Mass Spectrometric Identification of *N*- and *O*-Glycosylation Sites of Full-Length Rat Selenoprotein P and Determination of Selenide−Sulfide and Disulfide Linkages in the Shortest Isoform[†]

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

Mass Spectrometry Research Center, Department of Biochemistry, and Division of Gastroenterology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee 37232

Received April 22, 2003; Revised Manuscript Received June 26, 2003

ABSTRACT: Rat selenoprotein P is an extracellular glycoprotein of 366 amino acid residues that is rich in cysteine and selenocysteine. Plasma contains four isoforms that differ principally by length at the C-terminal end. Mass spectrometry was used to identify sites of glycosylation on the full-length protein. Of the potential *N*-glycosylation sites, three located at residues 64, 155, and 169 were occupied, while the two at residues 351 and 356 were not occupied. Threonine 346 was variably *O*-glycosylated. Thus, full-length selenoprotein P is both *N*- and *O*-glycosylated. The shortest isoform of selenoprotein P, which terminates at residue 244, was analyzed for selenide−sulfide and disulfide linkages. In this isoform, a single selenocysteine and seven cysteines are present. Mass spectrometric analysis indicated that a selenide−sulfide bond exists between Sec40 and Cys43. Two disulfides were also detected as Cys149−Cys167 and Cys153−Cys156. The finding of a selenide−sulfide bond in the shortest isoform is compatible with a redox function of this pair that might be analogous to the selenol−thiol pair near the C terminus of animal thioredoxin reductase. The disulfide formed by Cys153−Cys156 also has some characteristics of a redox active pair.

Selenoprotein P is an unusual extracellular glycoprotein present in plasma at relatively high concentrations (*1*). It is the only selenoprotein that has been shown to contain more than one selenium atom. Rat, mouse, and human selenoprotein P mRNAs have 10 UGAs in their open reading frames and 2 SECIS elements in their 3′-untranslated regions (*2*−*4*). UGA, the opal termination codon, directs insertion of selenocysteine when a SECIS element is present in an appropriate location (*5*). Thus, selenoprotein P would be expected to contain 10 selenocysteine residues in its primary structure.

Rat plasma selenoprotein P has been purified and shown to exist in four isoforms (*6*). One isoform is the full-length protein, with its amino acid sequence identical to the sequence implied by the cDNA (*6, 7*). Three shorter isoforms that terminate, respectively, at the positions of the 2nd, 3rd, and 7th UGAs of the mRNA were separated and characterized by mass spectrometry (*6*). It appears that the four isoforms are synthesized from the same mRNA. The presence of these isoforms implies that some of the internal UGAs have alternative functions of translation termination or insertion of selenocysteine.

Selenoprotein P mRNA is present in many tissues, but liver appears to be the primary source of plasma selenoprotein P (*8*). The functions of selenoprotein P have not been established with certainty. However, there is evidence it supplies selenium to tissues such as the brain and testis (*9*, *10*) and that it serves to protect endothelial cells against oxidative damage (*11*).

One of our strategies for determining the function of selenoprotein P is to characterize its isoforms. In this report we describe the characterization of the glycosylation of the full-length isoform and our identification of selenide−sulfide and disulfide bridges in the shortest isoform using MALDI-TOF MS and nano-electrospray MS/MS.

MATERIALS AND METHODS

Materials. Iodoacetamide, DTT, and mass calibration standards, des-Arg[1]-bradykinin and bovine insulin, were obtained from Sigma Chemical Co. (St. Louis, MO). The modified trypsin was purchased from Promega (Madison, WI). Endoproteinase Glu-C and Lys-C were obtained from Boehringer Mannheim (Indianapolis, IN). PNGase F was obtained from ProZyme (San Leandro, CA). Matrix material, CHCA, was purchased from Aldrich Chemical Co. (Milwaukee, WI). ^{75}Se-labeled selenite (800 mCi/mg) was purchased from the University of Missouri Research Reactor Facility (Columbia, MO).

[†] This work was supported by NIH Grants ES02497, DK26657, ES00267, and GM58008.

* Corresponding author. Address: C2104, Medical Center North, Vanderbilt Medical Center, Nashville, TN 37232-2279. Phone (615) 343-7740. Fax (615) 343-6229. E-mail raymond.burk@vanderbilt.edu.

[‡] Department of Biochemistry.

[§] Present address: Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, NJ 07033.

[||] Department of Medicine.

[1] Abbreviations: SECIS, selenocysteine insertion sequence; DTT, dithiothreitol; CHCA, α-cyano-4-hydroxy-cinnamic acid; MALDI, matrix-assisted laser desorption ionization; HPLC, high performance liquid chromatography; TOF, time-of-flight; Q-TOF, quadrupole time-of-flight; ESI, electrospray ionization; MS, mass spectrometry; MS/MS, tandem mass spectrometry; LCQ, ThermoFinnigan Quadrupole Ion Trap Mass Spectrometer; PNGase F, *N*-glycosidase F.

10.1021/bi0346300 CCC: $25.00
Published on Web 07/26/2003

Purification of Rat Selenoprotein P and Its Isoforms. Selenoprotein P was purified from rat plasma, and its four isoforms were separated using a heparin-sepharose column as described previously (*6*, *12*). The full-length isoform, peak 2, was used for characterization of the glycosylation sites, while the short isoform, peak 1b, was used for mapping the selenide−sulfide bond and disulfide linkages.

Enzymatic Digestion. The full-length selenoprotein P was reduced by DTT and alkylated by iodoacetamide. Then it was digested by Glu-C or trypsin. The short isoform of selenoprotein P was alkylated with iodoacetamide and then digested with trypsin followed by Glu-C, or with Lys-C alone. The procedures of enzymatic digestion were described previously (*6*). Lys-C digestion of the unreduced short isoform (modified by iodoacetamide) was performed in 50 mM phosphate buffer, pH 7.8, at 37 °C for 6 h with an enzyme/protein ratio of 1/40.

Deglycosylation. The glycopeptides isolated from proteolytic digestion of full-length isoform were treated with PNGase F by incubation in 50 mM ammonium bicarbonate at 37 °C for 6 h. The peptide mixtures from trypsin followed by Glu-C digestion of short isoform or from Lys-C digestion were further deglycosylated with PNGase F by incubation at 37 °C for 6 h.

Structural Analysis of Oligosaccharides. Possible oligosaccharide compositions of a glycan structure were calculated using GlycoMod (http://www.expasy.ch/tools/glycomod/) software (*13*), with constitutive monosaccharide residues limited to Hexose (Hex), *N*-acetyl hexosamine (HexNAc), Deoxyhexose (dHex), and *N*-acetyl neuraminic acid (NeuAc).

