

April 1998

# A New Human Selenium-containing Protein

Vadim N. Gladyshev

*University of Nebraska-Lincoln, vgladyshev1@unl.edu*

Kuan Teh Jeang

*National Institutes of Health, Bethesda, Maryland*

John C. Wootton

*National Institutes of Health, Bethesda, Maryland*

Dolph L. Hatfield

*National Institutes of Health, Bethesda, Maryland*

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



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

---

Gladyshev, Vadim N.; Jeang, Kuan Teh; Wootton, John C.; and Hatfield, Dolph L., "A New Human Selenium-containing Protein" (1998). *Vadim Gladyshev Publications*. 46.

<http://digitalcommons.unl.edu/biochemgladyshev/46>

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.

# A New Human Selenium-containing Protein

## PURIFICATION, CHARACTERIZATION, AND cDNA SEQUENCE\*

(Received for publication, October 6, 1997, and in revised form, February 2, 1998)

Vadim N. Gladyshev<sup>‡§¶</sup>, Kuan-Teh Jeang<sup>||</sup>, John C. Wootton<sup>\*\*</sup>, and Dolph L. Hatfield<sup>‡ ¶¶</sup>

From the <sup>‡</sup>Section on the Molecular Biology of Selenium, Basic Research Laboratory, NCI, the <sup>§</sup>Laboratory of Biochemistry, NHLBI, the <sup>||</sup>Laboratory of Molecular Microbiology, NIAID, and the <sup>\*\*</sup>National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20892

Selenium which occurs in proteins as the amino acid, selenocysteine, is essential for numerous biological processes and for human health. A prominent <sup>75</sup>Se-labeled protein detected in human T-cells migrated as a 15-kDa band by SDS-polyacrylamide gel electrophoresis. This protein subunit was purified and subjected to tryptic digestion and peptide sequence analyses. Sequences of tryptic peptides derived from the protein corresponded to a human placental gene sequence containing an open reading frame of 162 residues and a readthrough in-frame TGA codon. Three different peptide sequences of the 15-kDa protein corresponded to a nucleotide sequence located downstream of this codon, suggesting that the T-cell 15-kDa selenoprotein contains a selenocysteine residue encoded by TGA. Post-translational processing of the N-terminal portion of the predicted gene product to give the 15-kDa protein was suggested on the basis of molecular mass, amino acid analysis, and immunoblot assays of the purified protein. The 3'-untranslated region (UTR) of the gene encoding the 15-kDa protein contained a sequence that is very similar to the canonical selenocysteine-inserting sequence element. Computer analysis of transcript map data bases indicated that this gene was located on human chromosome 1. Its coding sequence showed no homology to known protein-encoding genes. The 15-kDa protein gene was expressed as mRNA in a wide range of tissues, with increased levels in the thyroid, parathyroid, and prostate-derived cells as evidenced by searches of partial cDNA sequences in public data bases. Genes corresponding to the 15-kDa selenocysteine-containing protein were found in mice and rats, while the corresponding genes in *Caenorhabditis elegans* and *Brugia malayi* contained a cysteine codon in place of TGA. The discovery of a new human selenoprotein provides an additional example of the role of selenium in mammalian systems.

Selenium has been implicated in immunological function and many other biological processes through various nutritional

\* 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) AF051894.

¶ Present address: Dept. of Biochemistry, University of Nebraska, Lincoln, NE 68588-0664. Tel.: 402-472-4948; Fax: 402-472-7842; E-mail: VNG@unlinfo2.unl.edu.

¶¶ To whom correspondence should be addressed: Basic Research Laboratory, NCI, National Institutes of Health, Bldg. 37, Rm. 2D09, 9000 Rockville Pike, Bethesda, MD 20892. Tel.: 301-496-2797; Fax: 301-435-4957; E-mail: hatfield@dc37a.nci.nih.gov.

and biochemical studies (1, 2). This trace element is a natural component of several prokaryotic and eukaryotic proteins. Although selenium occurs in prokaryotic proteins either as a cofactor or as a selenocysteine residue, mammalian selenoproteins identified thus far contain selenium only in the form of selenocysteine, which is the 21st naturally occurring amino acid in protein. A selenocysteine tRNA that decodes UGA has been found in all life kingdoms, suggesting that the use of UGA as a codon for selenocysteine is widespread in nature (2). The special conserved stem-loop structures in the 3'-untranslated regions of mammalian selenoprotein mRNAs are essential for recognition of UGA as a codon for selenocysteine, rather than a codon for termination of translation (3).

Of the eleven genes encoding different selenocysteine-containing proteins that have been found thus far in mammals, four encode various glutathione peroxidases (reviewed in Refs. 4 and 5), three encode different thyroid hormone deiodinases (6–8), and others encode thioredoxin reductase (9), selenophosphate synthetase 2 (SPS2) (10), selenoprotein P (11), and selenoprotein W (12).

Selenocysteine is located at the active center and is directly involved, or at least implicated, in the catalytic reactions catalyzed by glutathione peroxidases, thyroid hormone deiodinases, and selenophosphate synthetase 2. Thioredoxin reductase contains selenocysteine (13) in a novel C-terminal Gly-Cys-Sec-Gly redox motif (14). This center has been implicated in the peroxidase reaction catalyzed by the enzyme (14) and in a redox interaction with the N-terminal redox disulfide (15), although further studies are necessary to prove the suggested essential role of selenocysteine in this protein. Selenoprotein P, a protein of unknown function, is unusual in that it contains ten selenocysteine residues. The function of a selenoprotein W also remains unknown.

