelX indeed generated a major product at ~16 kDa, but also a minor one at ~10 kDa (Fig. 5*A*, *lane 4*), possibly arising from inefficient selenocysteine codon readthrough (6). Construct HASelXΔSECIS, as anticipated, produced almost exclusively the shortest form (*lane 5*). Obtaining the faint SECIS-independent 16-kDa band was reminiscent of what happened with GPx (*lane 3*) and other selenoproteins (5).

A 58-kDa product corresponding to the full-length protein produced by HASelN was expected. Indeed, synthesis of a 60-kDa protein was observed (Fig. 5*A*, *lane 6*). Even though a shorter product of 51 kDa showed up both in the presence and absence of the SECIS element (compare *lanes 6* and *7*), it must be stressed that the expected full-length 60-kDa protein appeared only in the presence of the SECIS element. Since the UGA codon is located at the penultimate position in the SelZ mRNA, we should not expect a difference in the mobilities of the full-length 48-kDa and UGA-terminated proteins. This is effectively what happened (*lanes 8* and *9*). We concluded from these experiments that the SECIS elements in the SelX and SelN mRNAs function to mediate readthrough of the selenocysteine codon, with the only ambiguity remaining for SelZ.

SelX, SelN, and SelZ Are Selenoproteins—To solve the SelZ ambiguity, but also to assert that the new cDNAs do encode selenoproteins, *in vivo* labeling was performed by growing transiently transfected COS-7 cells in a medium containing $Na_2{}^{75}SeO_3$. The HA-tagged proteins were immunoprecipitated from the cell extracts with the anti-HA antibody and fractionated by SDS-polyacrylamide gel electrophoresis. The immunoprecipitation and the difference in size arising from the tag enabled the specific detection of the recombinant selenoproteins. For SelX, SelN, and SelZ, a ^{75}Se-labeled product was obtained only with the SECIS-containing cDNAs (Fig. 5*B*, compare *lanes 4* and *5*, *6* and *7*, and *8* and *9*). The positions of the bands correlated with the protein sizes predicted from the cDNA lengths and with those on the Western blot in Fig. 5*A*. The variable intensities of the bands may be accounted for by differential mRNA or protein stabilities or by different activities carried by different SECIS elements, as previously observed in other contexts (20). In the control experiment, the full-length GPx protein was accompanied by a lower molecular mass product of ~22 kDa, which could arise from proteolysis (*lane 2*). Worth noting is the lack of detection of the full-length GPx, SelX, and SelN proteins that were observed on the Western blots in the absence of SECIS elements (Fig. 5*A*), even after long exposure (data not shown). It may well be that these selenium-lacking proteins originated from weak unspecific readthrough of the selenocysteine codon under our experimental conditions.

These results conclusively demonstrate that SelX, SelN, and SelZ are indeed selenoproteins. Because SelZ exists in two isoforms, this corresponds to four novel selenoproteins: SelX, SelN, SelZf1, and SelZf2. Since the corresponding cDNAs each contain an in-frame TGA codon and a SECIS element, the selenium labeling experiments strongly argue in favor of specific selenocysteine incorporation.

Searching Functions for the New Selenoproteins—Northern blot analysis was performed to determine possible tissue-specific expression of SelX, SelN, and SelZ (Fig. 3). SelN mRNA was ubiquitously expressed, with, however, a higher accumulation in the pancreas, ovary, prostate, and spleen. The distri-