Prediction of O-Glycosylation Sites. O-GLYCBASE, an updated database of information on glycoproteins and their *O*-linked glycosylation sites (*14*), was used to identify potential sites of *O*-glycosylation. This program (NetOGlyc) is available at Center for Biological Sequences CBS Prediction Server (http://www.cbs.dtu.dk).

Reverse-Phase HPLC. The proteolytic peptide mixtures were separated on a Vydac C18 column (2.1 × 250 mm) (Vydac, Hesperia, CA) using a HP 1100 HPLC system (Hewlett-Packard Co., Wilmington, DE). The HPLC conditions and gradient were the same as described previously (*6*).

Mass Spectrometry. MALDI-TOF mass spectra were obtained using a Voyager DE-STR MALDI-TOF mass spectrometer (Applied Biosystems, Foster City, CA), equipped with a 337 nm nitrogen laser, operated in linear mode. Mass calibration was accomplished using des-Arg^1-bradykinin (MW 903.46) and bovine insulin (MW 5733.58). CHCA, prepared at 10 mg/mL in 50% acetonitrile/49.9% H_2O/0.1% trifluroacetic acid, was used as the matrix.

MS/MS experiments were performed either on a Finnigan LCQ mass spectrometer (ThermoFinnigan, San Jose, CA) or a PE Sciex QSTAR (Q-TOF type) mass spectrometer (PE Sciex, Concord, Canada). Both mass spectrometers were equipped with a nanoelectrospray ion source (Protana A/S, Odense, Denmark) and operated on positive ion mode with a spray voltage of 800−900 V. The collision energy was set as 25−35% in LCQ and 30−45 eV in QSTAR mass spectrometer. The multiple stage MS/MS/MS experiment was performed on LCQ mass spectrometer.

RESULTS

N-Glycosylation Sites. N-Glycosidic linkage in proteins is through the amine nitrogen of asparagine in the tripeptide consensus sequence Asn-Xxx-Ser/Thr, where Xxx is any amino acid except proline. The presence of this sequence is not in itself sufficient to ensure glycosylation, however (*15*, *16*). There are five such potential *N*-glycosylation sites in selenoprotein P, located at asparagine residues 64, 155, 169, 351, and 356.

The basic strategy used in this study to characterize the glycosylation sites of the full-length selenoprotein P was to digest the reduced and alkylated protein with Glu-C or trypsin. The proteolytic peptides were separated by reverse-phase HPLC, and each fraction was then analyzed by MALDI-TOF MS. Those fractions containing glycopeptides were identified from their carbohydrate heterogeneity and were treated with PNGaseF. The deglycosylated peptides were analyzed by MALDI-TOF MS, and the peptide sequences were verified by nano-ESI MS/MS.

A peptide isolated by HPLC from a Glu-C digestion of full-length selenoprotein P was identified to have mass peaks at *m*/*z* 5183.2, 5473.7, 5516.1, and 5808.2 on MALDI-TOF mass analysis, as shown in Figure 1a. The ions at *m*/*z* 2592.0, 2737.2, 2758.7, and 2904.5 were the doubly charged ions of the respective glycopeptides above. After these peptides had been treated with PNGase F, which cleaves off asparagine-linked carbohydrates and converts the residue asparagine to aspartic acid, MALDI-TOF mass spectrometric analysis showed that the high mass peaks had disappeared, while a new peptide signal at *m*/*z* 2146.3 had appeared, as shown in Figure 1b. This new mass peak corresponds to the peptide of residues 165−184 (calculated *m*/*z* for $[M + H]^+$ 2145.3) with 1 Da mass increase from the conversion of asparagine at residue 169 to aspartic acid by PNGase F. A nano-ESI MS/MS experiment was performed on the doubly charged ion at *m*/*z* 1073.6. The fragmentation pattern verified the sequence as shown in Figure 2. Moreover, the masses of the fragment ions also confirmed the conversion of asparagine to aspartic acid at residue 169 because the mass difference between fragment ion y_{16}^+ and y_{15}^+ was 115 (residue mass for Asp) rather than 114 (residue mass for Asn), while the mass difference between fragment ion b_5^+ and b_3^+ was 243 (residue mass for LysAsp) rather than 242 (residue mass for LysAsn). Therefore, Asn169 was determined to be an *N*-glycosylation site.

The same methodology was used to identify the other *N*-glycosylation sites in selenoprotein P. One tryptic peptide showed mass peaks at *m*/*z* 5025.8, 5360.6, 5690.6, and 6023.0 on MALDI-TOF mass analysis, which were eliminated after incubation with PNGase F with appearance of a new peak at *m*/*z* 2738.0 (data not shown). This mass matched the mass of the peptide consisting of residues 57−80 (calculated *m*/*z* for $[M + H]^+$ 2737.0), with the conversion of asparagine at residue 64 to aspartic acid. Therefore, Asn64 is identified as an *N*-glycosylation site. The peptide sequence was further verified by a nano-ESI MS/MS experiment (data not shown). Similarly, a glycopeptide from Glu-C digestion showed mass peaks at *m*/*z* 5307.2, 5604.2, 5622.6, and 5913.5 from MALDI-TOF mass analysis. After incubation with PNGase F, these peaks disappeared and a new peak appeared at *m*/*z* 2654.2. This peptide was identified as

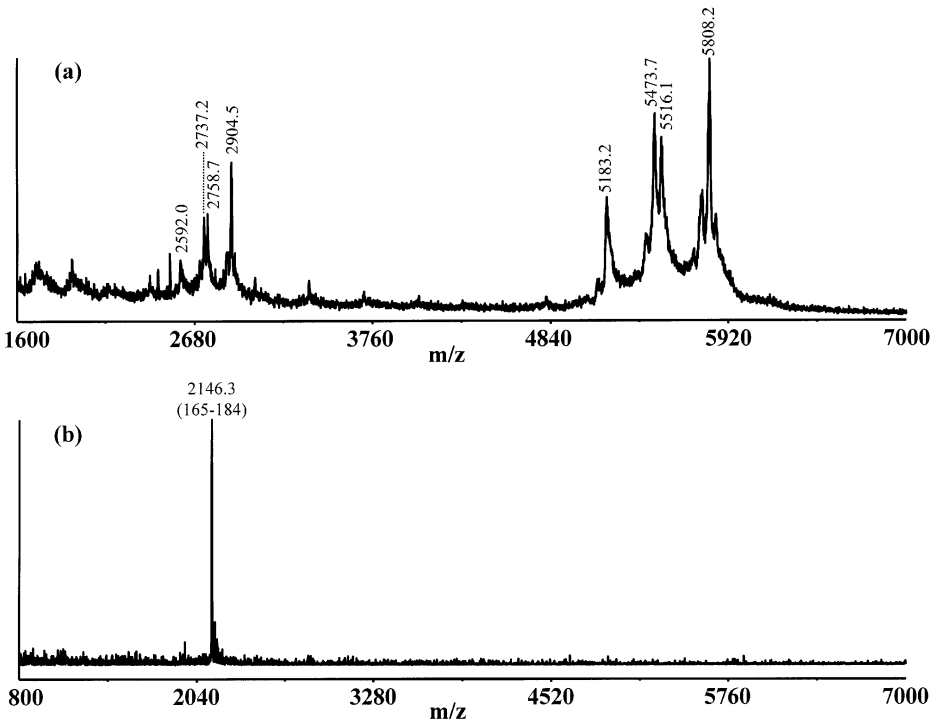

FIGURE 1: (a) MALDI-TOF mass spectrum of a peptide separated by HPLC from a Glu-C digest of reduced and alkylated selenoprotein P. (b) MALDI-TOF mass spectrum of the same peptide after incubation with PNGase F.

