

Strain-based Sequence Variations and Structure Analysis of Murine Prostate Specific Spermine Binding Protein Using Mass Spectrometry[†]

Pierre Chaurand,[‡] Beverly B. DaGue,[‡] Shuguang Ma,^{‡,§} Susan Kasper,^{||} and Richard M. Caprioli^{*,‡}

Department of Biochemistry and the Mass Spectrometry Research Center, Departments of Urologic Surgery and Cell Biology and the Vanderbilt Prostate Cancer Center, Vanderbilt University, Nashville, Tennessee 37232-6400

Received March 1, 2001; Revised Manuscript Received May 10, 2001

ABSTRACT: Mouse spermine binding protein (SBP) has been characterized using mass spectrometry, including its localization within the prostate, sequence verification, and its posttranslational modifications. MALDI (matrix-assisted laser desorption/ionization) mass spectrometry was employed for localization of proteins expressed by different lobes of the mouse prostate obtained after tissue blotting on a polyethylene membrane. The mass spectra showed complex protein profiles that were different for each lobe of the prostate. The prostate-specific spermine binding protein (SBP), primarily identified by its in-source decay fragment ion signals, was found predominantly expressed by the ventral lobe of the prostate. The MALDI in-source decay measurements combined with nanoESI (nanoelectrospray ionization) MS/MS measurements obtained after specific proteolysis of SBP, allowed the exact positioning of a single N-linked carbohydrate group, and the identification of a pyroglutamate residue at the sequence N-terminus. The N-linked carbohydrate component was further investigated and the general pattern of the N-linked carbohydrate identified. The presence of a disulfide bridge between cysteine₇₈ and cysteine₁₂₄ was also established. The full sequence characterization of SBP showed several strain-based sequence differences when compared to the published gene sequence.

Matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS)¹ and electrospray ionization (ESI) MS have greatly facilitated the studies of protein structure and function over the past 10 years (1–5). These techniques now provide the means for sequencing femtomole amounts of peptides, measuring the molecular weights of proteins up to 200 kDa in mass, analyzing protein posttranslational modifications (e.g., phosphorylation and glycosylation), and elucidating specific binding of ligands to proteins as well as protein–protein interactions. Protein identification is routinely performed using MS analysis of protein digests to obtain sequences of fragment peptides followed by the query of protein databases (6–10). Consequently the identification of variations in sequences and posttranslational modifications of proteins of nominally similar function may be achieved relatively easily.

MALDI ion sources are typically coupled to time-of-flight (TOF) mass analyzers that have delayed extraction capabilities (11). Thus, ions formed after the desorption/ionization events are kept in the source regions for several hundreds of nanoseconds. Some of the molecular species will fragment during this period in a process termed in-source decay (12). For proteins, many sequence-specific fragment ions are formed by cleavage of the protein backbone, with the most frequent being *c*_n, *a*_n and *y*_n ions (for fragment ion nomenclature see refs 13 and 14). From the mass differences one can deduce the sequence of a portion of the protein, and in turn, this information can be used to query protein databases and identify the protein.

Although in-source decay MALDI MS can provide partial amino acid sequence determination, only a few examples of peptide and protein identification done in this manner have been reported. Katta et al. studied the in-source decay behavior of several recombinant proteins and were able to precisely locate posttranslational modifications such as disulfide bonds (15). Reiber and colleagues successfully sequenced several peptides based upon their in-source decay fragment ion spectra and identified several large proteins (16, 17). Lennon and Walsh obtained sequence information using in-source decay on several peptides and proteins and successfully located posttranslational modifications such as disulfide bonds and sites of phosphorylations (18–20). More recently, Muscat and colleagues used in-source fragmentations to characterize polyesteramides (21).

The use of MALDI MS for protein profiling allows one to generate highly reproducible and specific protein patterns from fresh tissue sections (22, 23). Briefly, tissue sections

[†] This work was supported by the National Institutes of Health (Grant GM 58008), the T. J. Martell Foundation and the Vanderbilt Ingram Cancer Center.

* To whom correspondence should be addressed. Phone: (615) 322-4336. Fax: (615) 343 8372. E-mail: r.caprioli@vanderbilt.edu.

[‡] Department of Biochemistry.

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

^{||} Departments of Urologic Surgery and Cell Biology and the Vanderbilt Prostate Cancer Center.

¹ Abbreviations: SBP, spermine binding protein; MALDI, matrix-assisted laser desorption/ionization; ESI, electrospray ionization; NanoESI, nanoelectrospray ionization; MS, mass spectrometry; TOF, time-of-flight; MS/MS, tandem mass spectrometry; MW, molecular weight; *m/z*, mass to charge; CHCA, α -cyano-4-hydroxycinnamic acid; DHB, 2,5-dihydroxybenzoic acid; SA, sinapinic acid; IAA, iodoacetamide; DTT, dithiothreitol; TFA, trifluoroacetic acid; HPLC, high-performance liquid chromatography

are contact-blotted to an organic polymer membrane, which captures proteins through hydrophobic and electrostatic interactions. The blotted areas are then coated with a matrix and analyzed by MALDI MS. Typically, from such blots over 200 individual protein signals can be detected in a mass range up to 100 kDa (24). In the current paper, we have used these methods to profile and compare the proteins expressed or secreted by the different lobes of the mouse prostate. Because of the lobe sampling process, both cytoplasmic and secretory proteins are observed (24). Prostate-specific mouse spermine binding protein (SBP) (25) was identified by its in-source decay ions and found to be expressed in high abundance in the ventral prostate lobe. After purification by HPLC, SBP was treated with proteases and the peptide fragments analyzed by nanoelectrospray (nanoESI) tandem mass spectrometry (MS/MS). These data were used to verify the primary sequence of some of these fragments and to highlight variations between the sequence of the studied strain (CD1) and the database reference SBP sequence (C57/BK). Further, we have demonstrated the presence of an N-terminal pyroglutamate residue as well as the position of an N-linked glycosyl group. The MS/MS measurements also revealed the presence of a disulfide bridge in the protein.