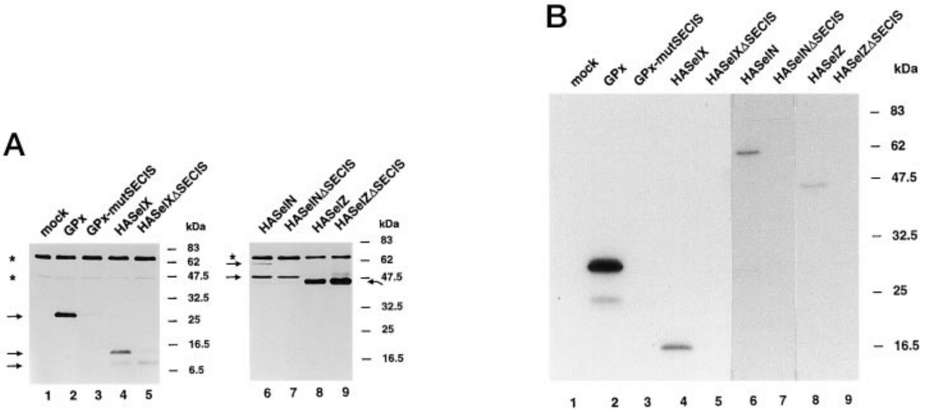


FIG. 5. **Translation of and ^{75}Se incorporation into SelN, SelX, and SelZ.** *A*, the SECIS elements mediate UGA readthrough from their own mRNA contexts. After transfection of the constructs (with SECIS (*lanes 4*, *6*, and *8*) and lacking SECIS (*lanes 5*, *7*, and *9*)) into COS-7 cells, the HA-tagged proteins were revealed by Western blot analysis with the anti-HA antibody. Control lanes are the same as described for Fig. 2*A*. Migrations in *lanes 1–5* and *6–9* were on 10 and 12% gels, respectively. *Arrows* point to the translation products mentioned under "Results"; *asterisks* indicate unspecific products. *B*, SelN, SelX, and SelZ are selenoproteins. Transfected COS-7 cells were cultured in the presence of ^{75}Se. The HA-tagged ^{75}Se-labeled proteins were immunoprecipitated, fractionated on a 12% gel, and revealed by autoradiography.

bution of the SelX mRNA was less homogenous than that of SelN, being preponderant in the liver and leukocytes, abundant in the pancreas, but low in the lung, placenta, and brain. SelZ mRNA showed more pronounced accumulation in the kidney, liver, testis, and prostate, but was low in the thymus.

In the course of this study, the cDNA for the selenoprotein TrxR2, a mitochondrion-specific thioredoxin reductase isoform, was cloned independently by several groups (21–23). Sequence comparisons between the SelZf1, SelZf2, and TrxR2 cDNAs, depicted schematically in Fig. 7*A*, indicated that they share a large common domain. The SelZf1 and TrxR2 cDNA sequences are identical from the 3′-end to residue 636 of TrxR2. In the SelZf2 cDNA, the region conserved with TrxR2 extends up to position 293 of TrxR2. The common region in the three cDNAs includes the 3′-part of the coding sequence with the in-frame TGA codon and the 3′-UTR, with sequence differences occurring at their 5′-ends. The three cDNAs encode three different proteins sharing a common core, but with different N-terminal domains.

Alignment of the human SelN DNA sequence with ESTs or of the SelN protein sequence with translated ESTs revealed the existence of a hypothetical ortholog in mouse and rat. The number of different ESTs was insufficient for reconstitution of complete cDNAs, but the partial assembled sequences showed conservation of the coding frames, in-frame TGA codons, and SECIS elements.

We next sought homologs to SelX. A mouse cDNA covering the entire length of the human SelX cDNA was reconstituted *in silico* by merging various overlapping mouse ESTs. The translated mouse cDNA showed 91% amino acid identity to the human SelX protein. Furthermore, data base searches found SelX sequence similarities to plant and *Drosophila* translated ESTs, but also to prokaryotic, yeast, and *Caenorhabditis elegans* ORFs indexed as hypothetical proteins of unknown function. Displayed in Fig. 6, these findings show striking amino acid identities between, for example, human SelX and *Escherichia coli* P39903 (24%), *C. elegans* P34436 (28%), and *Drosophila* EST AA540562 (28%). The comparison also stressed the 29% amino acid identity of the human and mouse SelX proteins to a domain of the *Neisseria gonorrhoeae*, *Hemophilus influenzae*, *Helicobacter pylori*, *Mycoplasma capricolum*, and *Streptococcus pneumoniae* PILB proteins, regulators of bacterial pilus formation (24). Although the sequences are similar over their entire lengths, the alignment highlights two blocks of higher sequence conservation: PWPAF (*1*)∝GLGHEF (*2*) in mammalian SelX and GWP(A/S)F (*1*)∝HLGHVF (*2*) in the SelX homologs (*blocks 1* and *2* in Fig. 6; the only two positions where *X* and M replaced L could originate from sequence uncertainties in the corresponding ESTs). It is striking that only the mammalian SelX proteins incorporate selenocysteine, whereas other organisms contain a cysteine instead. Sequence conservation is observed flanking the cysteine/selenocysteine (U): R(Y/H)C(I/V/M)N in SelX homologs and RFUIF in mammalian SelX.