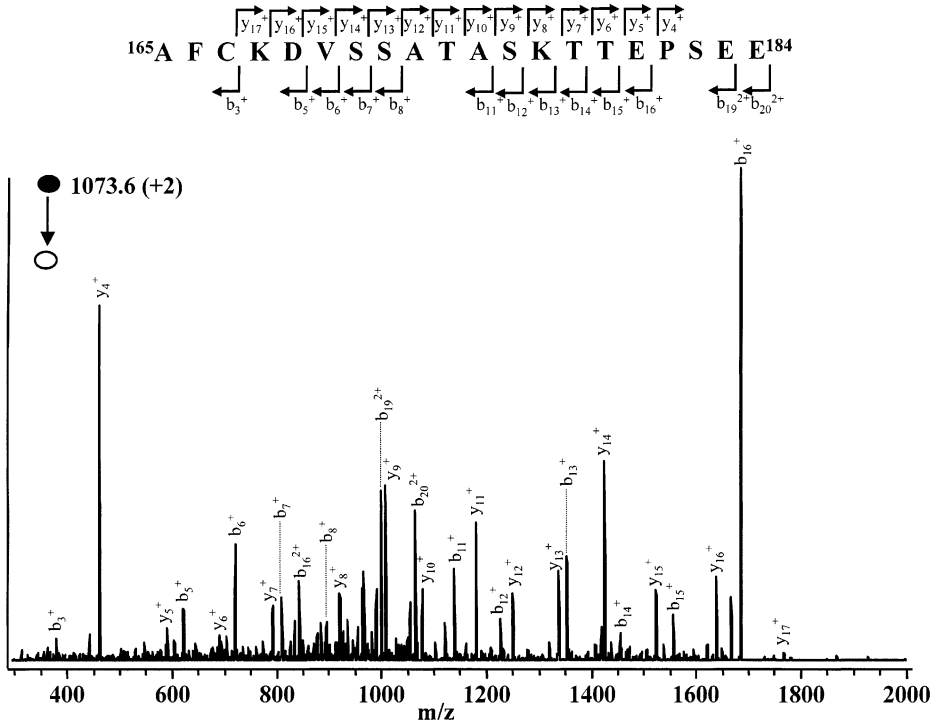

FIGURE 2: Nano-ESI MS/MS spectrum of the deglycosylated peptide studied in Figure 1b (residues 165−184), which was obtained from fragmentation of the doubly charged ion at *m/z* 1073.6 (experiment performed using a Finnigan LCQ mass spectrometer).

consisting of residues 143−164 (calculated *m/z* for $[M + H]^+$ 2653.0), with the modification of asparagine at residue 155 to aspartic acid, by nano-ESI MS/MS (data not shown). Therefore, Asn155 is also *N*-glycosylated.

Asparagine residues 351 and 356 are not *N*-glycosylated because a Glu-C peptide consisting of residues 348−366 (C-terminal peptide) showed no evidence of glycosylation (data not shown) (*6*).

The oligosaccharide compositions of various glycoforms attached to each of the three glycoslation sites were predicted using GlycoMod software, which mathematically finds all possible compositions of a glycan structure from its experimentally determined glycopeptide mass and the calculated mass from the peptide sequence. To simplify search results, only the four common monosaccharide residues, hexose, *N*-acetyl hexosamine, deoxyhexose, and *N*-acetyl neuraminic

Table 1: Possible Oligosaccharide Compositions for Different Glycoforms at *N*-Glycosylation Sites Asn64, Asn155, and Asn169, and One *O*-Glycosylation Site Thr346

