Mass Spectrometric Characterization of Full-length Rat Selenoprotein P and Three Isoforms Shortened at the C Terminus

EVIDENCE THAT THREE UGA CODONS IN THE mRNA OPEN READING FRAME HAVE ALTERNATIVE FUNCTIONS OF SPECIFYING SELENOCYSTEINE INSERTION OR TRANSLATION TERMINATION

Shuguang Ma‡‡, Kristina E. Hill‡‡, Richard M. Caprioli‡, and Raymond F. Burk‡‡**

From the □ Mass Spectrometry Research Center, Department of Biochemistry, ‡‡ Division of Gastroenterology, Department of Medicine, and ‡‡ Clinical Nutrition Research Unit, Vanderbilt University School of Medicine, Nashville, Tennessee 37232

Selenoprotein P is an abundant extracellular glycoprotein. Its mRNA contains 10 UGAs in an open reading frame terminated by a UAA. This predicts that full-length selenoprotein P will contain 10 selenocysteine residues. Full-length selenoprotein P and three smaller isoforms that have identical N termini have been demonstrated. Selenoprotein P was purified from rat plasma, and the four isoforms were separated by heparin chromatography and SDS-PAGE. Mass spectrometric peptide analysis of the full-length isoform verified 357 of its 366 predicted amino acid residues, including its C terminus and all 10 selenocysteines. The C termini of the smaller isoforms were characterized by mass spectrometry. The shortened isoforms terminated where the second, third, and seventh selenocysteine residues were predicted to be. This suggests that all isoforms arise from the same mRNA and that the UGAs that specify the second, third, and seventh selenocysteines in full-length selenoprotein P can alternatively serve to terminate translation, producing the shorter isoforms.

More than 15 animal selenoproteins have been characterized, and all of them contain selenocysteine. Selenocysteine incorporation is specified by a UGA in the open reading frame of the mRNA that is accompanied by a “selenocysteine insertion sequence” element in the 3′-untranslated region (1). This is an alternative function of the UGA codon, which usually terminates translation.

Selenoprotein P is an unusual extracellular glycoprotein that has been suggested to serve in oxidant defense (2) and in selenium homeostasis (3). It is synthesized in the liver and secreted into the plasma, where it is relatively abundant (30 μg of peptide/ml plasma) (5). The purified protein contained 7.5 ± 1 atoms of selenium per molecule in the form of selenocysteine (5).

The mRNA of selenoprotein P (deduced from its cDNA) contains 10 UGAs in its open reading frame, implying 10 selenocysteines in the primary structure of the protein (6). Thus, not all the predicted selenium was detected by analysis of the purified protein. This variance might be due to the presence of isoforms of the protein that contain different numbers of selenocysteine residues or to the absence of selenium from some of the predicted selenocysteine sites.

Isoforms of selenoprotein P are present in rat plasma, having been demonstrated by their differential binding to heparin-Sepharose (7). Two isoforms have been purified and characterized by amino acid analysis, conventional peptide sequencing, and C-terminal sequencing (8). Both of these isoforms have the same N-terminal amino acid sequence. One of them was shown to be the full-length protein, terminating at a “hard stop” UAA in the mRNA, and the other was shown to terminate at the predicted position of the second selenocysteine, i.e. the second UGA in the open reading frame of the mRNA (8). Two other isoforms were identified in that study, but they could not be purified in quantities sufficient for conventional amino acid sequencing. Aside from the demonstration that both of them had the same N-terminal amino acid sequence as the two isoforms that had been purified, these two isoforms have not been characterized.

The present study uses mass spectrometry to identify the C-terminal peptides of the two previously uncharacterized isoforms, revealing that they terminate at positions in the sequence where selenocysteine residues are predicted (UGAs). Mass spectrometry was also used to verify the amino acid sequence of the full-length isoform and to confirm that all 10 selenocysteines are present.