In the present study, the isolation of the 15-kDa protein and its gene sequence are described which characterize this protein as a member of a class of mammalian selenoproteins. Although this new human selenoprotein does not have homology to known proteins, genes putatively encoding this protein occur in other mammals as well as in nematodes and plants. Preliminary data on this protein have been presented elsewhere (16).

### EXPERIMENTAL PROCEDURES

**Materials**—<sup>75</sup>SeSelenious acid was obtained from the Research Reactor Facility, University of Missouri (Columbia, MO), ECL systems were from Amersham Pharmacia Biotech, EST clones were from ATCC (number 384717 from human placental cDNA library and number 409024 from human infant brain cDNA library), and other reagents were commercial products of the highest grade available.

**Cell Growth and Protein Purification**—A human Jurkat T-cell line, JPX9 (17), was grown and labeled with <sup>75</sup>SeSelenious acid (2 μCi/ml) as described (14). <sup>75</sup>Se-Labeled JPX9 cells were mixed with unlabeled cells, suspended in 2 volumes of 30 mM Tris-HCl, pH 7.5, 1 mM EDTA, 2 mM dithiothreitol, 1 mM MgCl<sub>2</sub>, 1 mM phenylmethylsulfonyl fluoride, and disrupted by sonication. Disrupted cells were centrifuged; the su-

pernatant was applied to a DEAE-Sepharose column, which had been equilibrated with 30 mM Tris-HCl, pH 7.5, 2 mM dithiothreitol, and 1 mM EDTA (buffer A); the column was washed with 2 volumes of buffer A; and proteins were eluted by application of a linear gradient from 0 to 500 mM NaCl in buffer A. Fractions containing  $^{75}\text{Se}$  were analyzed on SDS gels. Fractions containing the 15-kDa selenoprotein that eluted from the DEAE column with 350 mM NaCl were combined, concentrated, adjusted to a concentration of 0.5 M ammonium sulfate in buffer A, and applied to a phenyl-Sepharose column equilibrated in 1 M ammonium sulfate in buffer A; the column was washed by application of a linear gradient from 0.5 to 0 M ammonium sulfate in buffer A; and radioactive fractions corresponding to the 15-kDa protein were eluted by application of a linear gradient from buffer A to water. Radioactive fractions were combined, concentrated, and loaded on a  $\text{C}_{18}$  reversed-phase high performance liquid chromatography column that had been equilibrated in 0.05% trifluoroacetic acid, a gradient of 0 to 60% acetonitrile in 0.05% trifluoroacetic acid applied and  $^{75}\text{Se}$ -containing fractions corresponding to the 15-kDa protein eluted at 48% acetonitrile.

**Characterization of the 15-kDa Selenoprotein**—Fractions containing the 15-kDa selenoprotein from the  $\text{C}_{18}$  column were dried on a Speed-Vac SC110 (Savant), dissolved in SDS-PAGE<sup>1</sup> sample buffer, and analyzed by SDS-PAGE. The molecular mass of the 15-kDa selenoprotein was determined by electrospray and MALDI mass-spectrometry in fractions from the  $\text{C}_{18}$  column. Both mass spectra revealed a single strong signal of the 15-kDa protein. The native molecular mass of the 15-kDa selenoprotein purified on a DEAE-Sepharose column was determined using native PAGE and analytical high performance liquid chromatography gel filtration as described (18). The 15-kDa selenoprotein was detected as  $^{75}\text{Se}$ -labeled fractions from a gel-filtration column and as a  $^{75}\text{Se}$ -labeled band on native PAGE. For determination of sequences of internal peptides, the 15-kDa selenoprotein was separated by SDS-PAGE, transferred onto a polyvinylidene difluoride membrane, and stained with Ponceau S. The stained single 15-kDa band was removed and submitted to Harvard Microchem (Boston, MA), where the 15-kDa selenoprotein was digested with trypsin and the resulting peptides were separated on a reversed phase high performance liquid chromatography column and analyzed by MALDI mass-spectrometry and amino acid sequencing. Amino acid composition of the 15-kDa protein was also determined at Harvard Microchem. These data were carefully analyzed and found to be consistent with other experimental data (mass spectrometry, SDS-PAGE analyses, immunoblot assays).

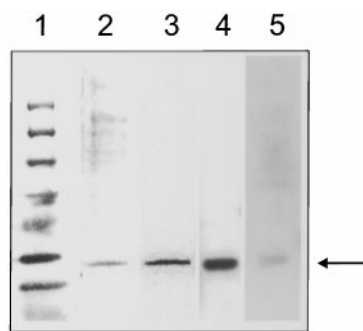
**Immunoblot Detection**—Rabbit antibodies were raised against a synthetic polypeptide corresponding to residues 145–162 of the human 15-kDa protein and used for Western blot analyses with ECL systems.

**cDNA Sequencing**—Plasmids were isolated according to the instructions provided with the plasmid purification kit (Qiagen), the sequencing reaction products were purified on separation columns as described by the manufacturer (Princeton Separations), and the nucleotide sequences of EST clones were determined using a Dye Terminator Cycle Sequencing kit as described by the manufacturer (Perkin-Elmer).