glycosylation site	calculated peptide $[M + H]^+$	peptide sequence	observed glycopeptide $[M + H]^+$	possible oligosaccharide composition
Asn64	2737.0	^{57}LENQGYFNISYIVVNHQGSPSQLK^{80}	5025.8	$(Hex)_5(dHex)_4+(Man)_3(GlcNAc)_2$
				$(HexNAc)_4(NeuAc)_2+(Man)_3(GlcNAc)_2$
				$(HexNAc)_4(dHex)_2(NeuAc)_1+(Man)_3(GlcNAc)_2$
			5360.6	$(Hex)_1(HexNAc)_7(dHex)_1+(Man)_3(GlcNAc)_2$
			5690.6	$(Hex)_{10}(dHex)_3+(Man)_3(GlcNAc)_2$
				$(Hex)_3(HexNAc)_2(dHex)_4(NeuAc)_2+(Man)_3(GlcNAc)_2$
				$(Hex)_5(HexNAc)_4(dHex)_1(NeuAc)_1+(Man)_3(GlcNAc)_2$
				$(Hex)_3(HexNAc)_2(dHex)_6(NeuAc)_1+(Man)_3(GlcNAc)_2$
			6023.0	$(Hex)_9(HexNAc)_1(dHex)_3(NeuAc)_1+(Man)_3(GlcNAc)_2$
				$(Hex)_2(HexNAc)_3(dHex)_4(NeuAc)_3+(Man)_3(GlcNAc)_2$
				$(Hex)_{11}(HexNAc)_3+(Man)_3(GlcNAc)_2$
				$(Hex)_9(HexNAc)_1(dHex)_5+(Man)_3(GlcNAc)_2$
				$(Hex)_4(HexNAc)_5(dHex)_1(NeuAc)_2+(Man)_3(GlcNAc)_2$
				$(Hex)_2(HexNAc)_3(dHex)_6(NeuAc)_2+(Man)_3(GlcNAc)_2$
Asn155	2653.0	^{143}AIKIAYCEKRCGNCSFTSLEDE^{164}	5307.2	$(Hex)_6(HexNAc)_1(dHex)_4+(Man)_3(GlcNAc)_2$
				$(Hex)_1(HexNAc)_5(NeuAc)_2+(Man)_3(GlcNAc)_2$
				$(Hex)_1(HexNAc)_5(dHex)_2(NeuAc)_1+(Man)_3(GlcNAc)_2$
			5604.2	$(Hex)_3(HexNAc)_2(NeuAc)_4+(Man)_3(GlcNAc)_2$
				$(Hex)_3(HexNAc)_2(dHex)_2(NeuAc)_3+(Man)_3(GlcNAc)_2$
			5622.6	$(Hex)_{11}(dHex)_2+(Man)_3(GlcNAc)_2$
				$(Hex)_4(HexNAc)_2(dHex)_3(NeuAc)_2+(Man)_3(GlcNAc)_2$
				$(Hex)_6(HexNAc)_4(NeuAc)_1+(Man)_3(GlcNAc)_2$
				$(Hex)_4(HexNAc)_2(dHex)_5(NeuAc)_1+(Man)_3(GlcNAc)_2$
			5913.5	$(Hex)_4(HexNAc)_2(dHex)_3(NeuAc)_3+(Man)_3(GlcNAc)_2$
				$(Hex)_{11}(dHex)_4+(Man)_3(GlcNAc)_2$
				$(Hex)_6(HexNAc)_4(NeuAc)_2+(Man)_3(GlcNAc)_2$
				$(Hex)_4(HexNAc)_2(dHex)_5(NeuAc)_2+(Man)_3(GlcNAc)_2$
Asn169	2145.3	^{165}AFCKNVSSATASKTTEPSEE^{184}	5183.2	$(Hex)_3(HexNAc)_1(NeuAc)_5+(Man)_3(GlcNAc)_2$
			5473.7	$(Hex)_3(HexNAc)_6(dHex)_3(NeuAc)_1+(Man)_3(GlcNAc)_2$
				$(Hex)_5(HexNAc)_8+(Man)_3(GlcNAc)_2$
				$(Hex)_3(HexNAc)_6(dHex)_5+(Man)_3(GlcNAc)_2$
			5516.1	$(Hex)_4(HexNAc)_9+(Man)_3(GlcNAc)_2$
				$(Hex)_2(HexNAc)_7(dHex)_5+(Man)_3(GlcNAc)_2$
			5808.2	$(Hex)_4(HexNAc)_9(dHex)_2+(Man)_3(GlcNAc)_2$
Thr346	2344.4	^{328}SCQCRSPPAAUHSQHVSPTE^{347}	2546.8	$(HexNAc)_1$
			2709.0	$(Hex)_1(HexNAc)_1$
			3000.2	$(Hex)_1(HexNAc)_1(NeuAc)_1$
			3291.0	$(Hex)_1(HexNAc)_1(NeuAc)_2$

acid, were included as the possible residues of the oligosaccharides. The search results are summarized in Table 1. Tandem mass spectrometry experiments of the intact glycopeptide were attempted but were unsuccessful because of the low intensity of the glycopeptides. Therefore, the structures of the *N*-glycans remain to be characterized directly.

O-Glycosylation Site. The most common type of *O*-glycosidic linkage is the attachment through *N*-acetyl hexosamine to the side chain of serine or threonine, although several other monosaccharides can also form *O*-glycosidic linkages (*17*). In contrast to *N*-linked glycosylation, no simple consensus sequence has been found to predict *O*-linked glycosylation, and hence any serine or threonine is capable of being *O*-glycosylated (*18*). This is consistent with each *O*-glycan linkage of monosaccharides and amino acids requiring its own enzyme (*13*).

Although there are no clear-cut acceptor sequences for *O*-glycosylation site selection, artificial neural networks, which take into consideration sequence context, secondary structure, and structure accessibility, have been developed for the prediction of *O*-glycosylation sites and found to be quite reliable (*19*, *20*).

Using this program, threonine residues 174, 178, 179, and 346 were identified as potential *O*-glycosylation sites. A glycopeptide with residues 165−184 was isolated from Glu-C digestion, and this peptide included three of the four potential *O*-glycosylation sites. However, all the carbohydrate components in this glycopeptide were removed by PNGase F, indicating that all the carbohydrate components in this peptide were *N*-linked through Asn169 (Figure 1). Therefore, threonine residues 174, 178, and 179 are not *O*-glycosylation sites. Thr346 was identified as an *O*-glycosylation site previously (*6*). Additional proof of this was obtained in this study. A Glu-C peptide showed masses at *m*/*z* 2343.6, 2546.8, 2709.0, 3000.2, and 3291.0, from MALDI-TOF mass analysis (data not shown). Nano-ESI MS/MS of the triply charged ion at *m*/*z* 1000.8 corresponding to the singly charged glycopeptide at *m*/*z* 3000.2 showed fragment ions corresponding to consecutive loss of *N*-acetyl neuraminic acid (291), hexose (162), and *N*-acetyl hexosamine (203) (Figure 3a). The fragment ions at *m*/*z* 1172.5 and 781.6 correspond to the peptide consisting of residues 328−347 with a charge state of +2 and +3, respectively. Tandem mass spectrometry confirmed the presence of carbohydrate at Thr346, while no peptide sequence information was obtained because the glycopeptide preferentially cleaved at *O*-glycosidic bonds and the peptide backbone remained intact. The peptide sequence was verified by a nano-ESI MS/MS/MS experiment carried out by further selecting and fragmenting