MATERIALS AND METHODS

Materials. Matrix compounds, α -cyano-4-hydroxycinnamic acid (CHCA), 2,5-dihydroxybenzoic acid (DHB), and 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid, SA) were obtained from Aldrich Chemical Co. (Milwaukee, WI). Peptides and proteins used as molecular weight calibration standards, ammonium bicarbonate, iodoacetamide (IAA), dithiothreitol (DTT), and endoproteinase Glu-C were obtained from Sigma Chemical Co. (St. Louis, MO). Biotech sequencing-grade 88% formic acid was obtained from Fisher (Fair Lawn, NJ) and sequencing-grade trifluoroacetic acid (TFA) was purchased from Burdick and Jackson (Muskegon, MI). HPLC-grade acetonitrile and methanol were EM Science Omnisolve brand (Merck KGaA, Darmstadt, Germany). Modified TPCK-treated porcine trypsin was obtained from Promega (Madison, WI), and N-deglycosidase PNGaseF from Prozyme (San Leandro, CA).

Preparation and MALDI MS Analysis of Tissue Blots. The prostate glands of five ten-week old male CD1 and C57/B6 mice (Harlan, Indianapolis, IN) were dissected and the different lobes carefully separated under a stereoscope. Animal sacrifice was performed under NIH guidelines. Special care was taken not to burst either the bladder or the seminal vesicle to avoid cross-contamination. The lobes were immediately placed on a carbon-imbedded polyethylene (C-PE) membrane (80 μ m thick, Goodfellow, Cambridge, England) mounted on a metal MALDI-MS sample plate using double-sided conductive tape (3M Co., Minneapolis, MN) (23). A glass microscope slide was placed on the tissue sections to achieve good contact between the tissue samples and the membrane and to avoid tissue dehydration. After 5 min at room temperature, the tissue sections were removed from the C-PE membrane and the blot areas were allowed to air-dry. Cellular debris, tissue fragments, physiological salts, and blood were then removed from the blot area by washing with 20 μ L water for about 30 s using an automatic pipet. The rinse water was retained for further analysis. The

blotted areas were air-dried and a 1- μ L drop of SA matrix (at 20 mg/mL in 5/5/0.01, V/V/V, acetonitrile/water/TFA) was applied to several regions of each blot. After drying under ambient conditions, analysis was performed using a Voyager MALDI DE-STR TOF MS (Applied Biosystems, Inc., Framingham, MA) equipped with a nitrogen (337 nm) laser. Positive ion mass spectra were acquired in the linear mode using a 25 kV accelerating potential under optimized delayed extraction conditions. Mass spectra were acquired and averaged from 1000–2000 laser shots randomly taken at various locations within a single matrix-coated spot (\sim 2 mm diameter). Mass calibration was achieved using the ions from α - and β -hemoglobins observed in all blots [α -hemoglobin, m/z 7498.5 ($z = 2$) and m/z 14 996.1 ($z = 1$); β -hemoglobin, m/z 7809.5 ($z = 2$) and m/z 15 617.9 ($z = 1$)]. Protein identification using amino acid sequence tags generated from in-source decay fragment ions was performed using the search engines available on the Proteomics LLC web site (www.proteomics.com) to search the NCBI nr protein database with the taxonomy *Mus musculus* (Proteomics, New York).

HPLC Purification of Spermine Binding Protein. Rinse water from several blots from the same tissue lobe was pooled and centrifuged at 16000g for 10 min to remove particulates. The cleared solution was fractionated at room temperature on a Narrowbore LC/MS C4 column (2.1 \times 250 mm, 5 μ m particle, 300 Å pore) from Vydac (Hesperia, CA) using an Agilent HP1100 HPLC system (Wilmington, DE). Elution solvents A and B were 0.1% TFA (aq) and 0.1% TFA in 95% acetonitrile (aq), respectively. Sample proteins were eluted from the column at a flow rate of 0.2 mL/min using the following gradient program: 5% B, 10 min isocratic; 5 to 60% B over 50 min, linear gradient; 60% B, 10 min isocratic; 60 to 80% B over 10 min, linear gradient; 80% B, 5 min isocratic. Protein peaks were detected by UV absorbance at 214 nm and collected manually in Eppendorf Safe-lock polypropylene microfuge tubes. HPLC fractions were lyophilized dry and reconstituted in 40 μ L 5/5/0.01 (V/V/V) acetonitrile/water/TFA for MALDI-MS analysis. A highly glycosylated protein subsequently identified as SBP eluted at 52–55 min (not corrected for system delay time).

Conditions for Enzymatic Digestion of HPLC-Purified SBP. The final concentrated HPLC fraction (elution time 52–55 min) was estimated to contain about 10 pmol of protein/ μ L on the basis of the UV absorbance signal. All digestions were carried out in 50 mM ammonium bicarbonate buffer (pH 7.8–8.0). Treatments were as follows: (1) *Trypsin digestion*. A 7.5- μ L portion of the concentrated HPLC fraction (approximately 20% of the total sample) was dried and reconstituted in 10 μ L buffer containing 0.05 μ g trypsin and the protein was digested for 6 or 24 h at 37 °C. (2) *Endoproteinase Glu-C digestion*. A 7.5 μ L portion of the concentrated HPLC fraction was dried and reconstituted in 10 μ L of buffer containing 0.05 μ g of enzyme. After 4.5 h reaction at 37 °C, another aliquot of 0.05 μ g of enzyme in 10 μ L buffer was added to the reaction and digestion was continued an additional 15.5 h. (3) *PNGaseF digestion* was used to remove N-linked carbohydrates from both a sample of the nonreduced protein as well as from a sample that had been preincubated with trypsin for 24 h. In each case, 1 unit of PNGaseF (reconstituted at 5 units/ μ L with water) was added to the sample for a 24 h digestion at 37 °C. (4)

NanoESI QTOF MS was performed on an MDS Sciex (Toronto, Canada) QStar instrument with a Protana (Protana A/S, Odense, Denmark) nanospray source. Samples were usually dissolved in methanol/water/88% formic acid (6/3.9/0.1, V/V/V). Positive ion spectra were acquired in the profile mode with an ion count accumulation time of 1 s, a TOF pulser frequency of 6.99 kHz, and pulse duration of 13 μ s. A stable spray was obtained at 800–900 V. For data acquisition in the TOF MS mode, the curtain and CAD nitrogen gas settings were 15 psi and 3–4 (arbitrary units), respectively. Peptide sequences were obtained from product ion spectra acquired for ions selected with Q1 and dissociated by increasing the CAD gas setting and Q0 offset potential. Protein spectra were deconvoluted using the Macintosh-based Q-Star software.