**Computer Analyses**—Three different peptide sequences from the 15-kDa selenoprotein were analyzed for matches to the data base of expressed sequence tags (dbEST) of partial cDNA sequences (19) using the BLAST (20, 21) and gapped BLAST-2 (22) programs. Multiple alignments of expressed sequence tag (EST) sequences and their translated products were viewed using the MSPcrunch/Blixem system (23). The Blixem alignments also revealed polymorphic sites in the human ESTs that were clearly distinct from sequencing errors.

## RESULTS AND DISCUSSION

**Detection, Purification, and Characterization of the 15-kDa Human Selenoprotein**—A human T-cell line, JPX9, was grown in the presence of  $^{75}\text{Se}$ , and the extracts of  $^{75}\text{Se}$ -labeled cells were analyzed by SDS-PAGE and PhosphorImager detection of radioactivity on the gels. One of the major  $^{75}\text{Se}$ -labeled proteins that migrated as a 15-kDa band on SDS-PAGE was purified initially on DEAE-Sepharose and phenyl-Sepharose columns and then further on a reversed-phase column as described under "Experimental Procedures." The resulting protein prep-



**FIG. 1. Characterization of  $^{75}\text{Se}$ -labeled human T-cell 15-kDa selenoprotein.**  $^{75}\text{Se}$ -labeled 15-kDa selenoprotein was purified as described under "Experimental Procedures" and analyzed by SDS-PAGE, PhosphorImager detection, and immunoblot assays. Lane 1, standards (from top to bottom: 98, 64, 50, 36, 30, 16, and 6 kDa); lane 2, SDS-PAGE analysis of purified 15-kDa protein; lane 3, PhosphorImager detection of  $^{75}\text{Se}$  in purified 15-kDa protein on SDS gel; lane 4, immunoblot detection of purified 15-kDa protein; lane 5, immunoblot detection of 15-kDa protein after DEAE purification step. Four individual images, which were generated by scanning the Coomassie Blue stained gel (lanes 1 and 2), scanning the PhosphorImager data (lane 3), and scanning the data on films from immunoblots (lanes 4 and 5) were used to assemble this figure. Location of the 15-kDa protein is indicated by the arrow.

aration was analyzed by SDS-PAGE. A major protein band migrating as a 15-kDa species was detected by staining the gels with Coomassie Blue and by PhosphorImager detection of radioactivity in the gel (Fig. 1). The purity of the 15-kDa protein was estimated to be about 50%. The native 15-kDa selenoprotein migrated as a 240-kDa species on a native gradient PAGE and was eluted as a 200-kDa species from a calibrated gel-filtration column (not shown). It was not determined whether the native 200–240 kDa protein is composed of multiple identical 15-kDa Se-containing subunits or is a hetero-oligomeric complex. The molecular mass of the 15-kDa selenoprotein subunit in fractions from the  $\text{C}_{18}$  column determined by MALDI mass-spectrometry was 14,830 Da. Electrospray mass-spectrometry of the same preparation revealed a molecular mass of 14,870 Da. The N terminus of the protein was blocked, which prevented determination of the N-terminal sequence. Amino acid analysis revealed the lack of internal methionine and histidine residues, as well as the hydrophobic character of the protein (Table I).

**Sequences of Tryptic Peptides**—The sequences of three different tryptic peptides and one overlapping peptide from the 15-kDa protein were determined (Fig. 2). Computer searches of the partial cDNAs in the dbEST using TBLASTN program revealed nucleotide sequences that corresponded to all three peptides in the same ORF.

**Characterization of cDNA and Polypeptide Sequences**—Several human dbEST sequences corresponded perfectly to peptides from the 15-kDa protein and were used to assemble an open reading frame (Fig. 2). The two cDNA clones containing the longest 5' sequences were obtained from IMAGE and sequenced. These clones revealed a continuous nucleotide sequence of 1268 nucleotides, containing a single open reading frame of 162 amino acid residues and a 3'-end poly(A) tail. A single ATG codon occurred in a nucleotide context, GCGATGG, that is similar to the Kozak consensus sequence for initiation of translation (24). This initiation ATG codon was followed by a 489-nucleotide open reading frame with an in-frame TAA termination codon. The obtained ORF included an in-frame TGA codon, suggesting the presence of a selenocysteine residue, Sec-93 (2). Three tryptic peptides for which sequences have been determined corresponded to deduced sequences located downstream of the TGA codon, indicating readthrough of the

<sup>1</sup> The abbreviations used are: PAGE, polyacrylamide gel electrophoresis; Sec, selenocysteine; ORF, open reading frame; EST, expressed sequence tag; dbEST, data base of expressed sequence tags; MALDI, matrix-assisted laser desorption ionization; 3'-UTR, 3'-untranslated region; SECIS element, selenocysteine insertion sequence element; GPX, glutathione peroxidase.

TGA codon rather than termination of translation. Although selenocysteine was not directly identified as a component of the 15-kDa protein, the labeling of the protein with  $^{75}\text{Se}$ , readthrough of the TGA codon, and the location of selenocysteine insertion sequence (SECIS) element in the untranslated area (Fig. 2) suggest the presence of selenocysteine in the protein. The predicted ORF encoded a protein of 17,790.6 Da. The mass of the purified 15-kDa protein was 14,870 Da, and this discrepancy suggested post-translational processing of the protein. Processing of the 15-kDa protein appears to occur at the N-terminal portion of the protein. That is, antiserum was raised to a synthetic peptide that was identical in sequence to

the eighteen C-terminal residues of the 15-kDa protein, and thus the antigenic site resides near the C terminus. As shown in Fig. 1, this antiserum recognized 15-kDa protein at different stages of purification. In addition, one of the sequenced tryptic peptides obtained from digests of the 15-kDa protein corresponded to residues 146–158, which are located near the C terminus according to the predicted gene sequence.

