Investigation of Angiotensin Glycosylation by MALDI-TOF and ESI Tandem Mass Spectrometry

**Soo-Jin Park, Deok-Hie Park,[a] Soohwan Sul,[b] Sunghwan F. Oh,[c] In-Sook Park,
Doo Soo Chung, Hie-Joon Kim,[*] Min-Sik Kim,[†] and Sang-Won Lee[†]**

School of Chemistry, Seoul National University, Seoul 151-747, Korea
[†]Department of Chemistry and Center for Electro- & Photo-Responsive Molecules,
Korea University, Seoul 136-701, Korea

Angiotensin I, a model decapeptide, was glycosylated and partially hydrolyzed with HCl (6 N, 80 °C, 4 h), aminopeptidase, and carboxypeptidase Y. A single peptide mass map obtained from truncated peptides in the partial acid hydrolysate of angiotensin and its glycosylation product mixture by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry enabled sequencing of angiotensin by a combinatorial procedure. MALDI-TOF and electrospray ionization (ESI) tandem mass spectrometric results indicate that both the N-terminal amino group of aspartic acid and the guanidinium group of the second residue arginine are glycosylated.

Key Words : Angiotensin, Glycosylation, Partial hydrolysis, MALDI-TOF MS, ESI MS/MS

Introduction

Proteins in eukaryotic cells are often modified after translation. Therefore, it is of great interest to determine the nature and the site of modifications. Mass spectrometric peptide sequencing techniques could be utilized, in principle, to investigate modification of amino acid residues. Electrospray ionization (ESI) tandem mass spectrometry (MS/MS)[1-4] is greatly useful for sequencing peptides. Tandem mass spectrometry following matrix-assisted laser desorption/ionization (MALDI) is becoming an alternative technique.[5]

In fact, identification of modification site by tandem mass spectrometry is not always straightforward. Suppose that the N-terminal amino group of a peptide is glycosylated. Without the amino group for protonation, the glycosylated terminal amino acid generated in tandem mass spectrometry may not be easily ionized and detected by either MALDI or ESI mass spectrometry. Or the information about the glycosylation site could be lost if collisional fragmentation leads to detachment of the glucose.

In case tandem mass spectrometry does not provide conclusive results, it is desirable to have an alternative or a complementary method for investigating protein modifications. In fact, mass spectrometric peptide sequencing is feasible without MS/MS. In 1994, Knierman *et al.* showed that a series of peptides successively truncated by one amino acid from either the N- or C-terminus could be obtained by partial acid hydrolysis.[6] Production of such peptide ladders by acid hydrolysis as well as by carboxypeptidase or aminopeptidase could be utilized for C- or N-terminal sequencing.[7,8] Schrä showed that 80% of the sequence of mellitin, a 26 residue peptide, could be determined by MALDI-linear TOF MS of both the N- and C-terminal peptide ladder generated by 1 min hydrolysis at 110 °C with 3 M HCl.[9] Gobom *et al.* used partial acid hydrolysis in the vapor phase to obtain specific internal backbone cleavage information as well as a partial sequence. They identified several proteins based on such informations.[10]

Utilizing the high mass accuracy and resolution of MALDI-TOF MS with delayed extraction and reflectron, one could also sequence a peptide without necessarily generating a sequential ladder. All amino acids, except the leucine/isoluecine and lysine/glutamine pairs, have mono-isotopic molecular weight different by at least 0.04 Da. Therefore, the mass of a small peptide can be attributed to a limited number of amino acid combinations. The combination can be found by brute force combinatorics.[11-14]

Using a small model peptide, angiotensin I, we tested the feasibility of using the combinatorial sequencing method for identifying the glycosylation site of a peptide. If partial hydrolysis produced a peptide and its truncation product, the mass difference could be used to obtain a unique combination or a small number of combinations of amino acids. The amino acid composition could yield several candidate partial sequences. And the partial sequence informations could be used to elongate the tentative sequence and eventually to construct the full sequence. In this paper, we demonstrate how this combinatorial approach could be used to sequence angiotensin I, a decapeptide, and to investigate the glycosylation site. ESI tandem mass spectrometric investigation of the glycosylation site is also discussed.

*Corresponding author: e-mail: hjkim1@snu.ac.kr; Fax: +82-2-889-1568
[a]Current address: LG Chem, Ltd. Daejeon 305-380, Korea
[b]Current address: Department of Chemistry, University of California, Irvine, CA 92697-2055, USA
[c]Current address: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA

Experimental Section

Materials. Angiotensin I (hereafter denoted as angiotensin), aminopeptidase, carboxypeptidase Y (CPY), and pronase were purchased from Sigma (St. Louis, MO, USA). Immobilized pH gradient (IPG) strip for isoelectric focusing (IEF) was from Bio-Rad (Hercules, CA, USA). Deionized water was prepared using a Barnstead purification system (Dubuque, IW, USA). Other chemicals were reagent grade.