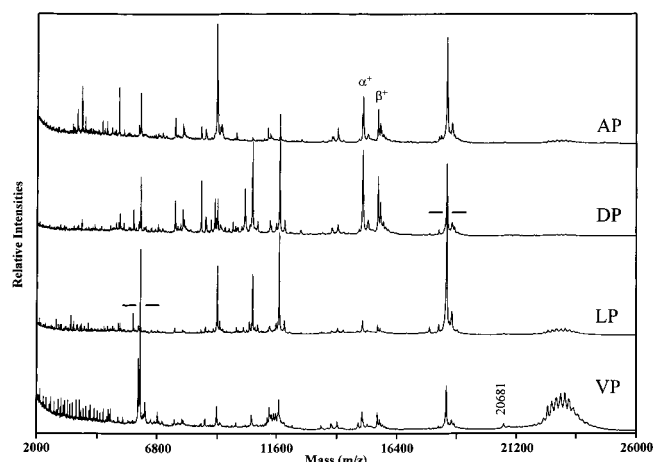


FIGURE 1: Protein profiles obtained by MALDI MS after fresh tissue blotting on a polyethylene membrane for the mouse anterior (AP), dorsal (DP), lateral (LP) and ventral prostate (VP) lobes. α^+ and β^+ are alpha and beta-hemoglobin, respectively.

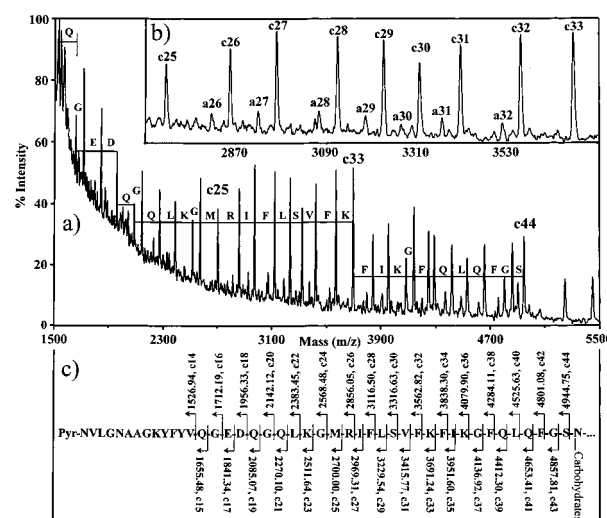


FIGURE 2: (a,b) Enlargement from Figure 1 of the protein profile obtained for the mouse ventral prostate lobe (VP) displaying in-source decay fragment ion signals. (c) Assignment of observed fragment ions. Numbering is relative to Swiss-Prot sequence P15501 for the processed mouse spermine-binding protein (SBP).

Protein Profiling and MALDI MS In-Source Decay. The

The mass spectra obtained from the ventral prostate also exhibited a unique set of low MW signals in the mass-to-charge (m/z) range from about 1500–5000 (Figure 2a). Because the mass differences between these low molecular weight signals corresponded to amino acid residue weights, the ions were assumed to be derived from in-source decay of a high abundance protein in the blot. Both c-type and

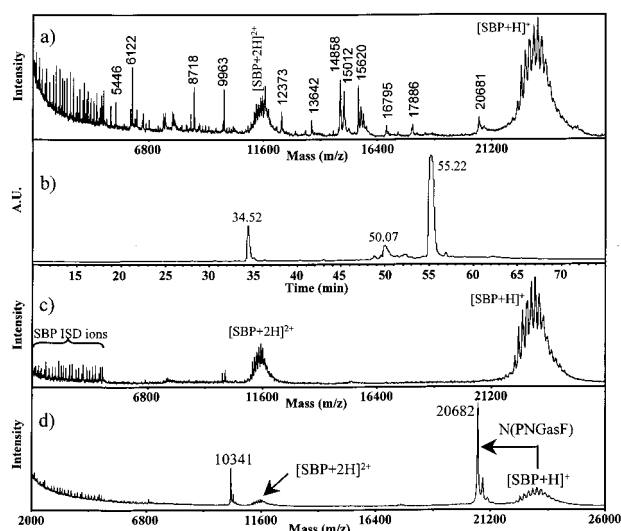


FIGURE 3: (a) Protein profile obtained by MALDI MS for the rinsing solution of the ventral prostate blot, (b) UV absorbance profile obtained from the HPLC separation of the proteins present in the ventral prostate lobe rinsing solution, (c) MALDI MS spectrum of the 55-min HPLC fraction, "SBP ISD ions" are spermine-binding protein in-source decay fragment ions, and (d) MALDI MS of the 55-min HPLC fraction after deglycosylation.

a-type in-source fragment ions were accounted for by the ion sequence (Figure 2b), and several sequence tags could be derived from the data. For example, $-F(I/L)SVFKF-$ was deduced from the ions in the range m/z 2969.3–3838.5. When used to query the NCBI database, "FLSVFKF" proved to be specific for SBP from mouse. A larger portion of the spectrum provided the amino acid sequence QGEDQGQLKGMRIFLSVFKFIKGFQLQFGS, corresponding to the segment [32–61] of pro-SBP. All of the observed ions were 111 Da greater than was expected based on the reported sequence of functional SBP, suggesting the presence of pyroglutamic acid at the sequence N-terminus, similar to that found in the rat SBP (26). These data are summarized in Figure 2c. The mass accuracy of the measured fragment ions was within one mass unit of the predicted average mass calculated for c-type N-terminal fragment ions. No in-source decay ion signals corresponding to other proteins were detected in the different protein profiles.

Interestingly, the in-source decay ion pattern abruptly terminated at m/z 4944.9, corresponding to the c-type ion formed after the amino acid serine in position 44 of SBP. The asparagine residue in position 45 of rat SBP had been shown to be glycosylated by Anderegg and co-workers (26). We, therefore, proposed that asparagine 45 of processed mouse SBP was N-glycosylated. No in-source decay ions with an ion distribution corresponding to the heterogeneous addition of sugar residues on SBP were detected.