The N-terminal portion of the putative precursor of the 15-kDa protein, as predicted from the gene sequence, had a stretch of hydrophobic amino acid residues, suggesting the presence of a signal peptide. Cleavage of these N-terminal amino acid residues is consistent with the amino acid composition of the protein (Table I) since the processed protein matches more closely the amino acid analysis data obtained for the purified 15-kDa protein than the full size 17-kDa protein. One possible site for post-translational processing is Ser-27, which coincides with the site of an exon-intron junction (not shown), making this residue the evolutionary favorable site for post-translational processing. The unknown nature of the N-terminal modifying group in the purified protein makes it difficult to predict the exact site of post-translational cleavage.

*Homologous Mouse, Rat, Brugia malayi, Caenorhabditis elegans, and Rice Gene Sequences*—Computer sequence analyses of the 15-kDa protein and its gene sequence revealed no homology to known proteins. However, a number of dbEST sequences from mouse, rat, *B. malayi*, *C. elegans*, and rice showed strong homology in TBLASTN searches with the 15-kDa human protein (Fig. 3). Mouse partial cDNA sequences were assembled into a full-length open reading frame, whereas rat, *C. elegans*, *B. malayi*, and rice sequences gave only partial coding sequences. Interestingly, although mouse and rat genes encode potential selenocysteine-containing 15-kDa proteins, the genes in *C. elegans* and *B. malayi* encode homologous proteins containing cysteine in place of selenocysteine. This is consistent with observations that nematode genes for glutathione peroxidase and thioredoxin reductase encode cysteine analogs of mammalian selenocysteine-containing proteins.

The regions flanking Sec-93 in the human 15-kDa protein had the highest degree of homology among proteins from different organisms, suggesting that the selenocysteine residue is located in a putative active center. In other mammalian sel-

TABLE I  
Amino acid composition of the human 15-kDa selenoprotein

The 162 residue sequence corresponds to a full size open reading frame. The 135 residue sequence corresponds to the open reading frame after removal of 27 N-terminal residues. Experimental data were obtained as described under "Experimental Procedures" and were normalized to a 14.87-kDa molecular mass for the purified protein determined by mass-spectrometry. U represents selenocysteine.

Amino acid	162 Residues	135 Residues	Experimental data
A	13	8	9.1
R	8	7	6.5
N	5	5	ND
D	8	8	ND
C	8	7	ND
E	13	13	ND
Q	8	7	ND
G	13	10	12.9
H	0	0	<0.5
I	5	5	5.0
L	26	20	19.2
K	9	9	9.3
M	1	0	<0.5
F	10	9	7.5
P	6	4	5.0
S	13	11	11.4
T	3	2	3.2
W	1	1	ND
Y	2	2	2.7
V	9	6	6.7
U	1	1	ND
E+Q	21	20	21.4
D+N	13	13	14.6

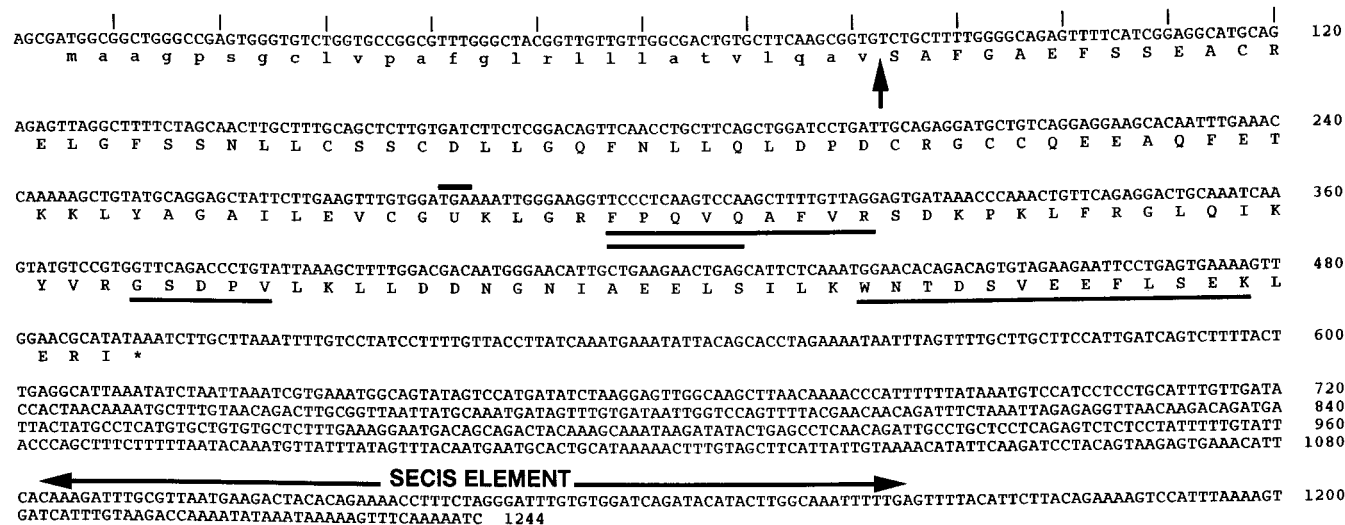
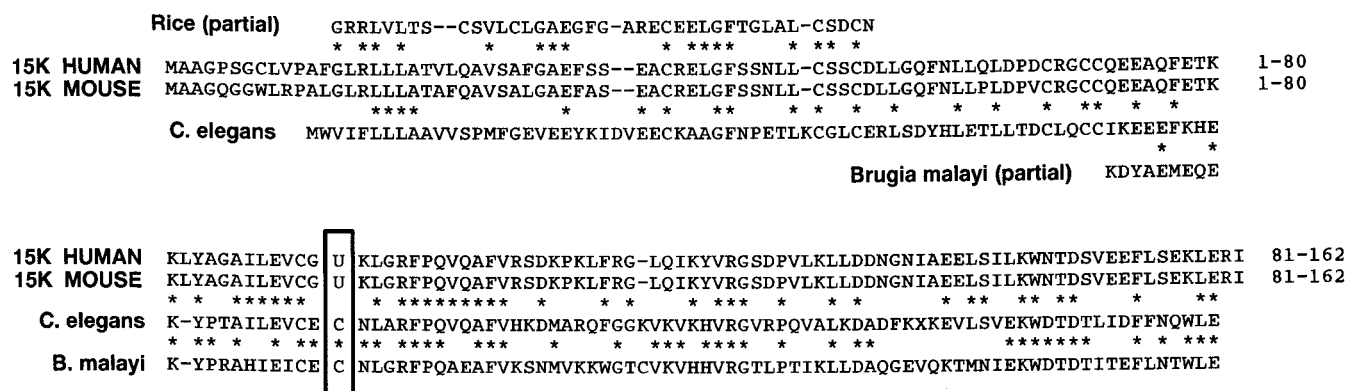