DISCUSSION

The objective of our study was the isolation of new selenoprotein cDNAs. The existence of selenoproteins other than those previously characterized was predicted by workers based on selenium labeling experiments, but did not lead to amino acid sequence data. To circumvent the lack of protein sequence information, we assumed that a number of the desired cDNA sequences were already deposited in the EST data bases. To exploit this information, our strategy took advantage of the obligatory presence of a SECIS element in all selenoprotein mRNAs. This differs from conventional screens in two respects. The SECIS hairpin being characterized more by the high conservation of its secondary structure than by the extent of invariant sequences, alignment methods such as BLAST and FASTA were inappropriate. The originality of our approach was the use of a program capable of detecting RNA foldings such as the SECIS consensus secondary structure. Another and probably the most important aspect of our screen is that selenoprotein cDNAs contain TGA codons, obviously rendering the identification of an ORF more challenging than in other cDNAs where TGA signals the end of the ORF. Notwithstanding, the strategy paid off since the RNA structure alone was sufficient to discover four novel different selenoproteins.

Seven SECIS candidates, out of the 21 selected *in silico*, indeed corresponded to functional SECIS elements. This came as a surprise since the inactive candidates harbored the features defined by the SECIS consensus structure. Several possibilities can explain this paradoxical situation. The SECIS losers may lack one or more essential sequences or base pairs that could have been unintentionally omitted in the SECIS descriptor because they were not yet identified in the then known SECIS elements. Alternatively, the SECIS losers may contain sequence or base pair anti-determinants preventing them from functioning. Finally, the sequences may fold *in vivo* into structures slightly different from the expected one.