For further studies, the protein was purified from the pooled water rinses of the ventral prostate tissue blot. As shown in Figure 3, HPLC fractionation of the pooled rinses yielded a highly purified preparation of SBP, confirmed by the intense signals for in-source decay ions. Treatment of the purified glycoforms with PNGaseF for 24 h converted approximately two-thirds of the total glycoform population to the deglycosylated protein at m/z 20 682 (± 2) (Figure 3d). It must be noted that a signal of low intensity at the same m/z value was also detected in the profile obtained from the

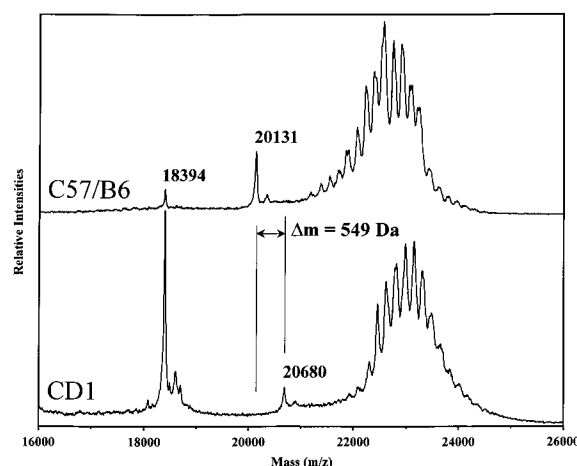


FIGURE 4: Partial protein profiles obtained by MALDI MS after fresh tissue blotting on a polyethylene membrane of the ventral prostate lobes from the CD1 and C57/B6 strains.

blot of the ventral lobe (Figure 1) and in its rinsing solution (Figure 3a). Upon deglycosylation of the glycoforms, the intensity of signals from in-source decay ions c14–c44 was reduced significantly in the MALDI MS. However, no strong signals corresponding to c45 and higher N-terminal derived in-source decay fragment ions were observed as the result of the production of carbohydrate-free SBP. These observations suggested that under the same experimental conditions, the carbohydrate-free protein undergoes very little in-source fragmentation and that the carbohydrate moiety plays an active role in the in-source decay charge remote fragmentation process(es). Other glycoproteins are being examined in order to determine if this is a general phenomenon or more specific to SBP.

Posttranslational Modification Characterization. The molecular mass observed for purified SBP after deglycosylation was 658 (± 2) Da greater than that expected based on the Swiss-prot database entry (P15501) for mouse SBP (shown in Figure 13). We had, however, used a different strain of mice (CD1) than the one used for the original sequence determination (C57/BK). The strain C57/BK could not be found but the very similar strain C57/B6 was investigated. Figure 4 compares the protein profiles obtained after blotting the ventral prostate from both the CD1 and C57/B6 strains. A mass difference of 549 Da was observed between these two strains indicating either sequence differences or different posttranslational modifications. The MW of the carbohydrate-free SBP from the C57/B6 mouse was found to be 111 (± 2) Da greater than that expected based on the SBP database entry potentially corresponding to the presence of an N-terminal pyroglutamic acid. This is also supported by the in-source decay profile observed for the C57/B6 SBP (not shown).

The strategy taken to characterize CD1 mouse SBP and to determine the source of the molecular weight difference was to utilize various proteases in combination with the N-deglycosidase PNGaseF to obtain peptide maps for the purified protein. Analysis of these peptide maps by MALDI MS and selected peptides by nanoESI QTOF MS/MS provided sufficient sequence information to account for the observed mass difference as described below and summarized in Table 1. Note that the amino acid residue numbers are based on the sequence results reported here.

Table 1: List of Peptides from SBP Expressed by CD1 Mice Sequenced by Mass Spectrometry

location	sequence	calcd mass (mono.)	MS/MS	ΔM (avg)	Figure
[1–10]	Pyr-NVLGNAAGK	953.5	477.8(+2)	+111.1	6
[11–23]	YFYVQGEDQGQLK	1573.7	787.9(+2)		
[24–26]	GMR	362.2	363.2(+1)		
[27–33]	IFLSVFK	852.5	427.3(+2)		
[34–36]	FIK	406.3	407.3(+1)		
[37–53]	<i>GFQLQFGSN*WTDVYGTR</i>	1974.9	659.6(+3) ^b		
[54–70]	SENFIDFLLEDGEHVIK	2004.0	669.0(+3)	+14	7
[71–88]	VEGSAVICLTSLTFTTNK	1883.0	971.0(+2) ^c	{–1}	8
[91–97]	VATFGVR	748.4	375.2(+2)		
[101–110]	YFSDTGGS DK	1075.4	538.7(+2)		
[111–126]	HLVTVNGMHAPGLCVK	1674.9	578.3(+3) ^c	–28, {–1}	8
[127–131]	GMGFK	538.3	539.3(+1)	+18	9
[132–149]	WGNINANGNDHYNNKEDK	2101.9	702.0(+3)		
[154–187]	DADNKDADNKDADNKDDGDEDDDDGNDDDDQKDES	3742.4	936.6(+4)	+543.6	
[148–187]	DKADNKDADNKDADNKDADNKDDGDEDDDDGNDDDDQKDES	4413.7	883.7(+5)	+543.6	10
[1–187]	carbohydrate free SBP (CD1 mouse)	20 678.5 (avg)		+658.7	13

^a ΔM represents the average mass difference for each peptide from the published sequence. ^b N[45] modified to D by deglycosidase in studied peptide. ^c Carbamidomethylation of cysteines in studied peptides. {Disulfide bridge subtracts 2 u}. Bold = sequence difference. Italic = sequence order not confirmed by MS/MS.

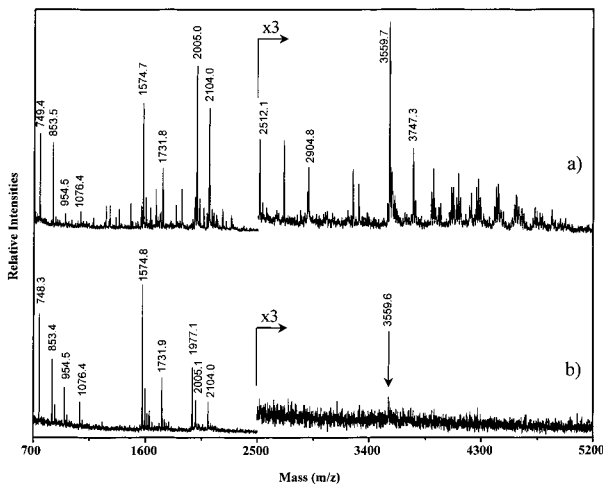


FIGURE 5: Peptide maps of SBP obtained by MALDI MS using 2,5-DHBA as matrix after (a) trypsin digestion and (b) trypsin plus PNGaseF digestion.