Fig. 2. Human cDNA sequence encoding a 15-kDa Se-containing protein. In the deduced amino acid sequence, the putative signal peptide is shown in lowercase, and the most probable site of post-translational cleavage is indicated by an upward arrow. The amino acid U represents selenocysteine 93 encoded by an in-frame TGA codon (overlined). The sequences of four tryptic peptides, for which amino acid sequences were experimentally determined, are underlined. Mass spectrometry data for these peptides are consistent with the amino acid sequences. In the 3'-UTR, the positions of the selenocysteine insertion sequence (SECIS element, see text and Fig. 4) and the poly(A) addition signal (dotted underline) are shown.



**FIG. 3. Alignment of the human 15-kDa selenoprotein sequence with homologs from mouse, nematodes, and rice.** The amino acid sequence of the mouse protein was deduced from the assembly of 39 independent partial cDNA sequences in dbEST. In addition, experimental confirmation of the 5' region encoding the mouse N-terminal sequence was made from partial cDNAs obtained from the IMAGE consortium. The *C. elegans* sequence was assembled from two partial cDNA clones (GenBank<sup>TM</sup> dbEST accession numbers C10051 and C08344) that are identical for an 81-base-pair region of overlap and encode the apparently complete reading frame shown. The partial amino acid sequence of the homolog from the filarial nematode, *B. malayi*, was translated from a single partial cDNA (GenBank<sup>TM</sup> dbEST accession number AA257328). Two rice partial cDNAs (GenBank<sup>TM</sup> dbEST accession numbers D47693, D47819) covered the translated region shown (in addition, shorter segments of similarity to the human sequence were noted in translations further downstream, but these were in error-prone regions of mismatch between the two ESTs and are not shown). All pairwise alignments were strongly significant, as shown by TBLASTX-2 (Washington University gapped blast, February 1997 release obtained from ftp://blast.wustl.edu/blast/executables). Typical EST pairs gave amino acid gapped E (expect) values (BLOSUM 62 matrix), using the sum statistics of Altschul and Gish (22) as follows (with the highest HSP score appended in parentheses): human/mouse,  $2 \times 10^{-35}$  (717); human/*C. elegans*,  $2 \times 10^{-20}$  (252); human/*B. malayi*,  $8 \times 10^{-12}$  (228); *C. elegans/B. malayi*,  $8 \times 10^{-21}$  (257); human/rice (including multiple short matches for scoring purposes),  $1 \times 10^{-2}$  (82).

encysteine-containing proteins in which the function is established, the selenocysteine residue is located at the active center, and it is essential for catalytic activity of the selenoenzyme (1, 2).

In addition to partial cDNA sequences summarized in Fig. 3, we have detected additional homologous human and mouse cDNAs relatively distantly related to the sequence for the human 15-kDa protein. These encode hypothetical proteins that have in-frame TGA in the position corresponding to the TGA encoding selenocysteine in the 15-kDa selenoprotein gene. The sequences of human and mouse genes for this protein were recently experimentally verified.<sup>2</sup> These observations suggest that the 15-kDa selenoprotein establishes a new class of eukaryotic selenium-containing proteins.

**Gene Expression**—Approximately 120 partial cDNA sequences in dbEST were found to match the human 15-kDa protein DNA sequence (within experimental error or expected frequencies of natural polymorphism). This sampling represents a sufficient abundance of independent clones to reveal the approximate tissue distribution of expression of this relatively highly expressed gene (expression as mRNA). cDNA libraries from 32 different adult, fetal, or embryonic tissues or organs were represented in this set of sequences. Table II shows the ranked incidence of these clones in tissues and organs for which at least one library has two or more independent 15-kDa protein cDNAs in dbEST.