Glycosylation Reaction. A 0.6 mL aliquot of an aqueous solution containing angiotensin and glucose, both at 0.38 mM concentration, was heated in an open vial at 90 °C for 2 h using heating block.

Isoelectric Focusing. 20 μL of the glycosylation reaction mixture was mixed with 110 μL of rehydration buffer consisting of 8 M urea, 4% CHAPS, 0.2% Bio-Lyte 3/10, 50 mM dithiothreitol and loaded on the IPG strip. After rehydration for 12 h, IEF was carried out using a Protean IEF cell (Hercules, CA, USA). Following IEF, the strip was cut into 0.5 cm pieces and the peptides were extracted by vortex mixing for 20 min with 10 μL of 50% acetonitrile/5% formic acid.

Hydrolysis with Protease. The glycosylation reaction mixture was hydrolyzed with aminopeptidase (10%, w/w) or carboxypeptidase Y (10%, w/w) at room temperature. Glycosylated angiotensin separated by IEF was also hydrolyzed with CPY.

Partial Acid Hydrolysis. The glycosylation reaction mixture, containing both angiotensin and its glycosylation product, was mixed with an equal volume of 12 N HCl in a glass ampoule. The ampoule was sealed and heated at 80 °C for 4 h.

MALDI-TOF Mass Spectrometry. Voyager-DE STR Biospectrometry workstation with delayed extraction and reflectron capability (PE Biosystems, Framingham, MA, USA) was used. It was equipped with a 337 nm nitrogen laser and a 2 m flight tube. Mass spectra were obtained in the positive ion mode. 20 kV accelerating voltage was used.

One μL of the peptide digest was mixed with 9 μL matrix solution (10 mg α-cyano-4-hydroxycinnamic acid in 1 mL acetonitrile/water (3 : 2) containing 0.1% trifluoroacetic acid). 1 μL of this mixture was applied to the stainless steel plate. Gold plate was used with the acid hydrolysate.

Data Base Search. TagIdent in Protein identification and characterization on the ExPaSy molecular biology server (Swiss Institute of Bioinformatics, Switzerland, http://www.expasy.org) was used to perform search using a tag.

ESI Tandem Mass Spectrometry. ESI tandem mass spectrometry was performed using a hybrid quadrupole/orthogonal acceleration time-of-flight mass spectrometer (Q-TOF Ultima Global, Micromass, Manchester, UK). The instrument was operated in the positive ion nanoelectrospray mode; the capillary and cone voltages were +3.1 kV and +115 V, respectively. The source temperature was 85 °C. The cone gas was set to 150 L/hr and the nanoflow gas to 0.7 bar. A sample dissolved in 0.1 vol % formic acid in 50 : 50 v/v acetonitrile/water was directly infused with a microsyringe

pump at a flow rate of 100 nL/min. MS/MS spectra of angiotensin-glucose adduct ions of +1 or +2 charge states (m/z 1458.72 or 729.86) were obtained with a scan range of m/z 80-1500, a scan time of 0.9 s, and an interscan delay of 0.1 s. The collision gas was Ar. For calibration, a 2 μg/μL solution of sodium iodide with cesium iodide was used.

ESI MS/MS experiments were also performed using a quadrupolar ion trap mass spectrometer (LCQ advantage, Finnigan, CA, USA) equipped with a home-built nanoESI interface that utilizes a xyz-translational stage (460A series, Newport, Irvine, CA, USA) to precisely adjust the position of the spray emitter relative to the desolvation capillary. Electrospray solutions were prepared by diluting the glycosylation mixture to 0.0145 mg/mL using methanol/water/acetic acid (50 : 49 : 1, v/v/v). A syringe pump (KDS100, kdScientific, Holliston, MA, USA) continuously injects the electrospray solution through the spray emitter at 300 nL/min flow rate. 3.0 kV was applied to initiate sprays, which were viewed by a stereo microscope at 80-fold magnification (Olympus, Tokyo, Japan). No nebulizer or counter flowing drying gas was used and the desolvation capillary was operated at 275 °C. MS/MS experiments were performed with CID energy of 35%. Helium was the collision gas.

Results and Discussion

Glycosylation of Angiotensin. Condensation of glucose with an amine takes place *via* Schiff base formation and leads to a series of complex reactions collectively known as the Maillard reaction. There are two potential glycosylation sites on angiotensin, one on the N-terminus and another on the basic amino acid arginine. Based on preliminary experiments, glycosylation reaction condition was chosen that produced an angiotensin-glucose adduct with a molar ratio of one glucose per angiotensin. We assumed no prior knowledge of the preferred glycosylation site.

MALDI-TOF mass spectrum obtained from the glycosylation reaction mixture before heating showed a monoisotopic peak at *m/z* 1296.68 ([angiotensin+H]$^+$; expected, 1296.68). After 2 h heating at 90 °C, the mass spectrum showed, in addition to the angiotensin peak, a glycosylation peak at 1458.72, 162.04 higher (glucose minus water) than 1296.68. No peak was observed corresponding to addition of 2 molecules of glucose.