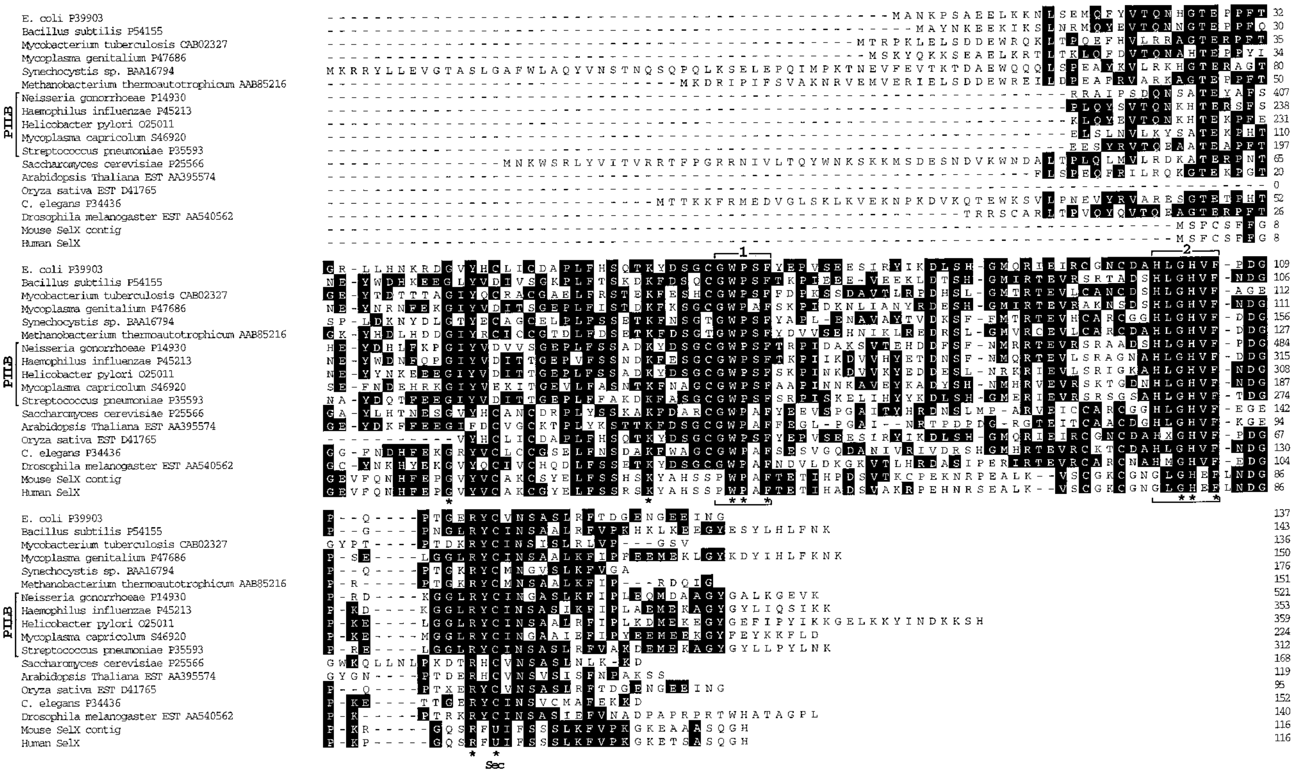

E. coli P39903
Bacillus subtilis P54155
Mycobacterium tuberculosis CAB02327
Mycoplasma genitalium P47686
Synechocystis sp. BAA16794
Methanobacterium thermoautotrophicum AAB85216
PILB
Neisseria gonorrhoeae P14930
Haemophilus influenzae P45213
Helicobacter pylori O25011
Mycoplasma capricolum S46920
Streptococcus pneumoniae P35593
Saccharomyces cerevisiae P25566
Arabidopsis Thaliana EST AA395574
Oryza sativa EST D41765
C. elegans P34436
Drosophila melanogaster EST AA540562
Mouse SelX contig
Human SelX
1
2
Sec

FIG. 6. **Sequence alignment of human SelX and similar proteins identified by data base searches.** The amino acid sequences deduced from ORFs or ESTs were aligned with the PILB proteins. The mouse SelX contig is the translation of a cDNA contig constructed from several overlapping ESTs. *Hyphens* indicate gaps. In the mouse and human SelX sequences, U stands for selenocysteine, marked *Sec* below the sequence. For the five PILB proteins, only the domain similar to SelX is shown. Identical amino acids are *shaded*; invariant positions are shown by *asterisks*. Conserved *blocks 1* and *2* mentioned under "Results" and "Discussion" are indicated. Amino acid positions are shown on the right.

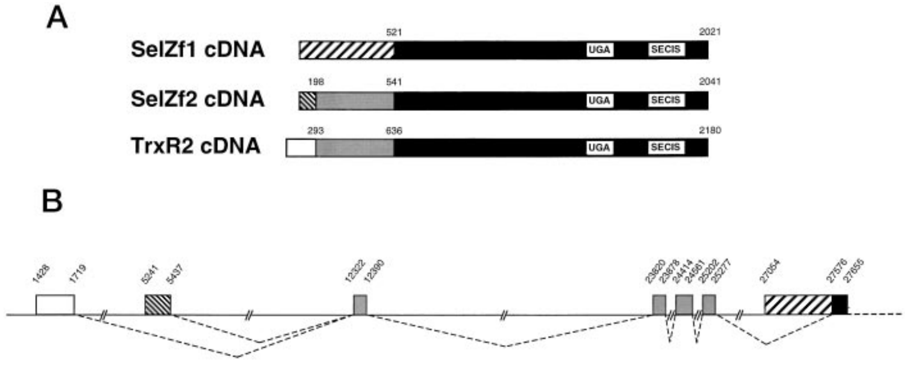


FIG. 7. **Homologies in SelZf1, SelZf2, and TrxR2 cDNAs.** Shown is a model for genomic organization. *A*, SelZf1, SelZf2, and TrxR2 cDNAs share common and specific regions. The similarly *boxed* common regions, the UGA selenocysteine codons, and the conserved SECIS elements are drawn. *B*, a BLAST search with the three cDNAs identified sequence similarities to cosmid 56c of chromosome 22q11.2 (GenBank™ accession number AC000090). The possible genomic organization, giving rise to the cDNAs in *A*, was obtained by joining the *boxes* according to the *dashed line*. Positions correspond to cosmid coordinates.

