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

Received for publication, November 30, 2001, and in revised form, January 15, 2002
Published, JBC Papers in Press, January 30, 2002, DOI 10.1074/jbc.M111462200

Selenoprotein P is an unusual 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 transport (3, 4). It was originally purified from rat 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 dithiothreitol were obtained from Sigma. Sequence-grade trifluoroacetic acid was purchased from Burdick and Jackson (Muskegon, MI). The modified T-tryosyl L-phenylalanineylchloromethylketone-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 deoxyribosylation 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, melittin, tryptsinogen, 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
ously, using a column prepared with the monoclonal antibody 8F11 (8). A heparin-Sepharose column was used to separate selenoprotein P isoform peaks from the purified protein preparation as was done previously (7). These peaks, 1a, 1b, and 2, were used for further characterization. Each separated peak was deglycosylated using N-glycosidase F and then digested with trypsin followed by in-gel digestion. The in-gel digestion was performed on the deglycosylated selenoprotein P isoform peaks as described previously (7).

**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 iodoacetic acid. The gel pieces were washed, dehydrated in acetone, 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 ammonium bicarbonate containing iodoacetic acid and incubated with iodoacetic acid for 30 min. Approximately 0.3 μg of Glu-C was added before incubation at 37 °C for 18 h. The digestion mixtures were concentrated with 1 μl of 5% trifluoroacetic acid, and peptides were separated on a reverse-phase C18 column.

The C-terminal peptide of each isoform of selenoprotein P was isolated from a proteolytic digest with different endopeptidases. Approximately 10 μg of each selenoprotein P isoform peak was reduced and alkylated before enzymatic digestion as described above. Peak 1a was digested with 0.2 μg of Asp-N in phosphate buffer (pH 8.0) for 18 h, and another sample of it was digested with 0.2 μg of modified trypsin in 50 mM ammonium bicarbonate for 18 h. Peak 1b was digested with 0.2 μg of Lys-C in 50 mM Tris-HCl buffer (pH 8.5), and peak 2 was digested with 0.25 μg of Glu-C in 25 mM ammonium bicarbonate. The digestion mixtures were neutralized with 1 μl of 5% trifluoroacetic acid and separated on a C18 column to isolate the C-terminal peptides.

**Reverse-phase HPLC**—The proteolytic peptide mixtures were separated on a Vydac microbore C18 column (2.1 × 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 peptide peaks 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 microcentrifuge tubes. HPLC fractions were lyophilized dry and stored at −20 °C. HPLC fractions 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 cytochrome c (MW 12356) for protein analysis. The selenium-containing peptides listed in Table I were further studied using mass spectrometry. 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 identified 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, selenocysteine was identified directly in those peptides by its isotope pattern. Selenium has a unique isotope distribution of $^{74}$Se (0.89%), $^{76}$Se (9.37%), $^{77}$Se (7.63%), $^{75}$Se (23.77%), $^{80}$Se (49.61%), and $^{82}$Se (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 peak 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 b-17 ([b-NH$_3$]$_7$) 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 full-length isoform of selenoprotein P was reduced, alkylated, and deglycosylated. Aliquot was subjected to SDS-PAGE, and the protein band was cut out and treated with trypsin before mass spectrometric analysis. b, aliquot was treated with Glu-C before mass spectrometric analysis.

**TABLE I**

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

| Partial sequence | Enzyme | Number of selenocysteine residues | Residue position(s) of selenocysteine | Peptide sequence<sup>a</sup> | |<sup>b</sup> | Mass accuracy<sup>b</sup> |
|-----------------|--------|----------------------------------|--------------------------------------|-----------------------------|---|-------------------|
| 28–51           | Glu-C  | 1                                | 40                                   | GTVTVALQASU                  |    | 0.08 Da.          |
|                 |        |                                   |                                      | YLCQQASRLE                   |    |                   |
| 239–260         | Trypsin| 1                                | 245                                  | QGHLESDMGASE                 |    |                   |
| 252–278         | Glu-C  | 1                                | 263                                  | GLQLSAA                     |    |                   |
| 298–324         | Trypsin| 2                                | 304, 316                             | SGAIAFEQCAENLP               |    |                   |
| 333–357         | Trypsin| 3                                | 338, 352, 354                        | SPPAAUVQHOOKST               |    |                   |
| 348–366         | Glu-C  | 4                                | 352, 354, 361, 363                   | ASPNUSUNKTKK                 |    |                   |

<sup>a</sup> All the cysteine and selenocysteine residues were modified by iodoacetamide (U = selenocysteine).

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

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 ± 200 Da. Peak 1b is the short isoform and has an average mass of 36,136 ± 200 Da. Peak 1a demonstrates two peaks with masses of 38,482 ± 200 Da and 48,842 ± 200 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.

The C-terminal peptide (residues 235–244) of the short isoform was identified in a Lys-C digestion, and its amino acid sequence was verified by MS/MS (Fig. S10 in the on-line supplement). A similar experiment (Fig. S8 in the on-line supplement) confirmed that the isoform in peak 2 was full-length.

Based on the mass of the protein measured by MALDI-TOF MS, the shorter isoform in peak 1a appeared to terminate at
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 \( \text{m/z} \) 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 2021.60 Da 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-glycolyneuraminic 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, which is doubly charged and contains one selenium atom.