Figure 1 shows separation, by IEF, of angiotensin-glucose adduct from unmodified angiotensin followed by MALDI-TOF MS detection.[15] The mass spectrum of peptides extracted from the IPG strip in the pH 8.0-8.5 range showed only unmodified angiotensin at 1296.69. Only glycosylated angiotensin was observed at 1458.77 in the pH 7.0-7.5 range. The observed shift in the pI is consistent with modification of the basic amino group by glucose. Both species of angiotensin were observed at pH 7.5-8.0 showing incomplete separation.

Terminal Amino Acid Analysis. After digestion of the glycosylation reaction mixture, containing both angiotensin
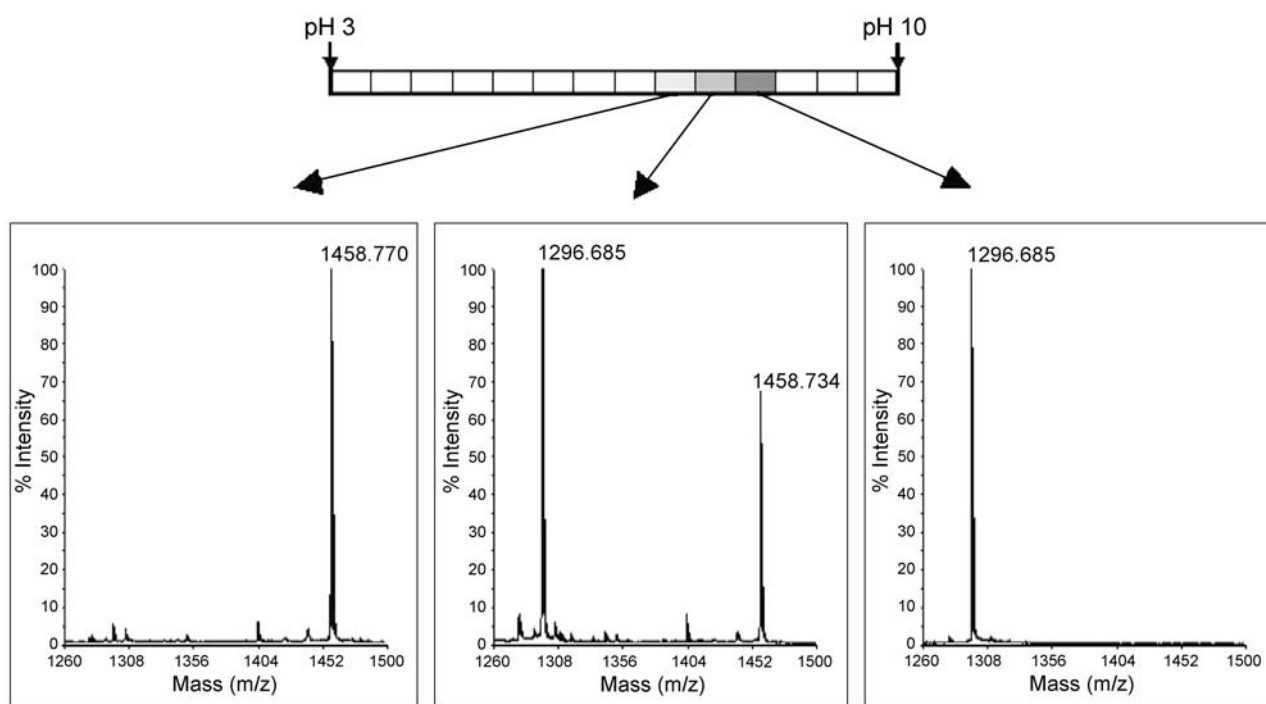
Figure 1. Separation of glycosylated angiotensin from unmodified angiotensin and MALDI-TOF mass spectrometric analysis of the separated peptides.

and its glucose adduct, with aminopeptidase for 3 min at room temperature, the parent and a truncated peptide were observed at 1296.69 and 1181.66, respectively (Fig. 2). The mass difference plus the mass of water is 133.04. Among the isotopic masses of 19 amino acids, the mass of aspartic acid (D, 133.03) is in agreement within 0.01 Da. Asparagine differs by 1 Da, which is much greater than the mass accuracy of the instrument. Therefore, D was identified as the N-terminal amino acid.

Figure 2 also shows a small peak at 1343.70 corresponding to removal of D from glycosylated angiotensin (1458.74). However, most of the glycosylated angiotensin remained unhydrolyzed. The intensity of glycosylated

angiotensin, which was initially lower than that of angiotensin, became higher than that of angiotensin. This reversal in intensity indicates that removal of the N-terminal amino acid by aminopeptidase proceeds much faster with unmodified angiotensin than with the angiotensin-glucose adduct. It is likely that glycosylation took place and introduced steric hindrance near the N-terminus.