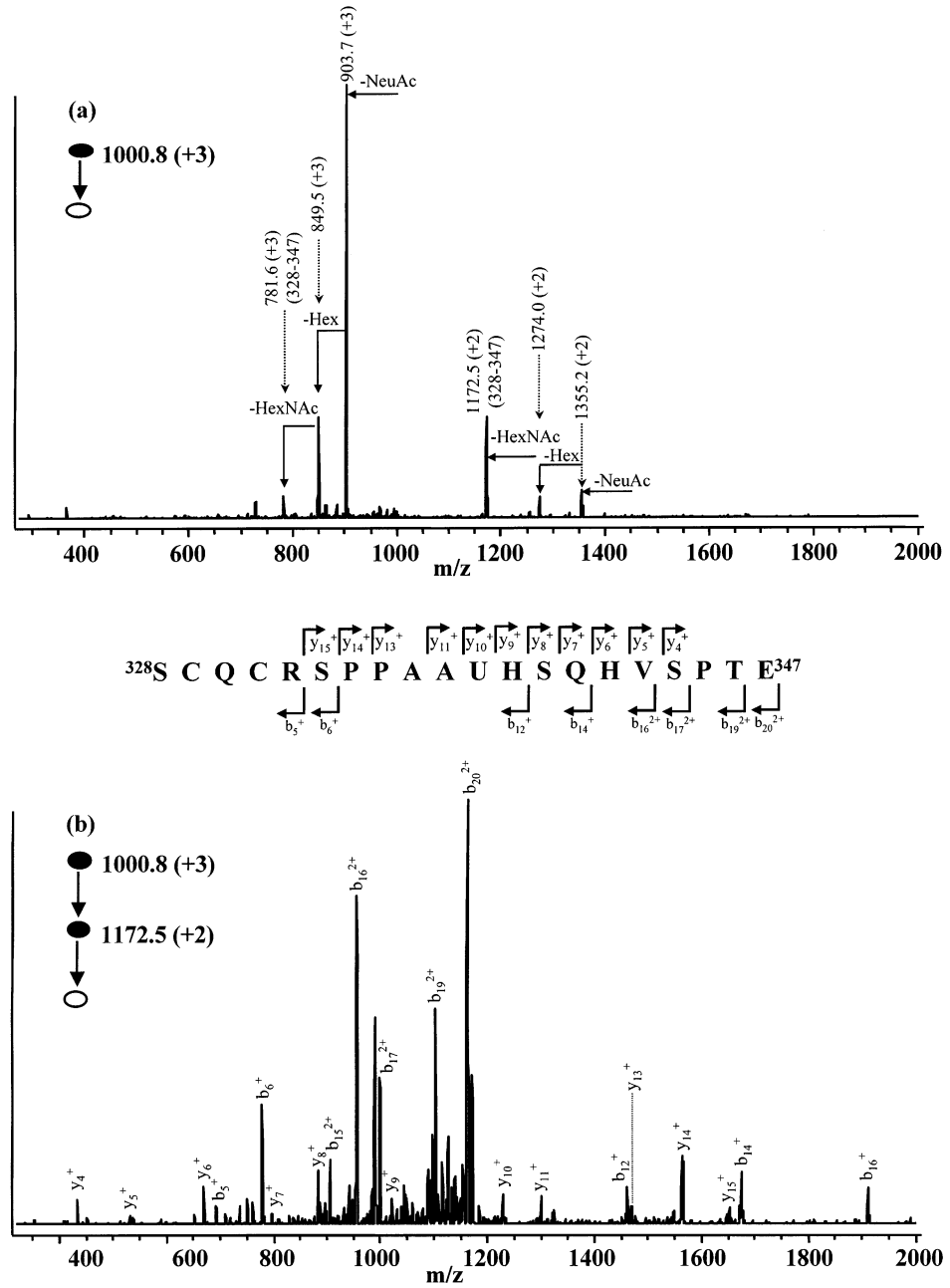

FIGURE 3: (a) Nano-ESI MS/MS spectrum of the *O*-glycosylated peptide consisting of residues 328−347, which was obtained from the fragmentation of the triply charged ion at *m*/*z* 1000.8. (b) Nano-ESI MS/MS/MS spectrum of the *O*-glycosylated peptide obtained from the fragmentation of the triply charged ion at *m*/*z* 1000.8, followed by further fragmentation of the doubly charged ion at *m*/*z* 1172.5 (experiment performed using a Finnigan LCQ mass spectrometer).

the doubly charged ion at *m*/*z* 1172.5. The masses of fragment ions confirmed the peptide sequence (Figure 3b). Therefore, Thr346 was identified as *O*-glycosylated in some full-length selenoprotein P molecules but not in all of them.

Overview of the Strategy of Disulfide Linkage Determination. The strategy for determining the oxidation state of cysteine residues was to alkylate all free cysteines and selenocysteines in the short isoform of selenoprotein P with iodoacetamide, digest with endoproteinase, deglycosylate with PNGase F, separate on reverse phase HPLC, and analyze by mass spectrometry.

The short isoform of selenoprotein P contains one selenocysteine as residue 40 and 7 cysteines as residues 9, 43, 121, 149, 153, 156, and 167. Mass spectrometric analysis of the Lys-C digest of the short isoform indicated that Cys9 and Cys121 were present in reduced form because peptides consisting of residues 1−10 (N-terminal peptide) and residues 113−145 were isolated and tandem mass spectrometric analysis showed that Cys9 and Cys121 were modified by iodoacetamide (data not shown).

Selenide−Sulfide Bond. A peptide with an isotope pattern indicating that it contained one selenium atom was isolated from a digestion with trypsin followed by Glu-C. The measured *m*/*z* value of 2355.1 (Table 2) was 2 Da less than the calculated *m*/*z* for the peptide consisting of residues 28−49 (calculated $[M + H]^+$ 2357.6 from the peptide sequence, data not shown). This implied that there is a selenide−sulfide bond between Sec40 and Cys43. A nano-ESI MS/MS experiment was performed on this peptide by fragmenting the triply charged ion at *m*/*z* 785.29 to verify the peptide

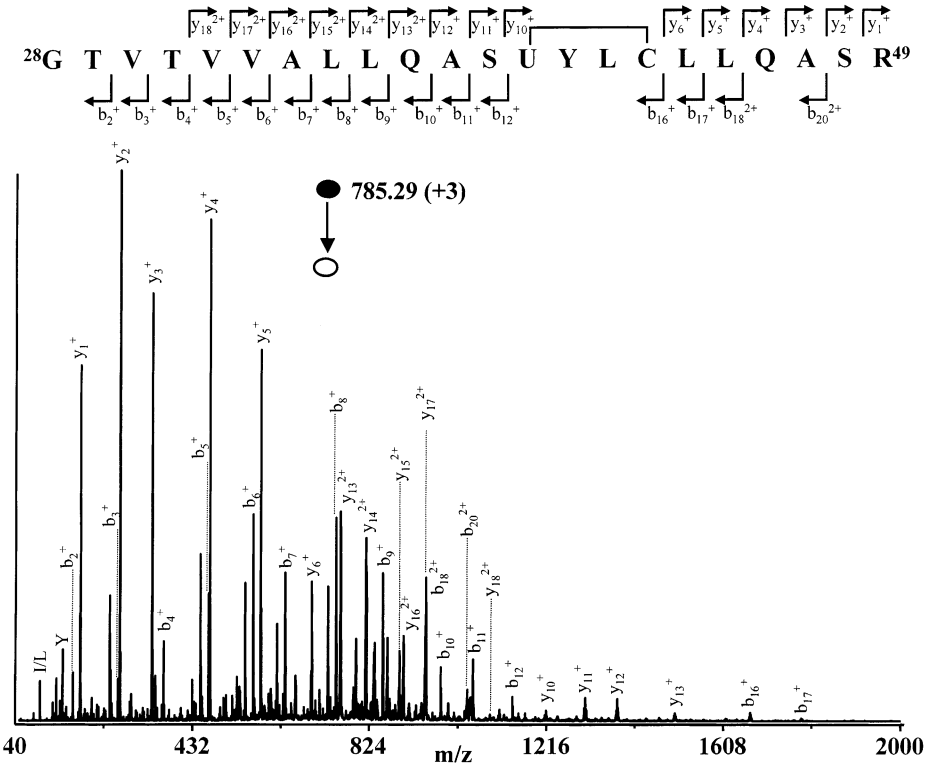

FIGURE 4: Nano-ESI MS/MS spectrum of the peptide consisting of residues 28−49 with a selenide−sulfide bond between Sec40 and Cys43. The spectrum was obtained from the fragmentation of the triply charged ion at *m*/*z* 785.29 (experiment performed using a PE Sciex QSTAR mass spectrometer).