EXPERIMENTAL PROCEDURES

Materials—iodoacetamide and diithiothreitol were obtained from Sigma. Sequence-grade trypsin was purchased from Burdick and Jackson (Muskegon, MI). The modified T-281-phenylalanylchromethyl ketone-treated porcine trypsin was obtained from Promega (Madison, WI). The other endoproteinases, Lys-C, Glu-C, and Asp-N, were purchased from Roche Molecular Biochemicals. The deglycosylation kit, containing N-glycosidase F, Endo-O-glycosidase, and sialidase A, was obtained from ProZyme (San Leandro, CA). Mass calibration standards, des-Arg1-bradykinin, neuropeptide Y, bovine insulin, mellitin, trypsinogen, and bovine serum albumin were purchased from Sigma. Matrix materials, CHCA1 and sinapinic acid, were purchased from Aldrich. 75Se-labeled selenite (800 mCi/mg) was purchased from the University of Missouri Research Reactor Facility (Columbia, MO). Other reagents used were analytical grade or better.

Rat Selenoprotein P Preparation—Rat plasma from which selenoprotein P was purified was purchased from Harlan Bioproducts for Science (Indianapolis, IN). Purification was accomplished as described previ-
In-gel Digestion—The protein gel band of interest was excised from the gel and digested with modified trypsin (9, 10). The destained gel slice was washed, crushed, reduced in dithiothreitol, and alkylated with iodoacetamide. The gel pieces were washed, dehydrated in acetonitrile, and swollen in 10 μl of 50 mM ammonium bicarbonate containing trypsin in a concentration of 0.1 μg/μl. After 15 min, additional 50 mM ammonium bicarbonate was added to cover the gel pieces, and digestion was allowed to proceed overnight. After digestion, ~1 μl of the supernatant was removed for MALDI mass spectrometry analysis. The remaining supernatant was removed, and the gel slices were extracted twice with 50 μl of 60% acetonitrile/0.1% trifluoroacetic acid. These extraction mixtures were dried and combined with the supernatant for injection onto a C18 column.

Enzymatic Digestion—In-solution Glu-C digestion of deglycosylated full-length selenoprotein P isoform (peak 2) was performed in 25 mM ammonium bicarbonate. Approximately 10 μg of protein was reduced in 10 μl of 50 mM DTT and alkylated with iodoacetamide, and swollen in 10 μl of 50 mM ammonium bicarbonate containing trypsin in a concentration of 0.1 μg/μl. After 15 min, additional 50 mM ammonium bicarbonate was added to cover the gel pieces, and digestion was allowed to proceed overnight. After digestion, ~1 μl of the supernatant was removed for MALDI mass spectrometry analysis. The remaining supernatant was removed, and the gel slices were extracted twice with 50 μl of 60% acetonitrile/0.1% trifluoroacetic acid. These extraction mixtures were dried and combined with the supernatant for isolation onto a C18 column.

Reverse-phase HPLC—The proteolytic peptide mixtures were separated on a Vydac microbore C18 column (2.1 x 250 mm) (Vydac, Hesperia, CA) using an HP 1100 HPLC system (Hewlett-Packard Co., Wilmington, DE). The separation was monitored with a UV detector at 214 nm. Solvent systems of A (0.1% trifluoroacetic acid) and B (94.9% acetonitrile/5.1% trifluoroacetic acid) were used. The peptides were eluted from the column with a flow rate of 0.2 ml/min using the following gradient program: 5% solvent B, 10 min, isocratic; 5–60% solvent B over 50 min, linear gradient; 60% solvent B, 10 min, isocratic; 60–80% solvent B in 10 min, linear gradient; 80% solvent B, 10 min, isocratic; then to 5% solvent B in 5 min, linear gradient. UV-monitored peptide peaks were manually collected in Eppendorf safe-lock polypropylene tubes. HPLC fractions were lyophilized dry and reconstituted in 40 μl of 50% acetonitrile/50% H2O/0.1% trifluoroacetic acid for mass spectrometry analysis.

Mass Spectrometry—MALDI mass spectra were obtained using a Perseptive DE-STR MALDI-TOF mass spectrometer (Applied Biosystems, Foster City, CA) equipped with a 337-nm nitrogen laser. The instrument was operated in the linear mode under optimized delayed extraction conditions for peptide and protein analysis. The selenium-containing peptides were analyzed in the reflector mode to obtain high resolution to identify the selenium isotope distribution. Mass calibration was accomplished using des-Arg1-bradykinin (MW 903.46) and bovine insulin (MW 5733.58) for peptide analysis and bovine trypsinogen (MW 23981) and bovine serum albumin (MW 66430) for protein analysis in the linear mode. In reflector mode, neurotensin (MW 1671.900) and melittin (MW 2844.754) were used as calibration compounds. A matrix of CHCA, prepared at 10 mg/ml in 50% acetonitrile/49.9% H2O/0.1% trifluoroacetic acid, was used for peptide analysis, and sinapinic acid at 10 mg/ml in 50% acetonitrile/49.9% H2O/0.1% trifluoroacetic acid was used as the matrix for protein analysis. The samples were prepared by dried-droplet method on a stainless steel MALDI plate.