The C-terminal sequence ladder in Figure 3, obtained after 3 min digestion of the glycosylation reaction mixture by CPY, showed that either leucine (L) or isoleucine (I) is the C-terminal amino acid. (Hereafter L is used to represent either L or I.) The mass spectrum in Figure 3 also shows additional C-terminal sequence (H, F, P). Terminal amino
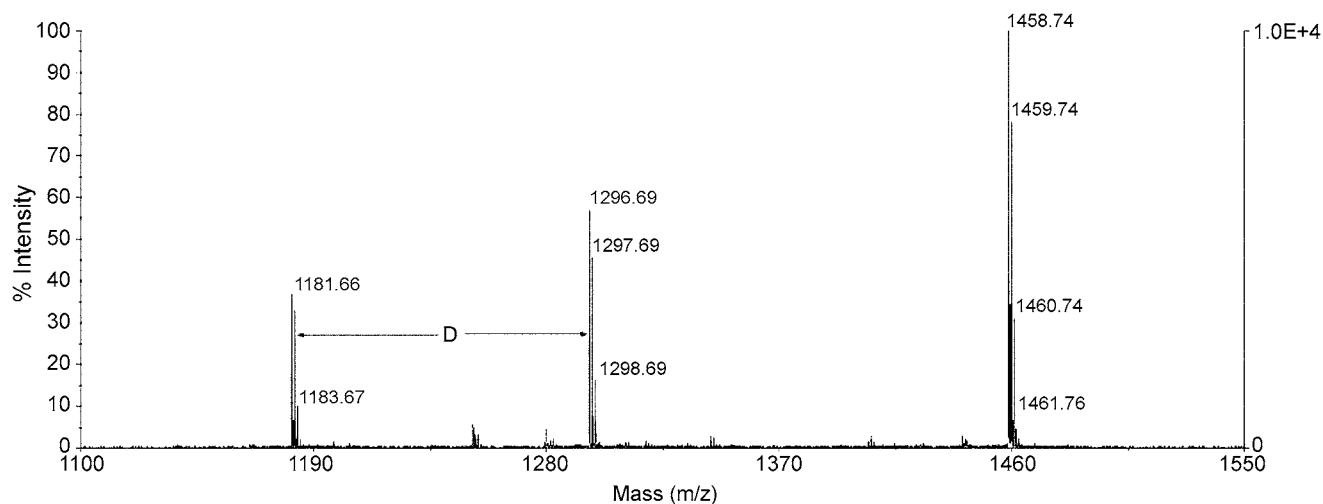

Figure 2. MALDI-TOF mass spectrum of aminopeptidase digest of the glycosylation reaction mixture showing undigested angiotensin-glucose adduct and truncation of the N-terminal aspartic acid.
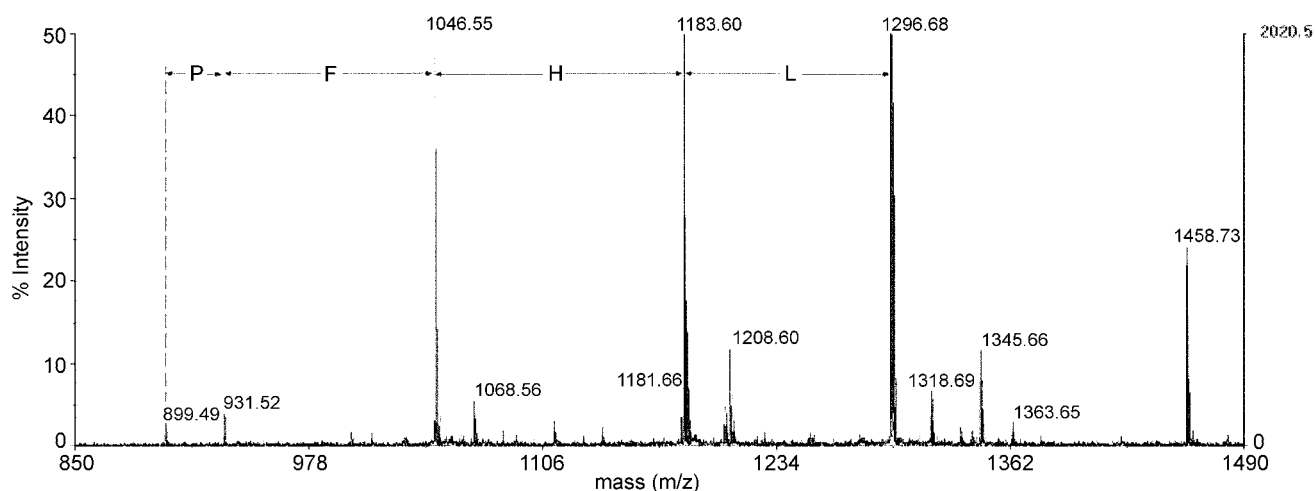
Figure 3. ALDI-TOF mass spectrum of carboxypeptidase Y digest of the glycosylation reaction mixture showing successive truncation of L, H, F, and P from the C-terminus from both unmodified angiotensin and the angiotensin α-glucose adduct.

acid analyses using aminopeptidase and CPY led to a tentative partial sequence, D( )PFHL.