Three SECIS elements among the seven winners led to the discovery of the SelN, SelX, and SelZ selenoprotein mRNAs, SelZ giving rise to the SelZf1 and SelZf2 isoforms. *In vivo* expression of the selenoprotein mRNAs indicated that selenocysteine incorporation was actually dependent on the presence of the SECIS element. No sequence similar to SelN could be found in protein or nucleotide sequence data bases. However, similarity searches were productive with SelX and SelZ. The amino acid comparisons in Fig. 6 underscored two prominent features of SelX. First, sequences similar to mammalian SelX were detected in all kingdoms. The human and mouse sequences had 24–28% amino acid identities to ORFs of unknown function in *E. coli* and *C. elegans* and in a *Drosophila* EST. Second, we found that mammalian SelX displayed 29% amino acid identity to a domain of PILB, a protein involved in pilus formation in the bacteria *N. gonorrhoeae*, *H. influenzae*, *H. pylori*, *M. capricolum*, and *S. pneumoniae* (Fig. 6). PILB possesses a peptide methionine-sulfoxide reductase activity (25). Sequence comparisons established that this activity resides in a PILB subdomain different from the SelX similarity. The situation differs in *E. coli*, where the peptide methionine-sulfoxide reductase activity is borne by MsrA, a polypeptide different from P39903, one of the hypothetical proteins identified by similarity to SelX (Fig. 6). From these observations, it looks as if SelX constitutes a functional module, acting *per se* or associated with peptide methionine-sulfoxide reductase in the bacterial polyprotein PILB. The conserved amino acids in *blocks 1* and *2* as well as the selenocysteine (Fig. 6) certainly play important roles in the function of SelX.

The C-terminal domains of SelZf1 and SelZf2 show clear homologies to the corresponding domain of the selenoprotein TrxR2. Interestingly, it was shown that the 293 bp at the 5′-end of the TrxR2 cDNA encode the mitochondrial targeting peptide (23), which is not found in SelZf2 (Fig. 7*A*). More surprisingly, the region of the cDNAs encoding the CVNVGC active site, common to the mitochondrial and cytoplasmic thioredoxin reductases and to the glutathione reductase (22), was found in the SelZf2 cDNA, but not in the SelZf1 cDNA. This suggests for SelZf1 a different function compared with SelZf2 and TrxR2. In the course of searching sequences similar to the SelZf1 and SelZf2 cDNAs, we identified genomic fragments (GenBank^{TM} accession numbers AC000079 and AC000080) with similarity to both cDNAs. An identical genomic fragment was also shown independently by others (23) to contain sequences encoding TrxR2. Alignment of the SelZf1, SelZf2, and TrxR2 cDNA sequences with the genomic sequence yielded the putative assembly pattern in Fig. 7*B*, obtained by removing the introns. We could see that those domains that differ between the three cDNAs (extending from the 5′-ends to positions 521, 198 and 541, and 293 and 636 in SelZf1, SelZf2, and TrxR2, respectively) correspond to distinct genomic segments. The three cDNAs should arise from the same gene, probably by alternative splicing resulting in the addition of different 5′-segments to a common core to generate three different selenoproteins with specialized functions or localizations.

Previous reports underscored the relevance of computational searches for identifying RNA structure motifs (26–28). Recently also, a computational screen using an original algorithm was employed to uncover methylation guide small nucleolar RNAs in the yeast genome (29). The peculiarity of our study resides in that the strategy employed led to the discovery of four novel selenoproteins. This once again illustrates the value of mRNA 3′-UTRs as a repository of functional RNA motifs instrumental in post-transcriptional control. Undoubtedly, with hundreds of new EST sequences deposited every day in the data bases and in the perspective of the completion of the human genome sequencing project, this strategy will enable more selenoproteins to be discovered. This could also be extended to the discovery in other organisms of mRNAs whose stability or localization is mediated by common structural motifs in the 3′-UTR.

Acknowledgments—We are grateful to P. Chambon, S. Elledge, F. Guillemot, J. L. Mandel, M. M. Y. Waye, and W. Yu for the gifts of various materials; D. Fagegaltier, M. Frugier, L. Jaeger, and K. Yamada for careful reading of the manuscript; and C. Loegler for skillful technical assistance.