Clearly, the 15-kDa protein gene exhibits a very broad spectrum of moderate expression in many tissues, and significantly higher levels of mRNA are shown by thyroid, parathyroid tumor, prostate, and pre-cancerous prostate cells. Expression estimates from dbEST library frequencies should be considered to be only semi-quantitative, considering that some libraries are normalized and variable levels of tissue contamination may exist. More quantitative representative estimates are given by the stringent CGAP (Cancer Gene Anatomy Project) libraries (25) prepared from small numbers of laser-microdissected cells, for example the pre-cancerous prostate library CGAP\_Pr2 (Ref. 26; Table II). Irrespective of the quantitative uncertainties, this

**TABLE II**  
*Incidence of the human 15-kDa protein gene expression*

Incidence of the human 15-kDa protein gene expression is based on the occurrence of two or more independent cDNA clones in dbEST libraries. In addition, 17 libraries from other tissues, including 3 distinct embryo libraries, contained only a single 15-kDa protein cDNA clone and are not tabulated here. For some clones, both 5' and 3' EST sequences are present in dbEST; these count as only a single cDNA in these calculations.

Library	Incidence per 10,000 ESTs (numbers/library size)
Thyroid	19.9 (4/2014)
Parathyroid tumor (Soares NbHPA)	18.4 (12/6511)
Prostate pre-cancerous cells (CGAP_Pr2)	15.4 (3/1945)
Prostate	11.2 (2/1792)
Fetal lung (Soares NbHL19W)	9.8 (9/9145)
Colon carcinoma (3 libraries)	5.6 (2/3358, 1/2791, 1/956)
Aorta	4.4 (2/4595)
Fetal retina (Stratagene 937202)	4.3 (2/4610)
Jurkat T-cells (2 libraries)	4.3 (2/3534, 1/3420)
Retina (2 libraries)	4.1 (3/8915, 2/3368)
Neuroepithelium (Stratagene 937231)	3.7 (2/5385)
Colon (Stratagene 937204)	3.3 (3/8974)
Testis (Soares NHT)	2.9 (4/13657)
Fetal heart (Soares NbHH19W)	2.3 (6/25708)
Germinal B-cells (CGAP_GCB1)	1.0 (2/19194)

large body of partial cDNA sequence data strongly demonstrates that the 15-kDa protein gene is expressed in a wide range of tissues, with increased levels of mRNA in the thyroid, parathyroid, and prostate-derived cells. Although in this report we describe the human 15-kDa protein gene expression as the expression of mRNA inferred from data base analysis, we detected the mouse analog of this human selenoprotein in immunoblot assays in prostate, heart, kidney, spleen, liver, and other mouse organs, with the highest level observed in prostate, suggesting the expression of both mRNA and the selenoprotein in many tissues and cell lines.<sup>2</sup>

**Selenocysteine Insertion Sequence Element**—Studies of the mechanism of selenocysteine incorporation into several eukaryotic selenoproteins have implicated related stem-loop structures, located in the mRNA 3'-UTR, as essential for selenocysteine insertion into proteins at a UGA codon in the

<sup>2</sup> V. Gladyshev and D. Hatfield, unpublished data.