Glycosylation Site, Tryptic Peptide [37–53]. Figure 5 shows the MALDI MS analyses of the limit tryptic digests of SBP purified from CD1 mouse obtained both with and without the addition of PNGaseF. In the absence of PNGaseF the presence of glycopeptides was evident from the ion series observed from m/z 3700–5000 where many species appeared to be related by mass differences of 162 (hexose residue) or 203 (*N*-acetyl hexosamine residue) daltons. Addition of PNGaseF to the tryptic digest eliminated these species and no evidence was seen for the retention of O-linked glycopeptides (sugar residues not cleaved by the PNGaseF). A unique peptide at m/z 1977.1 appeared in the deglycosylated tryptic digest. NanoESI QTOF MS/MS of the triple-charge ion of this peptide at m/z 659.6 was performed (not shown) and the sequence of (GF)QLQFGSDWTDVYGTR expected from the SBP database sequence was confirmed. Note that the method does not distinguish leucine from isoleucine or lysine from glutamine, and that asparagine (N_{45}) would be converted to (D_{45}) during the enzyme cleavage of the peptide-carbohydrate linkage. These observations, in conjunction with the in-source decay results described above, supported the

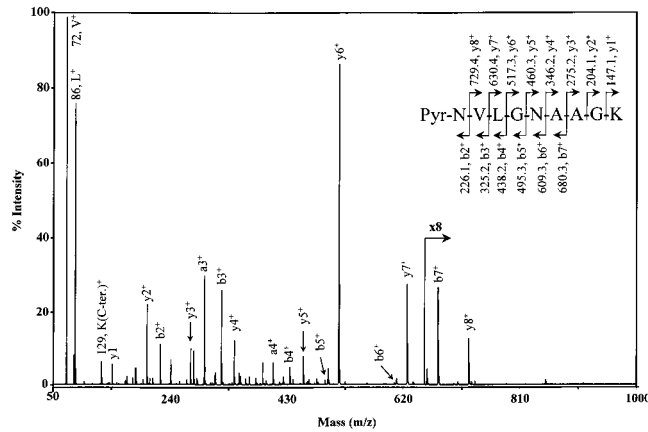


FIGURE 6: NanoESI QTOF MS/MS spectrum of SBP tryptic peptide [1–10] obtained from the double-charge ion detected at m/z 477.8.

conclusion that mature CD1 SBP had only a single N-glycosylation site at N_{45} .

***N*-Terminal Pyroglutamate, Tryptic Peptide [1–10].** The presence of *N*-terminal pyroglutamate would result in the tryptic fragment Pyr-NVLGNAAGK (monoisotopic molecular mass 953.5 Da). The MS/MS spectrum of the double-charge tryptic ion at m/z 477.8 obtained after HPLC fractionation of the digest is shown in Figure 6. This spectrum supported the proposed sequence. The *N*-terminal pyroglutamate residue was not directly recorded, due to the absence of b_1 and y_9 fragment ions. Nevertheless, the observed masses for the y_2 through y_8 fragments were within 0.002 (\pm 0.002) Da of the expected values. Comparison of the observed b fragments with those for the expected sequence and alternative isobaric sequences (QP)VLGNAAGK and (KP)VLGNAAGK showed average differences of only 0.003 (\pm 0.003) Da for the fit with the expected sequence, versus 0.037 (\pm 0.005) and 0.073 (\pm 0.005) Da for the respective alternative sequences.

Sequence Differences Found for CD1 SBP. Tryptic Peptide [54–70]. The predicted tryptic peptide ($[M + H]^+ = 1992.2$), was not observed. However, MALDI MS analysis showed a tryptic fragment at m/z 2005.0 (Figure 5). MS/MS fragmentation of the corresponding triple-charge ion

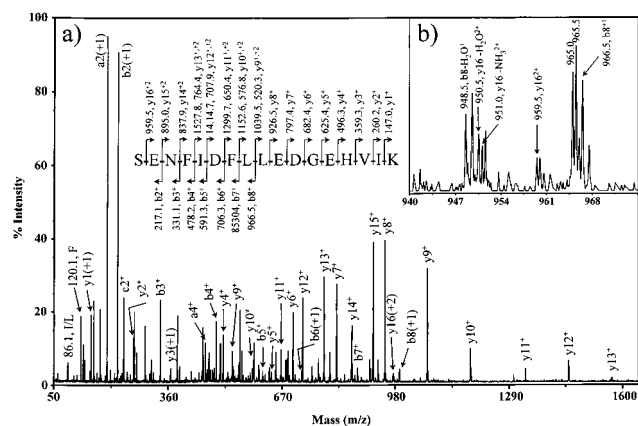


FIGURE 7: NanoESI QTOF MS/MS spectrum of SBP tryptic peptide [54–70] obtained from the triple-charge ion detected at m/z 669.0.