In contrast to aminopeptidase digestion, the relative intensity of glycosylated and unglycosylated peptides was maintained at all stages of truncation. This observation shows that CPY digestion proceeds unhindered even in glycosylated angiotensin and is consistent with the amino-peptidase result, which suggested that the added glucose is near the N-terminus.

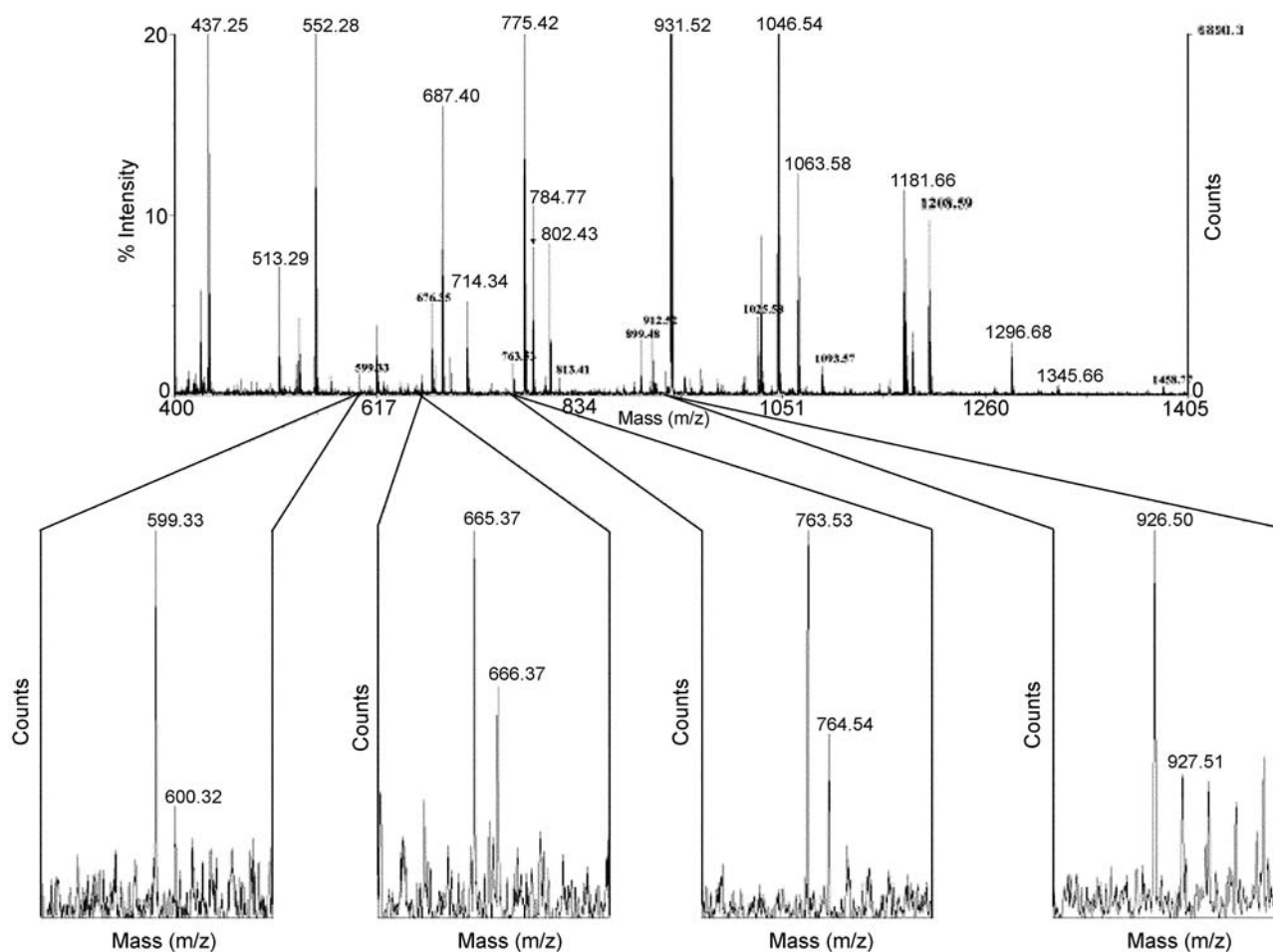**Sequencing the Peptide.** Figure 4 shows a MALDI-TOF

Figure 4. MALDI-TOF mass spectrum of partial acid hydrolysate (6 N HCl, 4 h at 80 °C) from the glycosylation reaction mixture.

mass spectrum obtained from a mixture of unmodified and glycosylated angiotensin partially hydrolyzed with 6 N HCl for 4 h at 80 °C. (Hydrolysis at 120 °C overnight leads to complete hydrolysis.) Several strong peaks were observed at *m/z* 1208.59, 1181.66, 1068.58, 1046.54, 1025.58, 931.52, etc. The small peak at 1296.68 corresponds to unhydrolyzed angiotensin. Other peptide fragments resulting from hydrolysis are also shown with varying signal intensity. Peaks separated by 162 Da were searched. The *m/z* values for five sets of peptides related by glycosylation are as follows: (1296.68, 1458.77), (1046.54, 1208.59), (931.52, 1093.57), (552.28, 714.34), (437.25, 599.33). The glycosylated peptides were set aside for subsequent identification of the glycosylation site. Unmodified peptides, which were more abundant and gave stronger signals, were used for sequence determination.