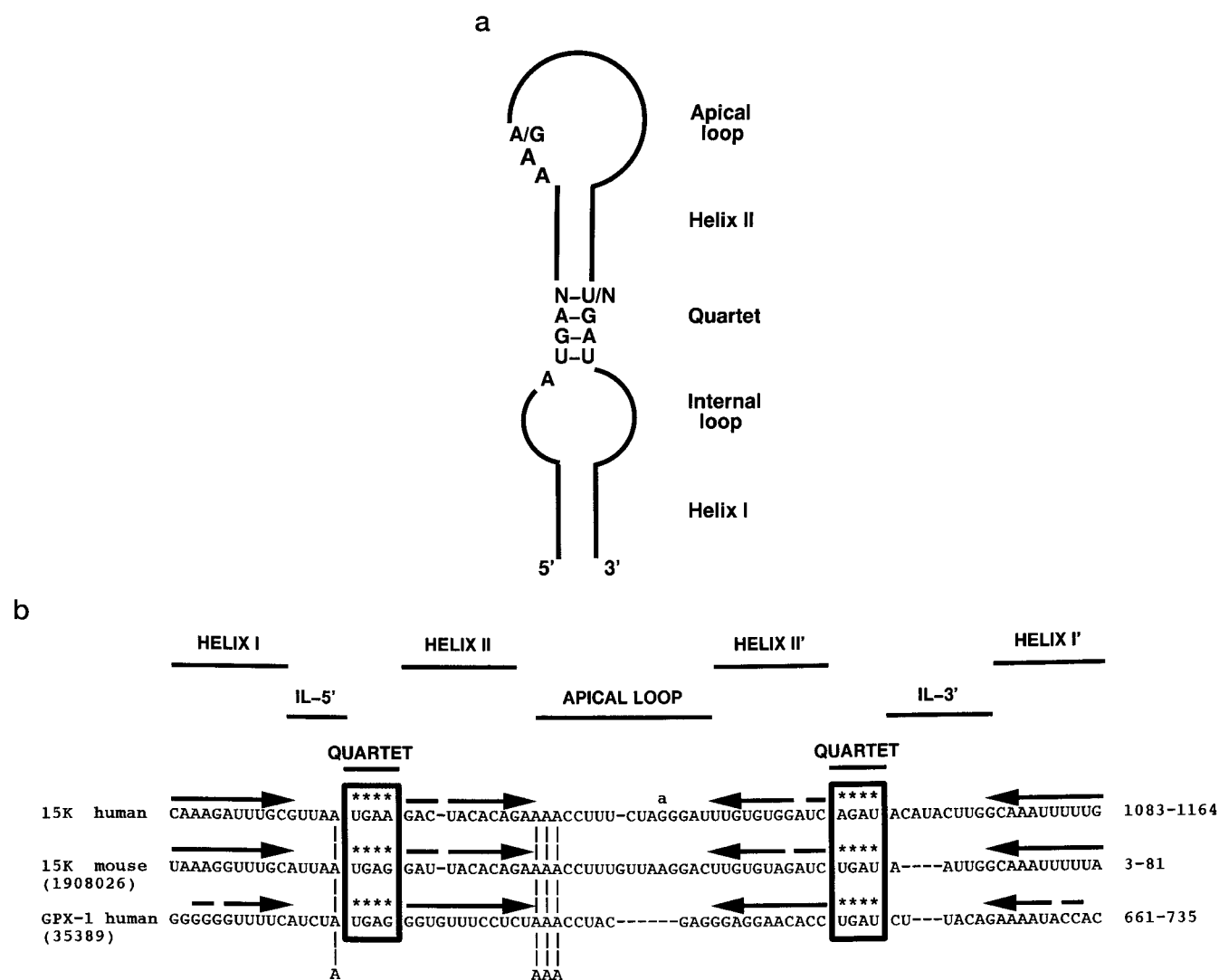


FIG. 4. **SECIS elements.** *a*, general features of eukaryotic SECIS elements used to identify a matching element in the 3'-UTRs of the mRNAs encoding human and mouse 15-kDa selenoproteins (see text). The helical segment of four non-Watson-Crick base pairs is labeled *Quartet*. *b*, alignment of the predicted SECIS elements of the human and mouse mRNAs encoding the 15-kDa protein with a typical experimentally verified example (human GPX-1). In helical stems, single base bulges or mismatches are shown by gaps in the arrows. A lowercase *a* residue above the human apical loop sequence indicates a polymorphism at position 1125.

coding sequence. The general structural features of this SECIS (selenocysteine insertion sequence) element have been deduced previously (3, 27), based on chemical probe experiments and sequence alignments, as summarized in Fig. 4*a*.

To locate potential SECIS elements in the 15-kDa protein mRNAs, the human and mouse cDNAs were searched for sequences meeting the following constraints (see Fig. 4*a*): *Helix I*, at least 4 base pairs; *Internal loop*, 3–9 nucleotides; *Quartet* (the non-Watson-Crick base-paired motif, UGAN (following A in Internal loop) . . . NGAN (following the downstream strand of Helix II); *Helix II*, 9–15 standard base pairs extending the Quartet; *Apical loop*, 10–20 nucleotides starting with AA(A/G). Single base mismatches or bulges were allowed within helices longer than 6 base pairs.

Sequences meeting these stringent criteria were found in both the human and mouse 3'-UTRs, ending approximately 60 nucleotides upstream of the poly(A) addition signal sequence (Fig. 2). Fig. 4*b* shows these human and mouse sequences aligned with the canonical SECIS element (3, 27) of the human glutathione peroxidase 1 (GPX-1) mRNA 3'-UTR. The 15-kDa protein mRNAs exhibit all the features known to be necessary in other eukaryotic selenoprotein mRNAs to promote seleno-

cysteine insertion. It was also noted from the alignments of multiple human sequences from dbEST that a probable G/A substitution polymorphism occurred at an Apical loop nucleotide (position 1125, Fig. 4*b*). Examination of the primary sequencer trace data (openly available on the Washington University server, [http://genome.wustl.edu/est/est\\_general/trace\\_intro.html](http://genome.wustl.edu/est/est_general/trace_intro.html)) confirmed that authentic G or A bases occur at this position and are not sequencing errors. An additional site of substitution (C/T) polymorphism was observed at the 811 position of the human 15-kDa protein cDNA sequence (not shown).

**Chromosomal Localization of the Gene for the 15-kDa Selenoprotein**—Computer analyses revealed the UNIGENE cluster of ESTs (28), corresponding to the 15-kDa human selenoprotein, maps to human chromosome 1, at the position 117–123 cM on the human transcript gene map, corresponding approximately to 1p31 (29).

**Concluding Remarks**—We have detected a new selenoprotein in human T cells designated as the 15-kDa protein. Although the function of this protein is not known, certain selenoproteins have been associated with a chemopreventive role in cancer (2). Interestingly, there are certain correlations be-

tween the occurrence of the newly described 15-kDa protein in cancer tissue and the effect of selenium on cancer that are worthy of discussion. For example, recent studies have shown that supplementation of the diet with selenium in human clinical trials resulted in 63% reduction in prostate cancer and, to a lesser extent, in the reduction of colon and lung cancers (30). Interestingly, the 15-kDa protein is highly expressed in prostate tissue (see Table II) and thus some of the protective properties of selenium in prostate cancer may be mediated through this protein. In addition, our preliminary data suggest that the levels of the 15-kDa protein and its mRNA are decreased in some cancers (e.g. liver of c-myc/TGF $\alpha$  transgenic mice<sup>2</sup>). Studies to determine whether the 15-kDa protein may play a role in delaying the progression of certain cancers including prostate cancer are in progress.