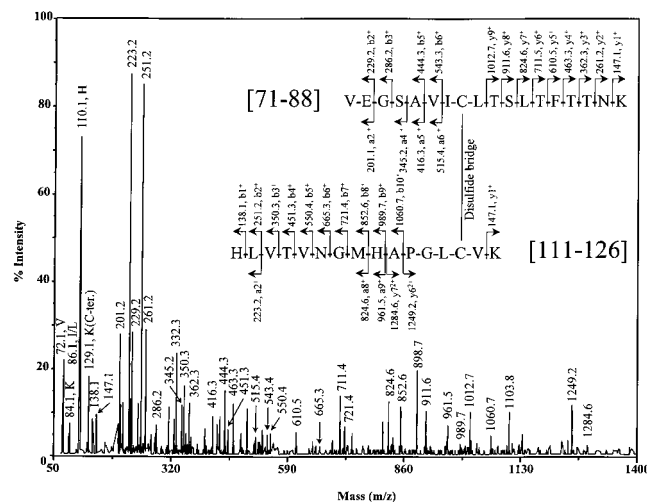
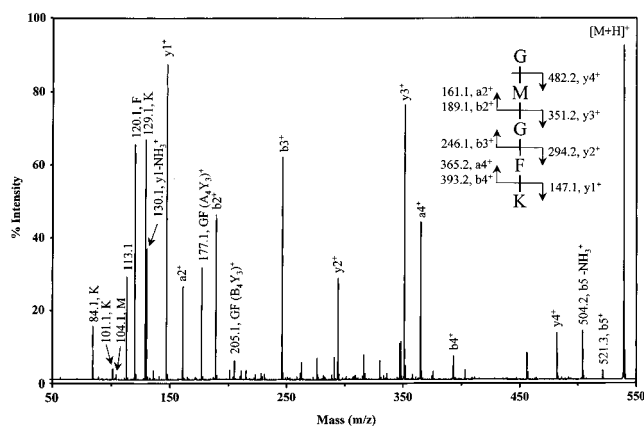
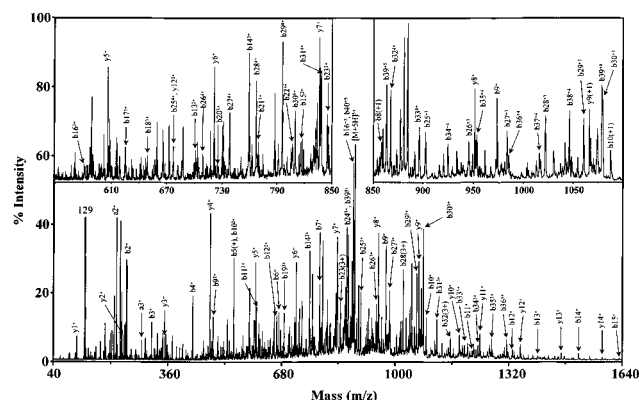


FIGURE 8: NanoESI QTOF MS/MS spectrum of SBP tryptic peptide [71–88]-S-S-[111–126] from the quadruple-charge ion observed at m/z 890.2.

observed in nanoESI QTOF MS at m/z 669.0, showed this peptide to be modified by the substitution of glutamic acid for aspartic acid at position 55 (Figure 7). This sequence change was confirmed by the presence of the b2 fragment ion (m/z 217.1), the double-charge y15 fragment ion (m/z 895.0), and the double-charge y16, y16-H₂O, and y16-NH₃ fragment ions (at m/z 959.5, 950.5, and 951.0, respectively). This modification of the protein resulted in an increase of the molecular weight by 14 Da.

Tryptic Peptides [71–88] and [111–126]. Each of these peptides contains a cysteine residue and they could possibly be linked to each other through disulfide bond formation. Neither of the individual peptides (expected monoisotopic $[M + H]^+$ values of 1884.0 and 1703.9, respectively) nor the disulfide-bridged peptide (average $[M + H]^+ = 3587.2$) was observed in the mass spectrum of the tryptic digest of nonreduced protein. However, an ion observed in the MALDI-MS spectrum at m/z 3558.4 (average mass) and as the quadruple-charge ion in the nanoESI spectrum at m/z 890.2, when fragmented by MS/MS, gave partial sequence coverage corresponding to the disulfide-bridged peptides. The fragmentation scheme is shown in Figure 8. The mass of the parent disulfide-bridged peptide was 28 Da lower than expected from the published sequence indicating a modification of one or both of the component peptides. Examination





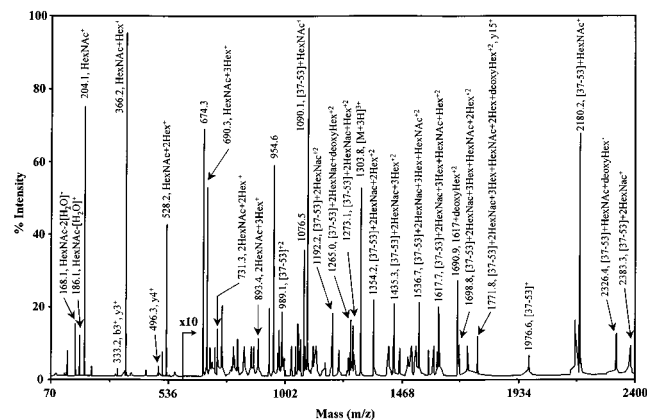


FIGURE 12: NanoESI QTOF MS/MS spectrum of the signal observed at m/z 3909.7 in Figure 11, corresponding to an SBP [37–53] tryptic glycopeptide. The MS/MS spectrum was obtained from the triple-charge ion detected at m/z 1303.8.

we have identified the presence of a pyroglutamate residue at the N-terminus of SBP. The in-source decay signals as well as the MS/MS data from the N-terminal tryptic peptide are in agreement with previous findings by Anderegg and co-workers who studied the rat SBP and reported the presence of a pyroglutamate at the N-terminal position of the processed SBP (26). The presence of a pyroglutamate residue at the N-terminal end of the processed protein formed by the rearrangement of the glutamine residue in position 18 of the proSBP sequence, also implies that the signal peptide of the proSBP consists of the amino acids from 1 to 17 and not from 1 to 18 as indicated in the published sequence (25). It is generally accepted that the presence of an N-terminal pyroglutamate residue inhibits or minimizes N-terminal protein degradation by exopeptidases, and in some cases has been shown to be important for enzymatic activity (28).