Nanoelectrospray was performed on a Finnigan LCQ (Finnigan, San Jose, CA) with a nanospray ion source (Prolec A/S, Odense, Denmark). The sample of 2 μl was loaded into the metal-coated glass capillary (Prolec A/S, Odense, Denmark). The capillary was positioned about 1 mm from the heated capillary. The spray voltage was set at 800 V. The heated capillary was kept at 150 °C. The capillary voltage was set at 45 V, and the tube lens was offset at ~10 V. The other parameters for ion optics were tuned to obtain the most intensity for the ion of interest. In MS/MS mode, the precursor ion of interest was isolated and fragmented at 25–55% collision energy, depending on the nature of the peptide. Theoretical isotope peaks were calculated using software that was supplied with the Perseptive DE-STR MALDI-TOF mass spectrometer.

RESULTS

Peptide Mapping—Mass spectrometric peptide mapping has been an important analytical tool for verifying protein sequence and identifying proteins by database search. This technique involves the digestion of the protein by specific endoproteases or chemicals. The resulting peptide mixture is subjected to analysis by mass spectrometry, for example MALDI-TOF MS. The comparison of the set of peptide molecular ion masses observed with the expected digestion fragment masses generally yields a sequence coverage of 60–90%. Using this technique, we have verified the sequence of selenoprotein P, originally deduced from its cDNA. Fig. 1 shows peptide maps of the full-length selenoprotein P isoform produced by MALDI-TOF MS after trypsin and Glu-C digestion. The masses of peaks in the spectra were compared with theoretical masses of predicted peptides (Tables S1 and S2 in the on-line supplement), and the peaks in the spectra that could be identified were labeled with the residue numbers of the peptides. Only three stretches of three amino acid residues each were not verified in this study. The sequence coverage was 97.5%, and it is shown in Fig. S1 in the on-line supplement.

Identification of Selenium in Peptides—Selenocysteine was confirmed to be present in the peptides predicted to contain it by showing that the determined peptide mass matched the predicted peptide mass (which included the mass of selenocysteine). From the peptide mapping, the 10 selenocysteines located at residues 40, 245, 263, 304, 316, 338, 352, 354, 361, and 363, respectively, have been identified. Table I summarizes the experiments that identified each of the selenocysteine residues in the full-length isoform.

In addition, selenium was identified directly in those peptides by its isotope pattern. Selenium has a unique isotope distribution of 74Se (0.89%), 76Se (9.37%), 77Se (7.63%), 78Se (23.77%), 80Se (49.61%), and 82Se (8.73%). Fig. 2 shows the isotopic detail of the mass peak of a peptide corresponding to residues 239–260 of selenoprotein P. Fig. 2a shows the expanded mass spectrum of this selenopeptide from MALDI-TOF mass spectrometry operated in reflector mode. Fig. 2b shows the theoretical expanded peak of this selenopeptide that was generated by a computer program. Fig. 2c shows the theoretical peak of the same peptide with a sulfur atom replacing the selenium atom. In Fig. 2, a and b are almost identical, whereas Fig. 2c has a different mass pattern. This allows the conclusion that the experimental peptide contains one selenium atom.

This selenopeptide was further studied by fragmenting the doubly charged peptide ion at m/z 1218.0 to verify the peptide sequence (Fig. 9). The inset shows the zoom scan of the parent ion, which indicates the charge state of +2 and its content of one selenium atom. With the arginine at the C-terminal as a favorable charge-carrying site, the abundant y ion series was observed. It is of interest to note that the b ion series in this case was a b17 ([b17-NH2]2) ion series due to the glutamine located at the N terminus. The pattern of fragments confirms the predicted sequence of the peptide.