**Acknowledgments**—The initial part of this work, including the protein purification, was done while V. N. Gladyshev was working in the laboratory of Dr. Thressa Stadtman (Laboratory of Biochemistry, NHLBI, National Institutes of Health). We thank her for support during the course of this work as well as for helpful suggestions during the manuscript writing. We also thank Harvard Microchem for highly professional protein microchemistry services, Dr. Rodney Levine for helpful discussions and MALDI analysis, Dr. Francesca Zappacosta for performing electrospray mass-spectrometry, Dr. Valery Bliskovsky for help with initial computer analyses, and Drs. Valentina Factor and Michael Jensen for helpful discussions and advice.

## REFERENCES

- Stadtman, T. C. (1996) *Annu. Rev. Biochem.* **65**, 83–100
- Hatfield, D. L., Gladyshev, V. N., Park, J., Park, S. I., Chittum, H. S., Baek, H. J., Carlson, B. A., Yang, E. S., Moustafa, M. E. & Lee, B. J. (1998) *Compr. Nat. Prod. Chem.*, in press
- Low, S. & Berry, M. (1996) *Trends Biochem. Sci.* **21**, 203–207
- Sunde, R. A. (1994) in *Selenium in Biology in Human Health* (Burk, R. F., ed) pp. 146–177, Springer-Verlag New York Inc., New York
- Ursini, F., Maiorino, M., Brigelius-Flohe, R., Aumann, K. D., Roveri, A., Schomburg, D. & Flohe, L. (1995) *Methods Enzymol.* **252**, 38–53
- Berry, M. J., Banu, L. & Larsen, P. R. (1991) *Nature* **349**, 438–440
- Croteau, H. W., Davey, J. C., Galton, V. A. & St. Germain, D. L. (1996) *J. Clin. Invest.* **98**, 405–417
- St. Germain, D. L., Schwartzman, R. A., Croteau, W., Kanamori, A., Wang, Z., Brown, D. D. & Galton, V. A. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 7767–7771
- Gasdaska, P. Y., Gasdaska, J. R., Cochran, S. & Powis, G. (1995) *FEBS Lett.* **373**, 5–9
- 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. & Zlotnik, A. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 15086–15091
- Hill, K. E. & Burk, R. F. (1994) in *Selenium in Biology in Human Health* (Burk, R. F., ed) pp. 117–132, Springer-Verlag New York Inc., New York
- Vendeland, S. C., Beilstein, M. A., Yeh, J.-Y., Ream, W. & Whanger, P. D. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 8749–8753
- Tamura, T. & Stadtman, T. C. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 1006–1011
- Gladyshev, V. N., Jeang, K.-T. & Stadtman, T. C. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 6146–6151
- Arscott, L. D., Gromer, S., Schirmer, R. H., Becker, K. & Williams, C. H. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 3621–3626
- Gladyshev, V. N., Jeang, K.-T., Hatfield, D. L. & Stadtman, T. C. (1997) *FASEB J.* **11**, 3 (A235)
- Nagata, K., Ohtani, K., Nakamura, M. & Sugamura, K. (1989) *J. Virol.* **63**, 3220–3226
- Gladyshev, V. N., Khangulov, S. V. & Stadtman, T. C. (1996) *Biochemistry* **35**, 213–223
- Boguski, M. S., Lowe, T. M. J. & Tolstoshev, C. M. (1993) *Nat. Genet.* **4**, 332–333
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410
- Altschul, S. F., Boguski, M. S., Gish, W. & Wootton, J. C. (1994) *Nat. Genet.* **6**, 119–129
- Altschul, S. F. & Gish, W. (1996) *Methods Enzymol.* **266**, 460–480
- Sonnhammer, E. L. & Durbin, R. (1994) *Comput. Appl. Biosci.* **10**, 301–307
- Kozak, M. (1997) *EMBO J.* **16**, 2482–2492
- Strausberg, R. L., Dahl, C. A. & Klausner, R. D. (1997) *Nat. Genet.* **15**, 415–416
- Krizman, D. B., Chuaqui, R. F., Meltzer, P. S., Trent, J. M., Duray, P. H., Linehan, W. M., Liotta, L. A. & Emmert-Buck, M. R. (1996) *Cancer Res.* **56**, 5380–5383
- Walczak, R., Westhof, E., Carbon, P. & Krol, A. (1996) *RNA* **2**, 367–379
- Boguski, M. S. & Schuler, G. D. (1995) *Nat. Genet.* **10**, 369–371
- Schuler, G. D., Boguski, M. S., Stewart, E. A., Stein, L. D., Gyapay, G., Rice, K., White, R. E., Rodriguez-Tome, P., Aggarwal, A., Bajorek, E., Bentolila, S., Birren, B. B., Butler, A., Castle, A. B., Chiannilkulchai, N., Chu, A., Clee, C., Cowles, S., Day, P. J., Dibling, T., Drouot, N., Dunham, I., Duprat, S., East, C., Hudson, T. J., et al. (1996) *Science* **274**, 540–546
- Clark, L. C., Combs, G. F., 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., Park, H. K., Sanders, B. B., Smith, C. L. & Taylor, J. R. (1996) *JAMA* **276**, 1957–1963