We have also identified the presence of a disulfide bridge in the sequence between cysteines 78 and 124 and a single N-glycosylation site on asparagine 45. The presence of the disulfide bridge also accounts for a difference of 2 mass units between the CD1 mouse carbohydrate free SBP and the calculated amino acid sequence (Swiss-Prot protein database, Primary Accession Number P15501). Table 1 lists the peptides from SBP expressed by CD1 mice, which were sequenced by mass spectrometry, and details the mass differences obtained after sequence corrections. These sequence variations account for a mass difference of +658.7 u (average mass), in agreement with the mass difference between the experimentally measured carbohydrate-free SBP and the published SBP sequence. On the basis of our measurements, Figure 13 presents the amino acid sequence of murine SBP as identified in the CD1 mouse strain compared to the sequence published by Mills et al. (25). Only three amino acids ($_{98}\text{RGR}_{100}$) out of 187 were not detected by MS/MS. However, the agreement (within the mass accuracy) between the measured ($\text{MW } 20\,680 \pm 2$) and the calculated mass ($\text{MW } 20\,678.5$) of the carbohydrate-free SBP (assuming no spontaneous rearrangement of N_{138} to D_{138}) suggests that this partial sequence is correct. The three amino acids, which were found different when comparing the CD1 and C57/BK SBP sequences, can all be explained by single point mutations in the SBP gene. $\text{Asp55}_{(\text{C57/BK})}$ to $\text{Glu55}_{(\text{CD1})}$ corresponds to the mutation of the codon $\text{GAT}_{(\text{C57/BK})}$ to

```

1  UNVLGNAAGKYFYVQGEDQGQLKGMRIFLS  30
   -nvlgnaagkyfyvqgedqgqlkgmrifls

31  VFKEIKGFLQFGSNWTDVYGTRSENFIDF  60
   vfkfikgflqfgsnwtdvygtrsdnfdif

61  LLEDGEHVIKVEGSAVICLTSLTFTTNKGR  90
   lledgehvikvegsavicltsltfttnkgr

91  VATFGVRRgrYFSDTGGSDKHLVTVNGMHA 120
   vatfgvrrgrYfSDtggSDkhlvtvngmha

121 PGLCVKGMGFKWGNINANGNDHYNNKEDKA 150
   pglcvrgigfkwgninangndhyNNkedka

151 DNKDADNKDADNKDADNKDDGDEDDDDGND 180
   dnk-----dadnkdadnkddgdedddgndd

181 DDQKDES
   ddqkdes

```

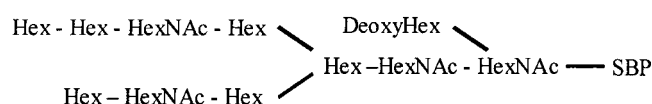
FIGURE 13: Single letter amino acid sequence of the murine spermine-binding protein (SBP) as identified in the CD1 mouse strain. U indicates the N-terminal pyroglutamate, bold characters indicate differences from the published sequence (small caps), characters in italics indicate the amino acid order was not confirmed by MS/MS, and lower case italics highlight residues not detected by MS. SBP possesses an N-glycosylation site at N_{45} and a disulfide bridge between C_{78} and C_{124} .

$\text{GA[A/G]}_{(\text{CD1})}$, $\text{Arg126}_{(\text{C57/BK})}$ to $\text{Lys126}_{(\text{CD1})}$ corresponds to the mutation of the codon $\text{AGA}_{(\text{C57/BK})}$ to $\text{AAA}_{(\text{CD1})}$, and finally, $\text{Ile128}_{(\text{C57/BK})}$ to $\text{Met128}_{(\text{CD1})}$ corresponds to the mutation of the codon $\text{ATT}_{(\text{C57/BK})}$ to $\text{ATG}_{(\text{CD1})}$. The codon sequences for the C57/BK strains were obtained from the SBP gene sequence published by Mills and colleagues (25). Because at this point we have only studied two forms of SBP, we cannot tell if these mutation sites are hyper-variable.

SBP, primarily expressed in the ventral lobe, was detected in its mature fully processed glycoforms. SBP was also detected in the other prostate lobes displaying the same isoforms but expressed at much lower intensities. A low abundance signal at m/z 20 681 corresponding to carbohydrate-free SBP, was observed in blots from the mouse ventral prostate, in the rinsing solution recovered after tissue blotting as well as in the 56-min HPLC fraction sampled on the tail end of the UV signal corresponding to SBP. These results suggest that SBP is present in the ventral prostate in both the glycosylated and unglycosylated isoforms. Because of the mass accuracy of ± 2 units on the measurement of the carbohydrate-free SBP, we cannot detect with certainty if the rearrangement of the asparagines 116 and 138 to aspartic acid detected on the [111–126] and [132–149] tryptic peptides, respectively, is taking place on the protein. However, it is noted that the N-deglycosylation reaction was performed in ammonium bicarbonate buffer potentially favoring these rearrangements.

Both MALDI and ESI MS have been used effectively for the study of glycoproteins (29–32). From the width of the signal distribution of the glycoforms, measured in both the MALDI and nanoESI MS protein profiles, it is estimated that the carbohydrate entity is composed of up to 20 sugar residues with a mean value centered on 14. From the MS/MS data obtained from the [37–53] glycopeptide of MW 3908.7 (average), a general structure for the carbohydrate entity of the molecule is proposed. We have used the

molecular weight of the carbohydrate entity of the [37–53] glycopeptide (MW 1930.7 + 18 = 1948.7, monoisotopic mass) to scan the GlycoSuite DB carbohydrate database (33) (www.glycosuite.com). Only one known N-linked carbohydrate structure composed of 11 sugar residues matching our experimental data was reported for several proteins in mammalian species. Bovine brain ribonuclease (34), porcine major seminal plasma glycoproteins PSP-I and PSP-II (35), and various mouse/human hybrid immunoglobulin G glycoproteins (36) have been found modified with N-linked oligosaccharides with this same sugar composition. The structure below is proposed as the structure of the carbohydrate entity of the [37–53] tryptic glycopeptide (corresponding to the SBP signal at MW 22 610). It is in agreement with the experimental MS/MS data and our proposed carbohydrate composition. From our data, however, we cannot identify the nature of the different carbohydrate residues nor their interconnecting chemistries.



One of the striking properties of SBP is its potential to generate N-terminal in-source decay ions in high abundance. Several structural features can lead to the favored formation of N-terminal in-source decay ions. In particular, the protein requires one or more potential ionization sites near or at its N-terminus (37, 38). The potential N-terminal ionization sites are essentially the protein N-terminus, and lysine and arginine (basic) amino acid residues. In the case of the mouse SBP protein, the N-terminal residue is a pyroglutamate, which because of its cyclic structure is not likely to capture the MALDI ionizing proton compared to a free N-terminal primary amine. This statement is supported by previous experimental data generated by Takayama and co-workers who have analyzed by MALDI in-source decay MS several peptides possessing a pyroglutamate N-terminal residue (37, 38). Considering the N-terminal sequence of SBP and the mass of the first observed in-source decay signal (at m/z 1527), the most probable location for the ionizing proton in the sequence is on the side chain of the lysine residue in position 10.