Generally, when sequencing a peptide using a peptide ladder generated by a protease or by collision as in MS/MS, the difference in mass of successive peaks is compared with the mass of the known amino acids. When dealing with a very complex mixture of peptides, some of which might

Table 1. Tentative sequence determined from the difference between masses of two peptides

starting m/z	aa	obs'd m/z	tentative sequence	
1296.68				
	D	1181.66	D̲( )	
	DL	1068.58	DL( ), D( )L, D( )DL, D( )LD	
	EV	1068.58	D( )EV, D( )VE	
	MP	1068.58	D( )MP, D( )PM	
	NN	1068.58	D( )NN	
	HL	1046.54	D( )HL, D( )LH	
			→ DL( )HL, DL( )LH, D( )HL	
	DR	1025.58	D̲R̲( )H̲L̲	
1181.66				
	L	1068.58	DR( )HL	
	R	1025.58	DR( )HL	
	GV	1025.58	→ inconsistent	
	CF	931.52	→ inconsistent	
	HL	931.52	DR( )HL	
	SY	931.52	→ inconsistent	
	RV	926.50	DRV̲( )HL	
	LR	912.52	DRV( )HL	
1068.58				
	H	931.52	DRV( )HL	
	AA	926.50	→ inconsistent	
	R	912.52	DRV( )HL	
	GV	912.52	→ inconsistent	
	RV	813.41	DRV( )HL	
	CY	802.43	→ inconsistent	
	EH	802.43	DRV( )EHL	→ undetermined
	FH	784.77	DRV( )FHL	→ undetermined
	HR	775.42	DRV( )HL	

Table 1. Continued

starting m/z	aa	obs'd m/z	tentative sequence
1046.54			
	D	931.52	DRV( )(E/F)HL
	F	899.48	DRV( )F̲HL
	DE	802.43	→ inconsistent
	FP	802.43	DRV( )PFHL
	LM	802.43	→ inconsistent
	DF	784.77	DRV( )PFHL
	DR	775.42	DRV( )PFHL
1025.58			
	V	926.50	DRV( )PFHL
	L	912.52	DRV( )PFHL
	DP	813.41	→ inconsistent
	LV	813.41	DRV( )PFHL
	CF	775.42	→ inconsistent
	HL	775.42	DRV( )PFHL
	SY	775.42	→ inconsistent
	DF	763.53	→ inconsistent
	MM	763.53	→ inconsistent
	VY	763.53	DRVY̲( )PFHL
931.52			
	E	802.43	→ inconsistent
	F	784.77	DRVY( )PFHL
	R	775.42	DRVY( )PFHL
	GV	775.42	→ inconsistent
	AP	763.53	→ inconsistent
	DE	687.40	→ inconsistent
	FP	687.40	DRVY( )PFHL
	LM	687.40	→ inconsistent
	RV	676.35	DRVY( )PFHL
	HL	676.35	DRVY( )PFHL
926.50			
	L	813.41	DRVY( )PFHL
	Y	763.53	DRVY( )PFHL
912.52			
	V	813.41	DRVY( )PFHL
	H	775.42	DRVY( )PFHL
	HV	675.35	DRVY( )PFHL
899.48			
	P	802.43	DRVY( )PFHL
	D	784.77	DRVY( )PFHL
	DP	687.40	DRVY( )PFHL
	HP	665.37	DRVY( )H̲PFHL
813.41			
	H	676.35	DRVY( )HPFHL
802.43			
	D	687.40	DRVY( )HPFHL
	H	665.37	DRVY( )HPFHL
	H	552.28	DRVY( )L̲HPFHL

show peaks with a low S/N ratio, it would be more straightforward to subtract the mass of the 19 amino acids or their combinations and look for their presence in the mass spectrum. In this work, we limited the consideration to truncation of 2 amino acid combinations. It turned out that full sequencing and identification of the modification site was possible using up to 2 amino acid combinations.

Combination of amino acids with similar combined mass include: (N, 132.05; GG, 132.05) (R, 174.11; GV, 174.10) (DL, 246.12; EV, 246.12; MP, 246.10; NN, 246.10) (FP, 262.13; LM, 262.14) (CF, 268.09; HL, 268.15; SY, 268.11) (DF, 280.11; MM, 280.09; VY, 280.14) (CY, 284.08, EH, 284.11). Standard notation for amino acids and their monoisotopic mass are used.