Table 2: Calculated and Measured Masses of the Selenide−Sulfide and Disulfide Bonded Peptides in the Short Isoform of Selenoprotein P.

sequence	sequence position	$[M + H]^+_{calculated}$	$[M + H]^+_{measured}$
GTVTVVALLQASUYLCLLQASR	28−49	2355.6	2355.1
IAYCEK	146−151	726.3	726.4
RCGDCSFTSLEDEAFCK	152−168	1908.8	1908.9
IAYCEK RCGDCSFTSLEDEAFCK	[146−151] + [152−168]	2633.8	2633.1

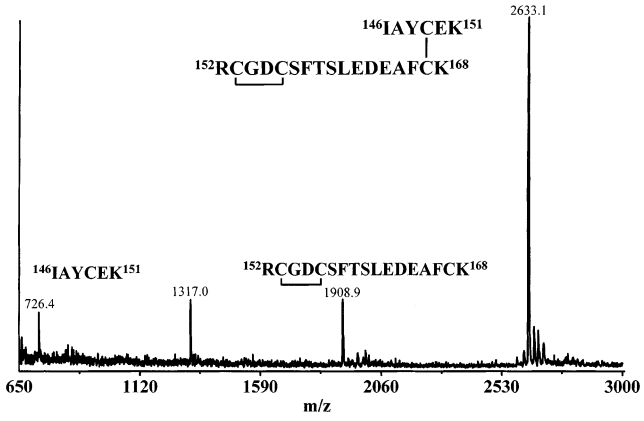

FIGURE 5: MALDI-TOF mass spectrum of a protein fragment consisting of two peptides (residues 146−151 and residues 152−168) linked by a disulfide bond.

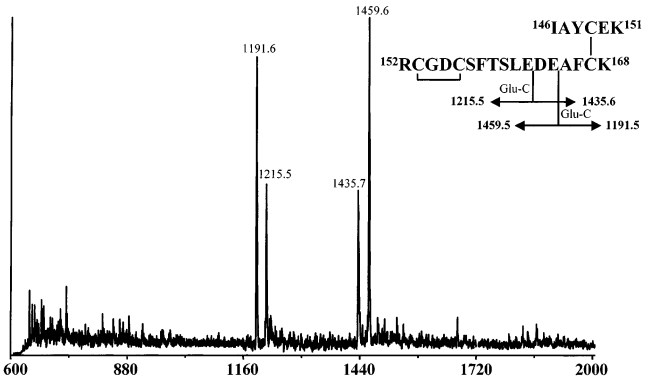

FIGURE 6: MALDI-TOF mass spectrum of the disulfide-linked protein fragment studied in Figure 5 after it had been digested by Glu-C. The calculated masses of the proteolytic peptides from Glu-C digestion are indicated adjacent to the amino acid sequence.

sequence and Se−S linkage. As shown in Figure 4, the fragment ion masses of the b-type ions, b_{16}^+, b_{17}^+, b_{18}^{2+}, and b_{20}^{2+}, confirmed the presence of a selenide−sulfide bond between Sec40 and Cys43. Also, the same conclusion was reached by analyzing the y-type fragment ion masses of y_{10}^+ to y_{18}^{2+}. The fragment ions corresponding to b_{13}^+ to b_{15}^+ and y_7^+ to y_9^+ were not observed. This is because the appearance of these fragment ions would involve the cleavage of the peptide backbone together with the selenide−sulfide bond, while under the low energy collision induced dissociation, the conversion of collision energy to internal energy is not sufficient to break these bonds (*21*, *22*).

Disulfide Bonds. MALDI-TOF MS peptide mapping of the short isoform of selenoprotein P showed a peptide ion

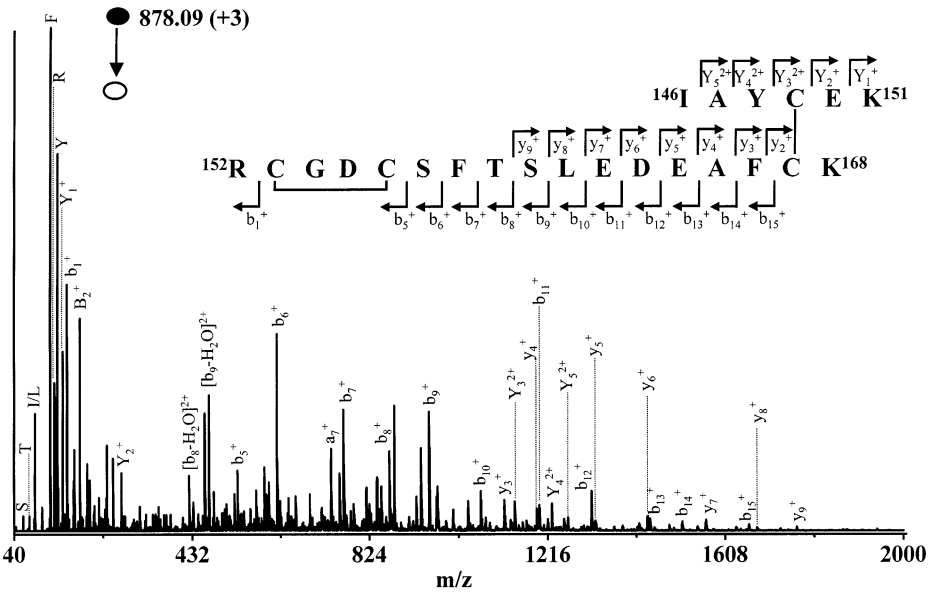

FIGURE 7: Nano-ESI MS/MS spectrum of the disulfide-linked protein fragment studied in Figure 5 (experiment performed at a PE Sciex QSTAR mass spectrometer).