ACKNOWLEDGMENT

The mouse prostates were a kind gift from Dr. Robert Matusik of the Department of Urologic Surgery (Vanderbilt University). We also thank Dr. Robert Matusik for the use of his facility for animal storage and dissection.

SUPPORTING INFORMATION AVAILABLE

Supplemental material includes a figure presenting the nanoESI MS/MS spectrum of deglycosylated SBP [37–53] tryptic peptide obtained from the triple-charge ion detected at m/z 659.6 and a figure presenting the MALDI MS spectrum of the SBP [132–149] tryptic peptide, displaying monoisotopic signals at m/z 2102.9 and m/z 2103.9, suggesting the rearrangement of N₁₃₈ to D₁₃₈. This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES

1. Karas, M., Bachmann, D., Bahr, U., and Hillenkamp, F. (1987) *Int. J. Mass Spectrom. Ion Processes* 78, 53–68.

2. Hillenkamp, F., Karas, M., Beavis, R. C., and Chait, B. T. (1991) *Anal. Chem.* 63, 1193A–1202A.
3. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) *Science* 246, 64–71.
4. Roepstorff, P. (1997) *Curr. Opin. Biotechnol.* 8, 6–13.
5. Yates, J. R. (1998) *J. Mass Spectrom.* 33, 1–19.
6. Mortz, E., Vorm, O., Mann, M., and Roepstorff, P. (1994) *Biol. Mass Spectrom.* 23, 249–261.
7. Chen, Y. J., Wall, D., and Lubman, D. M. (1998) *Rapid Commun. Mass Spectrom.* 12, 1994–2003.
8. Chaurand, P., Luetzenkirchen, F., and Spengler, B. (1999) *J. Am. Soc. Mass Spectrom.* 10, 91–103.
9. Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., and Yates, J. R. (1999) *Nat. Biotechnol.* 17, 676–682.
10. Borchers, C., Peter, J. F., Hall, M. C., Kunkel, T. A., and Tomer, K. B. (2000) *Anal. Chem.* 72, 1163–1168.
11. Brown, R. S., and Lennon, J. J. (1995) *Anal. Chem.* 67, 1998–2003.
12. Brown, R. S., and Lennon, J. J. (1995) *Anal. Chem.* 67, 3990–3999.
13. Roepstorff, P., and Fohlmann, J. (1984) *J. Biomed. Mass Spectrom.* 11, 601.
14. Biemann, K. (1990) *Methods Enzymol.* 193, 455–479.
15. Katta, V., Chow, D. T., and Rohde, M. F. (1998) *Anal. Chem.* 70, 4410–4416.
16. Reiber, D. C., Brown, R. S., Weinberger, S., Kenny, J., and Bailey, J. (1998) *Anal. Chem.* 70, 1214–1222.
17. Reiber, D. C., Grover, T. A., and Brown, R. S. (1998) *Anal. Chem.* 70, 673–683.
18. Lennon, J. J., and Walsh, K. A. (1997) *Protein Sci.* 6, 2446–2453.
19. Walsh, K. A., and Lennon, J. J. (1997) *FASEB J.* 11, 2955.
20. Lennon, J. J., and Walsh, K. A. (1999) *Protein Sci.* 8, 2487–2493.
21. Muscat, D., Henderickx, H., Kwakkenbos, G., van Benthem, R., de Koster, C. G., Fokkens, R., and Nibbering, N. M. M. (2000) *J. Am. Soc. Mass Spectrom.* 11, 218–227.
22. Caprioli, R. M., Farmer, T. B., and Gile, J. (1997) *Anal. Chem.* 69, 4751–4760.
23. Chaurand, P., Stoeckli, M., and Caprioli, R. M. (1999) *Anal. Chem.* 71, 5263–5270.
24. Todd, P. J., Schaaff, T. G., Chaurand, P., and Caprioli, R. M. (2001) *J. Mass Spectrom.* 36, 355–369.
25. Mills, J. S., Needham, M., and Parker, M. G. (1987) *Nucleic Acids Res.* 15, 7709–7724.
26. Anderegg, R. J., Carr, S. A., Huang, I. Y., Hiiipakka, R. A., Chang, C., and Liao, S. (1988) *Biochemistry* 27, 4214–4221.
27. Brennan, T. V., and Clarke, S. (1995) in *Deamidation and Isoaspartate Formation in Peptides and Proteins* (Aswad, D. W., Ed.) pp 65–90, CRC Press, Boca Raton, FL.
28. Abraham, G. N., and Podell, D. N. (1981) *Mol. Cell. Biochem.* 38, 181–190.
29. Tsaropoulos, A., Bahr, U., Pramanik, B. N., and Karas, M. (1997) *Int. J. Mass Spectrom.* 169, 251–261.
30. Nemeth, J. F., Hochensang, G. P., Marnett, L. J., and Caprioli, R. M. (2001) *Biochemistry* 40, 3109–3116.
31. Harvey, D. J. (2000) *J. Am. Soc. Mass Spectrom.* 11, 900–915.
32. Harvey, D. J. (2000) *J. Mass Spectrom.* 35, 1178–1190.
33. Cooper, C. A., Harrison, M. J., Wilkins, M. R., and Packer, N. H. (2001) *Nucleic Acids Res.* 29, 332–335.
34. Katoh, H., Ohgi, K., Irie, M., Endo, T., and Kobata, A. (1990) *Carbohydr. Res.* 195, 273–293.
35. Nimtz, M., Grabenhorst, E., Conradt, H., Sanz, L., and Calvete, J. J. (1999) *Eur. J. Biochem.* 265, 703–718.
36. Lund, J., Takahashi, N., Nakagawa, H., Goodall, M., Bentley, T., Hindley, S. A., Tyler, R., and Jefferis, R. (1993) *Mol. Immunol.* 30, 741–748.
37. Takayama, M., and Tsugita, A. (1998) *Int. J. Mass Spectrom.* 181, L1–L6.
38. Takayama, M., and Tsugita, A. (2000) *Electrophoresis* 21, 1670–1677.

BI010424L