The sequencing steps are summarized below as well as in Table 1. No prior knowledge of the N- or C-terminal amino acid was assumed.

1. We started with the parent peptide with *m/z* 1296.68 for [M+H]$^+$. The parent peptide is designated as (1296.68) implying that there is no information about the amino acid residues at this point. Monoisotopic masses of 19 amino acids minus 18.01 (water) were subtracted from 1296.68. The only peak observed within 0.05 Da tolerance was 1181.66 (Fig. 4) corresponding to truncation of aspartic acid (D). D must be either the N- or the C-terminal amino acid. The amino acid residue identified at the particular step is underlined in Table 1. A separate experiment using amino-peptidase identified the N-terminus as D (Fig. 2). Even without such an information, one could arbitrarily assign D to the N-terminus and carry out sequencing, because the whole sequence could be reversed at the end in confirming the sequence from the data base. Thus, the tentative sequence of the parent peptide is designated as D(1181.66), where (1181.66) represents the truncated peptide with *m/z* 1181.66.

2. Subtraction of the masses of 19×19 combinations of 2 amino acids from 1296.68 resulted in 6 matches with the observed mass. Subtraction of DL matched 1068.58. Four possible sequences consistent with the previously assigned tentative sequence, D(1181.66), are listed in Table 1. DL could have been removed from the N-terminus, in which case L should be the second residue from the N-terminus. D and L could have been removed from the N- and C-terminus, respectively. The DL dipeptide could have been removed from the C-terminus. Four possible tentative sequences consistent with the previous assignment are listed. EV, MP, and NN combinations have the same mass as DL within 0.05 Da. Tentative sequences with D at the N-terminus are also listed for these combinations.

Another 2 amino acid truncation was observed at 1046.54 for histidine-leucine (HL) combination. Two tentative sequences, D( )HL and D( )LH, are possible. This assignment eliminates EV, MP, and NN combinations above. Only 3 tentative sequences, DL( )HL, DL( )LH and D( )HL, remain considering both DL and HL combinations.

Observation of a DR combination at 1025.58 clearly indicated that arginine (R) is the second residue from the N-

terminus, which in turn reduced the number of possible tentative sequences to only one, DR( )HL.

Analysis of the mass data so far clearly shows the value of the accurate mass measurement possible using high resolution MALDI-TOF mass spectrometer. The procedure could be extended to $19 \times 19 \times 19$ combinations of 3 amino acids. However, we wanted to see if the full sequence could be determined using 2 amino acid combinations.

3. The procedure was repeated starting from 1181.66, second highest in mass among peptides without glucose. Designation of the parent peptide as D(1181.66) is shown on the same row. Truncation of L, R, or HL was consistent with the tentative sequence, DR( )HL. Truncation of GV, CF, or SY was inconsistent with the tentative sequence and disregarded. Observation of RV truncation at 926.50 identified valine as the third amino acid from the N-terminus, leading to DRV( )HL. Truncation of LR corresponds to removal of R and L from both ends of (1181.66).

4. Truncation of H, R, RV, or HR from (1068.58) was consistent with the tentative sequence, DRV( )HL. Truncation of AA, GV, or CY was inconsistent and discarded. Observation of 802.43 (glutamic acid, E-histidine, H truncation) and 784.77 (phenylalanine, F-histidine truncation) opened two possibilities as tentative sequences, DRV( )EHL and DRV( )FHL.

5. Truncation of D from (1046.54) left the E/F question unresolved. However, observation of F truncation clearly showed that phenylalanine is the third amino acid from the C-terminus, because F cannot be the N-terminal amino acid. Observation of DF truncation confirmed the assignment. Observation of FP truncation at 802.43 added proline (P) as the fourth amino acid from the C-terminus. The 802.43 peak could also be interpreted as DE or LM truncation. However, they were inconsistent with the tentative sequence established and, therefore, discarded. The tentative sequence becomes DRV( )PFHL.

6. Repeating the procedure starting from 1025.58, tyrosine (Y) was identified, based on the observation of the 763.53 peak, as the fourth amino acid from the N-terminus. The tentative sequence is DRVY( )PFHL.

7. Repeating the procedure starting from 931.52, 926.50, and 912.52 only confirmed the sequence DRVY( )PFHL.

8. Additional information was obtained starting from 899.48. Observation of the 665.37 peak, corresponding to truncation of HP, was critical in identifying histidine (H) as the fifth amino acid from the C-terminus. The tentative sequence is DRVY( )HPFHL.

9. A peak observed at 552.28 showed truncation of HL from 802.43. Since it was established above that the 802.43 peptide has D as N-terminal amino acid, the HL combination has to be at the 5-6 position from the C-terminus. H was already identified as the fifth amino acid. Therefore, the sequence becomes DRVY( )LHPFHL.