that could not be addressed as a simple proteolytic peptide but appeared to be two peptides linked by a disulfide bond. For example, a peptide signal was observed at *m*/*z* 2633.1, which matched the combined masses of peptides 146−151 and 152−168 minus 4 Da (Table 2). This 4 Da difference indicated that there were two disulfide bonds, suggesting that peptide 146−151 was linked to peptide 152−168 through a disulfide bond while there was an additional disulfide bond within peptide 152−168 (Figure 5). Note that Asn155 in the peptide sequence was converted to aspartic acid by deglycosylation with PNGase F. The ion at *m*/*z* 1317.0 corresponded to the doubly charged ion of the major signal at *m*/*z* 2633.1. It is interesting to note that fragment ions matched to masses of the two individual constituent monomer peptides were also observed. The ion at *m*/*z* 726.4 matched the mass of peptide 146−151, while the ion at *m*/*z* 1908.9 matched the mass of peptide 152−168 containing a disulfide bond. These ions resulted from the *in situ* decomposition during MALDI conditions of the disulfide bond that linked the peptides 146−151 and 152−168. The precise mechanism of disulfide cleavage during MALDI-MS analysis is still unknown, but the fragmentation appears to originate during the initial ionization event (*23*, *24*). The decomposition of this protein fragment revealed the disulfide pairings, providing a simple method for mapping disulfide bonds. However, this result did not indicate which one of the three cysteines of peptide 152−168 was linked to Cys149 in peptide 146−151. Therefore, further experiments were needed to reveal the disulfide bond linkages.

After the two peptides connected by a disulfide were further treated with endoprotease Glu-C for 2 h, mass spectrometric analysis showed four ion signals, at *m*/*z* 1191.6, 1215.5, 1435.7, and 1459.6, as seen in Figure 6. The ion at *m*/*z* 1191.6 was assigned to the peptide 165−168 linked to peptide 146−151 through a disulfide bond between Cys167 and Cys149 arising from the cleavage at the C-terminus of Glu164. The ion of the remainder of the molecule was observed at *m*/*z* 1459.6, corresponding to the peptide 152−164 with a disulfide bond between Cys153 and Cys156. Similarly, the ion at *m*/*z* 1435.7 was assigned to the peptide 163−168 linked to peptide 146−151 through a disulfide bond between Cys167 and Cys149, and the ion of the remainder of the molecule was observed at *m*/*z* 1215.5, corresponding to the peptide 152−162 with a disulfide bond between Cys153 and Cys156. This experiment concluded that there were two disulfide bonds in this protein fragment. They are Cys149-Cys167, and Cys153-Cys156.

The disulfide linkages were also confirmed by nano-ESI MS/MS experiments, as shown in Figure 7. The absence of fragment ions of b_2^+, b_3^+, and b_4^+ implied that there is a disulfide bond between Cys153 and Cys156. The measured fragment ion masses for other b^+ ion series, b_5^+ to b_{15}^+, which were 2 Da less than the corresponding calculated masses from the peptide sequence, confirmed the existence of the disulfide bond between Cys153 and Cys156. The masses of fragment ions of the y^+ ion series, y_2^+ to y_9^+, accounted for the mass of fragment ions of y-ion series from peptide 152−168 together with the mass of the peptide with sequence 146−151, which indicated the existence of another disulfide bond between Cys149 and Cys167. These data provide a complete verification of the proposed disulfide linkage in these disulfide-linked peptides.

DISCUSSION

Studies using mice with deletion of the selenoprotein P gene have shown conclusively that this selenium rich protein supplies selenium to the testis and the brain (*9*). Additional functions are postulated on the basis of studies showing enzymatic activity (*25*) and association with selenium dependent protection against oxidative damage (*26*). Moreover, the identification of four isoforms of selenoprotein P in rat plasma, the shortest of which contains only one selenium atom per molecule, suggests functions beyond transport of the element (*6*). We are characterizing the selenoprotein P molecule in order to gain insight into its function.

Glycosylation is an important posttranslational modification of proteins. The functions of carbohydrate attached to proteins range from effects on folding of the protein to

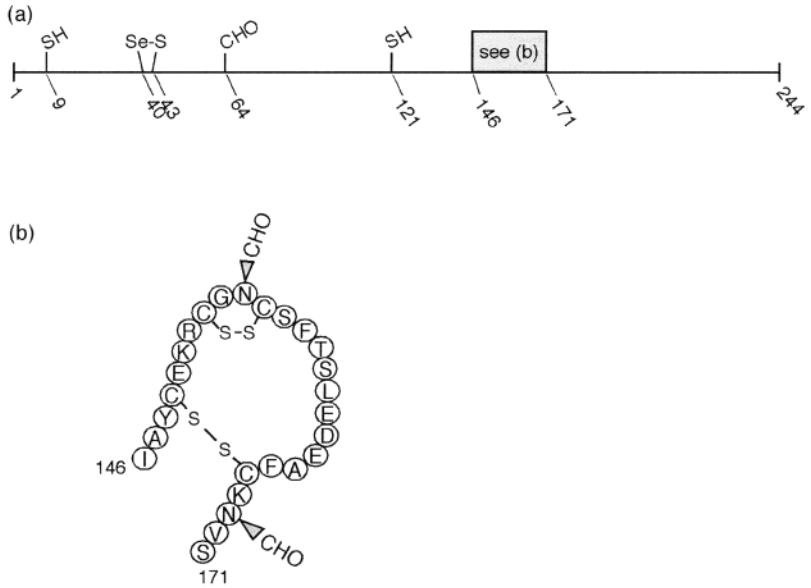

FIGURE 8: (a) Schematic view of shortest isoform of selenoprotein P showing structural features elucidated in this study. (b) An enlargement of the sequence consisting of residues 146−171.

formation of recognition sites on it. Three of the five potential *N*-glycosylation sites in the full-length form of selenoprotein P located at asparagine residues 64, 155, and 169 were determined to be glycosylated. The possible compositions of the *N*-glycans have been obtained from computer calculations, but their precise structures remain to be characterized. Selenoprotein P was also partially *O*-glycosylated at Thr346. The *O*-linked glycan structures were identified to be $(Hex)_1$, $(Hex)_1(HexNAc)_1$, $(Hex)_1(HexNAc)_1(NeuAc)_1$, and $(Hex)_1(HexNAc)_1(NeuAc)_2$, respectively.

The shorter selenoprotein P isoforms are also glycosylated. The masses of modifications to these isoforms can be estimated to be the difference between their determined masses (*6*) and their calculated peptide masses. Predicted masses of modifying carbohydrate in the three shorter isoforms approximate the carbohydrate masses estimated for the full-length protein. Thus, the three shorter isoforms of selenoprotein P can be predicted to be glycosylated in a manner similar to the full-length protein.

Enzymatic activity of selenoproteins almost invariably depends on the redox properties of selenium in the active site (*27*). Disulfide linkages are common in extracellular proteins and are important in establishing and maintaining tertiary structure. We chose to study the selenocysteine and cysteine residues of the shortest isoform of selenoprotein P because of its relative simplicity when compared with the longer isoforms that contain more selenium. The shortest isoform of selenoprotein P terminates after amino acid residue 244; it contains seven cysteine residues and one selenocysteine residue.