REFERENCES

1. Beck, M. A., and Levander, O. A. (1998) *Annu. Rev. Nutr.* **18,** 93–116
2. Stadtman, T. C. (1991) *J. Biol. Chem.* **266,** 16257–16260
3. Burk, R. F., and Hill, K. E. (1999) *Bioessays* **21,** 231–237
4. Gladyshev, V. N., Jeang, K. T., Wootton, J. C., and Hatfield, D. L. (1998) *J. Biol. Chem.* **273,** 8910–8915
5. 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
6. Low, S. C., and Berry, M. J. (1996) *Trends Biochem. Sci.* **21,** 203–208
7. Walczak, R., Westhof, E., Carbon, P., and Krol, A. (1996) *RNA* (*N. Y.*) **2,** 367–379
8. Walczak, R., Carbon, P., and Krol, A. (1998) *RNA* **4,** 74–84
9. Bösl, M. R., Takaku, K., Oshima, M., Nishimura, S., and Taketo, M. M. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 5531–5534
10. de Haan, J. B., Bladier, C., Griffiths, P., Kelner, M., O'Shea, R. D., Cheung, N. S., Bronson, R. T., Silvestro, M. J., Wild, S., Zheng, S. S., Beart, P. M., Hertzog, P. J., and Kola, I. (1998) *J. Biol. Chem.* **273,** 22528–22536
11. Behne, D., Kyriakopoulos, A., Weiss, N. C., Kalckloesch, M., Westphal, C., and Gessner, H. (1996) *Biol. Trace Elem. Res.* **55,** 99–110
12. Gautheret, D., Major, F., and Cedergren, R. (1990) *Comput. Appl. Biosci.* **6,** 325–331
13. Laferriere, A., Gautheret, D., and Cedergren, R. (1994) *Comput. Appl. Biosci.* **10,** 211–212
14. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680
15. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402
16. Huang, X. (1992) *Genomics* **14,** 18–25
17. Croteau, W., Davey, J. C., Galton, V. A., and St-Germain, D. (1996) *J. Clin. Invest.* **98,** 405–417
18. Buettner, C., Harney, J. W., and Larsen, P. R. (1998) *J. Biol. Chem.* **273,** 33374–33378
19. Kozak, M. (1997) *EMBO J.* **16,** 2482–2492
20. Kollmus, H., Flohé, L., and McCarthy, J. E. G. (1996) *Nucleic Acids Res.* **24,** 1195–1201
21. Gasdaska, P. Y., Berggren, M. M., Berry, M. J., and Powis, G. (1999) *FEBS Lett.* **442,** 105–111
22. Lee, S. R., Kim, J. R., Kwon, K. S., Yoon, H. W., Levine, R. L., Ginsburg, A., and Rhee, S. G. (1999) *J. Biol. Chem.* **274,** 4722–4734
23. Miranda-Vizuete, A., Damdimopoulos, A. E., Pedrajas, J. R., Gustafsson, J. A., and Spyrou, G. (1999) *Eur. J. Biochem.* **261,** 405–412
24. Taha, M. K., So, M., Seifert, H. S., Billyard, E., and Marchal, C. (1988) *EMBO J.* **7,** 4367–4378
25. Wizemann, T. M., Moskovitz, J., Pearce, B. J., Cundell, D., Arvidson, C. G., So, M., Weissbach, H., Brot, N., and Masure, H. R. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 7985–7990
26. Dandekar, T., and Hentze, M. (1995) *Trends Genet.* **11,** 45–50
27. Dandekar, T., Beyer, K., Bork, P., Kenealy, M.-R., Pantopoulos, K., Hentze, M., Sonntag-Buck, V., Flouriot, G., Gannon, F., Keller, W., and Schreiber, S. (1998) *Bioinformatics* **14,** 271–278
28. Pesole, G., Liuni, S., Grillo, G., Ippedico, M., Larizza, A., Makalowski, W., and Saccone, C. (1999) *Nucleic Acids Res.* **27,** 188–191
29. Lowe, T. M., and Eddy, S. R. (1999) *Science* **283,** 168–171