Expanded peaks of selenopeptides containing two, three, and four selenium atoms have characteristic isotope patterns (Fig. S2 in the on-line supplement). The isotope pattern of each peak closely approximated the pattern predicted by a computer program (predicted peaks containing two, three, and four selenium atoms, shown in Fig. S3 in the on-line supplement). All the selenopeptides listed in Table I were further studied using...
MS/MS to verify the predicted amino acid sequence (Fig. 3, as well as Figs. S4–S8 in the on-line supplement). The pattern of fragments confirmed the predicted sequence of each of the selenopeptides.

**Masses of Isoforms**—Purified selenoprotein P can be separated into three peaks using heparin chromatography (7). Two of the peaks, 1b and 2, contain one isoform of the selenoprotein each. The other peak, 1a, contains two isoforms, and these have not been characterized beyond the demonstration that they have the same N terminus as the other two isoforms.

MALDI-TOF mass spectra of the three peaks from the heparin column were obtained (Fig. S9 in the on-line supplement). Peak 2 is the full-length isoform and has an average mass of 50,474 Da. Peak 1b is the short isoform and has an average mass of 36,136 Da. Peak 1a demonstrates two peaks with masses of 38,482 Da and 48,842 Da, which are both intermediate in size between the other two isoforms. This indicates that peak 1a contains two additional isoforms. All four mass peaks are broad, which is consistent with microheterogeneity, that is presumably the result of their glycosylation.

**Determination of C Termini of Isoforms**—Scheme 1 depicts the strategy for determining the C termini of the isoforms. Each isoform was digested with a particular protease, and the predicted C-terminal peptide was sought using mass spectrometry. After identification of the peak by MALDI-TOF mass spectrometry, the amino acid sequence was determined by MS/MS to confirm that it was indeed the predicted C-terminal peptide.

<table>
<thead>
<tr>
<th>Partial sequence</th>
<th>Enzyme</th>
<th>Number of selenocysteine residues</th>
<th>Residue position(s) of selenocysteine</th>
<th>Peptide sequence</th>
<th>[M + H]_{exp}</th>
<th>[M + H]_{calc}</th>
</tr>
</thead>
<tbody>
<tr>
<td>28–51</td>
<td>Glu-C</td>
<td>1</td>
<td>40</td>
<td>GTTVVALLQASU</td>
<td>2714.38</td>
<td>2714.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>YLCLLQASRLE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>239–260</td>
<td>Trypsin</td>
<td>1</td>
<td>245</td>
<td>QHILESUDMGASE</td>
<td>2435.08</td>
<td>2435.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GLQLSLAQR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>252–278</td>
<td>Glu-C</td>
<td>1</td>
<td>263</td>
<td>GLQSLAQRKLUR</td>
<td>3292.62</td>
<td>3292.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>QGCINQLCKLSEE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>298–324</td>
<td>Trypsin</td>
<td>2</td>
<td>304, 316</td>
<td>SGSATUQCAENLP</td>
<td>3113.28</td>
<td>3113.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SLCSUQGLFAEEL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>333–357</td>
<td>Trypsin</td>
<td>3</td>
<td>338, 352, 354</td>
<td>SSPAUIHSSQVSP</td>
<td>2879.02</td>
<td>2878.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EASPNUSUNNK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>348–366</td>
<td>Glu-C</td>
<td>4</td>
<td>352, 354, 361, 363</td>
<td>ASPNUSUNNKTKKK</td>
<td>2487.84</td>
<td>2487.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>UKUNLN</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a* All the cysteine and selenocysteine residues were modified by iodoacetamide (U = selenocysteine).

*b* This m/z number is for the isotope peak of the peptide that has the greatest abundance.

**Table I**

*Summary of the identification of the 10 selenocysteine residues using MALDI-TOF mass spectrometry operated in the reflector mode*

Mass accuracy is <0.08 Da.
the third UGA in the mRNA. To confirm this, peak 1a was subjected to endoproteinase Asp-N digestion. The C-terminal peptide (residues 246–262) was isolated by HPLC and was subjected to MS/MS experiments by fragmenting the doubly charged ion with m/z 909.50 to verify the sequence, as shown in Fig. 4. These results confirm that this isoform indeed terminates at the third UGA codon.