Figure 8a represents the oxidation state of the selenocysteine and cysteine residues in the shortest isoform. In the purified protein that had not been protected from oxygen, the selenocysteine residue in position 40 was present in an oxidized form linked to the cysteine in position 43. Animal thioredoxin reductase provides a precedent for such a selenocysteine−cysteine pair (*28*). Activity of that enzyme depends on redox cycling of the selenol−thiol pair at the C terminus of the protein. It seems unlikely that the pair in selenoprotein P serves a structural function because of its selenium content and because of the nearness of the residues to one another. Thus, we propose that this is an active site for enzymatic activity. Already, selenoprotein P has been shown to be capable of reducing hydroperoxides (*25*), and thioredoxin can serve as a reducing substrate in that reaction (*29*). The proposed active site at residues 40−43 would appear to be suited to be responsible for this activity.

The disulfide between Cys153 and Cys156 is also a candidate to be a redox center. Once again, thioredoxin reductase provides a precedent for such a pair (*28*). In thioredoxin reductase, a redox-active disulfide is reduced by NADPH acting through a flavin, and in turn, the thiol pair reduces the selenide−sulfide at the C terminus to the active selenol-thiol pair. We have no proof that such a mechanism is active in selenoprotein P, but this analysis of its physical characteristics raises the possibility that this occurs. The fact that one of the occupied *N*-glycosylation sites is present between the cysteine residues of this pair would appear to lessen the likelihood that this site is redox active. The disulfide between Cys149 and Cys167 creates a loop that contains the disulfide between Cys153 and Cys156 (Figure 8b). This loop might be important in positioning the latter cysteine pair. The remaining cysteine residues at positions 9 and 121 were reduced in the protein as isolated. Thus, little can be speculated about their function.

Results presented here indicate that full-length selenoprotein P has three occupied *N*-glycosylation sites and one variably occupied *O*-glycosylation site. There is microheterogeneity of the glycosylation at all sites. Mass measurements indicate that the carbohydrate can account for all the modification of the polypeptide but do not rule out small modifications of other types. The shortest isoform contains a selenide−sulfide and a disulfide that are speculated to serve as redox active pairs. These results support the proposal that selenoprotein P functions as an enzyme with redox functions.

ACKNOWLEDGMENT

The authors are grateful to Ansley S. Tharpe, Todd B. Peterson, and Troy E. Morris for preparation of selenoprotein P isoforms.

REFERENCES

1. Burk, R. F., and Hill, K. E. (1999) *BioEssays 21*, 231−237.
2. Hill, K. E., Lloyd, R. S., Yang, J.-G., Read, R., and Burk, R. F. (1991) *J. Biol. Chem. 266*, 10050−10053.
3. Hill, K. E., Lloyd, R. S., and Burk, R. F. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 537−541.
4. Steinert, P., Ahrens, M., Gross, G., and Flohé, L. (1997) *Biofactors 6*, 311−319.
5. Low, S. C., and Berry, M. J. (1996) *TIBS 21*, 203−208.
6. Ma, S., Hill, K. E., Caprioli, R. M., and Burk, R. F. (2002) *J. Biol. Chem. 277*, 12749−12754.
7. Himeno, S., Chittum, H. S., and Burk, R. F. (1996) *J. Biol. Chem. 271*, 15769−15775.
8. Kato, T., Read, R., Rozga, J., and Burk, R. F. (1992) *Am. J. Physiol. 262*, G854−G858.
9. Hill, K. E., Zhou, J., McMahan, W. J., Motley, A. K., Atkins, J. F., Gesteland, R. F., and Burk, R. F. (2003) *J. Biol. Chem. 278*, 13640−13646.
10. Schomburg, L., Schweizer, U., Holtmann, B., Flohe, L., Sendtner, M., and Kohrle, J. (2003) *Biochem. J. 370*, 397−402.
11. Atkinson, J. B., Hill, K. E., and Burk, R. F. (2001) *Lab. Invest. 81*, 193−200.
12. Chittum, H. S., Himeno, S., Hill, K. E., and Burk, R. F. (1996) *Arch. Biochem. Biophys. 325*, 124−128.
13. Cooper, C. A., Gasteiger, E., and Packer, N. H. (2001) *Proteomics 1*, 340−349.
14. Hansen, J. E., Lund, O., Rapacki, K., and Brunak, S. (1997) *Nucleic Acids Res. 25*, 278−282.
15. Krogh, T. N., Bachmann, E., Teisner, B., Skjodt, K., and Hojrup, P. (1997) *Eur. J. Biochem. 244*, 334−42.
16. Bause, E., and Legler, G. (1981) *Biochem. J. 195*, 639−44.
17. Clausen, H., and Bennett, E. P. (1996) *Glycobiology 6*, 635−646.
18. Elhammer, A. P., Poorman, R. A., Brown, E., Maggiora, L. L., Hoogerheide, J. G., and Kezdy, F. J. (1993) *J. Biol. Chem. 268*, 10029−38.
19. Hansen, J. E., Lund, O., Tolstrup, N., Gooley, A. A., Williams, K. L., and Brunak, S. (1998) *Glycoconjugate J. 15*, 115−30.
20. Hansen, J. E., Lund, O., Engelbrecht, J., Bohr, H., Nielsen, J. O., Hansen, J.-E. S., and Brunk, S. (1995) *Biochem. J. 308 (Pt 3)*, 801−13.
21. Bean, M. F., and Carr, S. A. (1992) *Anal. Biochem. 201*, 216−226.
22. Yen, T.-Y., Joshi, R. K., Yan, H., Seto, N. O., Palcic, M. M., and Macher, B. A. (2000) *J. Mass Spectrom. 35*, 990−1002.
23. Crimmins, D. L., Saylor, M., Rush, J., and Thoma, R. S. (1995) *Anal. Biochem. 226*, 355−61.
24. Patterson, S. D., and Katta, V. (1994) *Anal. Chem. 66*, 3727−32.
25. Saito, Y., Hayashi, T., Tanaka, A., Watanabe, Y., Suzuki, M., Saito, E., and Takahashi, K. (1999) *J. Biol. Chem. 274*, 2866−2871.
26. 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.
27. Stadtman, T. C. (1996) *Annu. Rev. Biochem. 65*, 83−100.
28. Zhong, L., Arner, E. S., and Holmgren, A. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 5854−9.
29. Takebe, G., Yarimizu, J., Saito, Y., Hayashi, T., Nakamura, H., Yodoi, J., Nagasawa, S., and Takahashi, K. (2002) *J. Biol. Chem. 277*, 41254−8.

BI0346300