Formal sequencing of the shorter three isoforms was not un-
dertaken 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 19–18, 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 57–120, 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 isoforms by Edman chemistry yielded only sequences predicted from the cDNA (8). 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. Two corresponding isoforms have the same amino acid sequence and differ from one another by terminating at different sites within that sequence (Scheme 2). A previous study supports this conclusion for selenoprotein P10 (Se-P10) because it terminates at the hard stop UAA in the mRNA and contains 10 selenium atoms (Scheme 2). The other three isoforms terminate at positions of UGA codons in the open reading frame, sites where selenocysteine is present in selenoprotein P10. Respectively, the shorter isoforms terminate at the second, third, and seventh UGAs. Evidence suggests that all four isoforms share the same amino acid sequence and differ from one another by terminating at different sites within that sequence (Scheme 2). A previous study supports this conclusion for selenoprotein P10 and selenoprotein P10 (8). Each of those two isoforms was 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 isoforms by Edman chemistry yielded only sequences predicted by the cDNA (8).

In the present study, the high sensitivity of mass spectrometry allowed further characterization of all four isoforms. The amino acid sequence of selenoprotein P10 was demonstrated to match the predicted sequence with only nine amino acid residues not being verified (Fig. S1 of the on-line supplement). Of those nine unverified residues, five had been verified in the previous study (8), so only four (residues 121, 122, 325, and 326) of the total of 366 predicted remain unverified by direct analysis of the protein. Significant portions of the sequences of the three shorter isoforms were verified also. These findings are consistent with the hypothesis that all four isoforms arise from the same mRNA.

The dual nature of UGA as a termination codon and a sense codon has been studied before by several groups (11–13). Using cell lines transfected with selenoprotein cDNA constructs, these groups have shown that the nucleotide context of the UGA codon affects its read-through efficiency. The two nucleotides downstream and the codon upstream from the UGA have significant effects. In the systems studied, a purine in the +1 position (downstream) promotes termination (11). The second, seventh, eighth, ninth, and tenth UGAs in selenoprotein P mRNA are followed by purines. A recent report has shown that the +2 nucleotide can also increase termination if a +1 pyrimidine is followed by a +2 purine (13). In selenoprotein P, the first, third, fourth, fifth, and sixth UGA codons are followed by such a pyrimidine-purine combination. Thus, all 10 UGA codons of selenoprotein P mRNA are present in downstream termination contexts. Changes in the upstream codon have generally caused increased termination (12), and that is not applicable to the present discussion. Thus, current knowledge of nucleotide context cannot explain the production of the isoforms found in this study. It seems likely that trans-acting factors that can be produced independent of the mRNA will be responsible for the termination leading to isoform production. Selenoprotein P provides an unique opportunity to study in vivo factors involved in the alternative functions of the UGA codon.

It remains a theoretical possibility that the shorter isoforms are produced from the full-length protein by proteolysis. If that were the case, however, there would have to be a protease that cleaves the protein at specific selenocysteine residues. Because no such protease is known, it seems more likely that termination of translation is responsible for production of the isoforms. This question remains open, however.

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 physiology 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. Tharpe, 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

1. Berry, M. J., Banu, L., Chen, Y., Mandel, S. J., Kieffer, J. D., Harney, J. W., and Larsen, P. R. (1993) Nature 363, 273–276.
2. Burk, R. F., Hill, K. E., Awad, J. A., Morrow, J. D., Kato, T., Cockell, K. A., and Lyons, P. R. (1995) Hepatology 21, 561–569.
3. Motaenbocker, M. A., and Tappe, A. L. (1992) Biochim. Biophys. Acta 1139, 147–153.
4. Burk, R. F., Hill, K. E., Read, R., and Belloe, T. (1991) Am. J. Physiol. 261, E26–E30.
5. Read, R., Belloe, T., Yang, J.-G., Hill, K. E., Palmer, I. S., and Burk, R. F. (1990) J. Biol. Chem. 265, 17899–17905.
6. Hill, K. E., Lloyd, R. S., Yang, J.-G., Read, R., and Burk, R. F. (1991) J. Biol. Chem. 266, 10050–10053.
7. Chittum, H. S., Himone, S., Hill, K. E., and Burk, R. F. (1996) Arch. Biochem. Biophys. 325, 124–128.
8. Himone, S., Chittum, H. S., and Burk, R. F. (1996) J. Biol. Chem. 271, 15769–15775.
9. Rosenfeld, J., Capdevielle, J., Guillermet, J. C., and Ferrara, P. (1992) Anal. Biochem. 203, 173–179.
10. Heldman, U., Wernstedt, C., Gonzalez, J., and Heldin, C. H. (1995) Anal. Biochem. 224, 451–455.
11. McCaughan, K. K., Brown, C. M., Dalphon, M. E., Berry, M. J., and Tate, W. P. (1995) Proc. Natl. Acad. Sci. U.S.A. 92, 5431–5435.
12. Nasim, M. T., Jenecke, S., Belduz, A., Kollau, H., Plohe, L., and McCarthy, J. E. G. (2000) J. Biol. Chem. 275, 14846–14852.
13. Grander-Culemann, E., Martin III, G. W., Tujebajeva, R., Harney, J. W., and Berry, M. J. (2001) J. Mol. Biol. 310, 699–707.
14. Burk, R. F., Hill, K. E., Boeglin, M. E., Ehner, F. F., and Chittum, H. S. (1997) Histoch. Cell Biol. 108, 11–15.
15. Kato, T., Read, R., Rozga, J., and Burk, R. F. (1992) Am. J. Physiol. 262, G854–G868.

SCHEME 2. Representation of selenoprotein P mRNA and isoforms of the protein. The mRNA is shown as a horizontal line with the open reading frame between the AUG and the UAA codons. UGA codons are located by vertical lines above the mRNA. Selenocysteine insertion sequence elements are stem-loop structures in the 3-untranslated region. Selenoprotein P isoforms are shown as Se-P with the number of selenium atoms in them as a subscript. Asterisks indicate selenocysteine residues.