The longer isoform in peak 1a was predicted to terminate at the seventh or eighth UGA based on its mass (Fig. S9 of the on-line supplement). However, when a trypsin digest of it was studied by MALDI-TOF MS (not shown), neither predicted C-terminal peptide could be identified. Therefore the digest was subjected to HPLC, and each chromatographic peak was evaluated by MALDI-TOF MS. Fig. 5 shows the spectrum of one HPLC peak. In Fig. 5a, a complex pattern is seen with the lowest mass being 2021.6 followed by signals at mass increments of 203, 162, 291, and another 291. These mass increments correspond to the masses of N-acetylgalactosamine, galactose, and N-glycolyneraminic acid, respectively. This indicates that this peptide is an O-linked glycopeptide. The glycopeptide was further treated with endo-O-glycosidase and sialidase A. The glycopeptide signals disappeared, and the intensity of the peptide peak at 2021.60 Da increased (Fig. 5b). This is the predicted mass of the C-terminal peptide that terminates at the seventh UGA. MS/MS verification of the sequence of the deglycosylated peptide was obtained (S11 of the on-line supplement). The inset of that figure shows the zoom scan of the parent ion (+2), which indicates that the peptide contains one selenocysteine residue (SeCys338). These results indicate that this peptide is variably O-glycosylated and that the isoform terminates at the seventh UGA.

Formal sequencing of the shorter three isoforms was not un-
The work undertaken in this study, but some of their peptides were identified by mass (data not shown) in the experiment performed to determine their C termini. For selenoprotein P1, 59% of the predicted amino acid sequence (residues 1-18, 86-120, 124-152, 178-232) was identified by mass. For selenoprotein P2 and selenoprotein P6, respectively, 42% (residues 55-120, 178-193, 233-260) and 62% (residues 57-121, 124-145, 178-260, 292-324) of the predicted amino acid residues were identified.

Thus, four isoforms of selenoprotein P have been identified. They share the same N terminus and appear to be derived from the same amino acid sequence. One isoform is full-length, but the other three terminate at positions predicted to be occupied by selenocysteine residues (UGA codons in the open reading frame of the mRNA).

**DISCUSSION**

Selenoprotein P is the only selenoprotein identified so far that contains more than one selenium atom. The presence of 10 UGAs in the open reading frame of rat selenoprotein P mRNA implies that the full-length protein contains 10 selenocysteine residues. Because there is no precedent for multiple selenocysteines in a protein and it is at least conceivable that a partic-
The amino acid sequence of selenoprotein P10 was demonstrated to allow further characterization of all four isoforms. The reading frame, sites where selenocysteine is present in selenoprotein P, and contains 10 selenium atoms (Scheme 2). The other three isoforms were shown to have the expected amino acid composition and appropriate N- and C-terminal amino acid sequences.

Moreover, amino acid sequencing of a mixture of all two isoforms was shown to have the expected amino acid composition and appropriate N- and C-terminal amino acid sequences. Furthermore, amino acid sequencing of a mixture of all isoforms by Edman chemistry yielded only sequences predicted from the cDNA (8).

The antibody-based methods for purification and measurement of selenoprotein P do not allow distinction among the isoforms. This suggests that the available antibodies are directed toward epitopes in the N-terminal region of the protein. It will be important to develop methods to detect and measure individual isoforms of selenoprotein P. That will allow determination of the cells that produce each isoform and assessment of the physiological function of each isoform. Correlation of physiological studies (4, 14, 15) with the molecular properties of the isoforms of selenoprotein P should provide insights into their functions.

Acknowledgments—We thank Robert J. Hondal, Ph.D., for performing preliminary experiments; Ansley S. Harpe, Todd B. Peterson, and Troy E. Morris for preparation of selenoprotein P isoforms; and Eric F. Howard of the protein chemistry core for Edman sequencing.

REFERENCES


Mass Spectrometric Characterization of Full-length Rat Selenoprotein P and Three Isoforms Shortened at the C Terminus: EVIDENCE THAT THREE UGA CODONS IN THE mRNA OPEN READING FRAME HAVE ALTERNATIVE FUNCTIONS OF SPECIFYING SELENOCYSTEINE INSERTION OR TRANSLATION TERMINATION

Shuguang Ma, Kristina E. Hill, Richard M. Caprioli and Raymond F. Burk

doi: 10.1074/jbc.M111462200 originally published online January 30, 2002

Access the most updated version of this article at doi: 10.1074/jbc.M111462200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2002/04/09/277.15.12749.DC1

This article cites 15 references, 5 of which can be accessed free at
http://www.jbc.org/content/277/15/12749.full.html#ref-list-1