10. At each step, the molecular weight of the amino acids in the tentative sequence was added and the resulting molecular weight minus the appropriate number of water molecules was compared with the molecular weight of the

parent peptide. This match was obtained with the sequence DRVY( )LHPFHL. Thus, the full sequence of the peptide was determined as DRVYLHPFHL.

11. Additional peaks confirmed the above sequence. Low molecular weight peptides were also observed corresponding to the truncation indicated by the mass difference.

Data Base Search. Determination of the full sequence by partial hydrolysis may not be always feasible even for a small peptide. Thus, the value of a partial sequence in identifying a peptide was tested. The *m/z* 802.43 peak corresponds to truncation of PFHL from the parent peptide. It has been established already that L not P is the terminal residue. It has also been established that the opposite terminus is D. If we did not know that D is the N-terminal amino acid, the parent peptide could be written as either (D802.43)PFHL or LHFP(802.43D). Observation of the 665.37 and 552.28 peaks indicated that HL is at the opposite end from D in the 802.43 peptide.

When peptides or proteins were searched using a PFHL sequence tag, 186 candidates were found in the entire pI and molecular weight range among 105667 peptides and proteins in the SWISS-PROT data base. Among the 186 candidates, only 5 had either LH or IH next to P. All of the 5 were angiotensins (human, horse, sheep, rat, mouse) with the DRVYIHPFHL sequence. Thus the search not only confirmed the sequence determined above but also identified the 5th residue as isoleucine not leucine. On the other hand, 140 candidates were found with the LHFP tag. Among the 140 candidates, none had the HL combination next to P. These search results show that an arbitrary assignment of a terminal residue to either N- or C-terminal at the beginning might be acceptable in many cases as long as there is sufficient sequence information.

Data base search was also conducted to see how the correct sequence could be identified when an ambiguity is encountered. For example, in step 4 above the C-terminal sequence was determined to be either FHL or EHL. When peptides up to 2000 Da with FHL sequence were searched, seven were found. Among the seven peptides, six were angiotensins. The remaining one had FHL in the middle and was discarded. Among the six angiotensins, only bovine angiotensin had valine instead of isoleucine as in the five angiotensins mentioned above. There were three peptides with EHL sequence. Only one had EHL on the C-terminus, but it had valine on the opposite end and was discarded. This observation suggests that a computer algorithm could be developed for selecting the correct residue when multiple options are encountered while sequencing a peptide by this combinatorial approach.

This example demonstrates that, utilizing the high mass accuracy and resolution of MALDI-TOF MS, one could sequence a peptide without necessarily generating a sequential ladder. All amino acids, except the leucine/ isoluecine and lysine/glutamine pairs, have monoisotopic molecular weight different by at least 0.04 Da. Therefore, the mass of a small peptide can be attributed to a limited number of amino acid combinations. The combination can

be found by brute force combinatorics.[11-14]

Identification of the Glycosylation Site. In general, lysine with an -amino group is known to be more reactive than other amino acids toward glycosylation. For example, addition of 7 glucose molecules to lysozyme, which has 7 lysine residues, has been reported.[15] Angiotensin does not have a lysine residue. From the sequence of angiotensin determined above, it appears that the likely glycosylation sites with basic group are the N-terminal aspartic acid and the arginine next to it.

The high intensity of the glycosylated angiotensin peak at m/z 1458.74 (Fig. 2) shows that glycosylated angiotensin is resistant to hydrolysis by aminopeptidase. This observation suggests that the N-terminal amino group is glycosylated, in which case aminopeptidase will not be able to recognize the terminal amino group and start hydrolyzing the peptide bonds from the N-terminus.

Figure 2 also shows a small peak at 1343.70 corresponding to removal of D from glycosylated angiotensin. In this case, clearly an amino acid residue other than D is glycosylated. Results to be discussed below show that this glycosylated residue is arginine (R). The intensities of the 1458.74 and the 1343.70 peaks do not necessarily reflect the abundance of glycosylated angiotensin (hereafter denoted as [1-10]' to signify modification of the peptide with residues 1-10) and glycosylated angiotensin with D removed, [2-10]'. It is well known that peptides with R at the C-terminus, obtained after tryptic digestion, are desorbed and ionized with high efficiency. Thus the 1458.74 peak might be intense due to the surviving R, whereas the 1343.70 peak is small due to the modification of R. The relative reactivity of the terminal amino group and arginine needs to be investigated further.

Two pairs of modified peptides related by the N-terminal aspartic acid (D) were observed from the acid hydrolysate (Fig. 4). The peak observed at 1208.59 was a glucose adduct of the 1-8 peptide ([1-8]'), and the 1093.57 peak was a glucose adduct of the 2-8 peptide ([2-8]'). Another pair was observed at 714.34 and 599.33 corresponding to [1-4]' and [2-4]', respectively. Glycosylated DR ([1-2]') and DRV ([1-3]') were observed at *m/z* 452.4 and 551.7, respectively, after 20 min CPY digestion of the glycosylated angiotensin
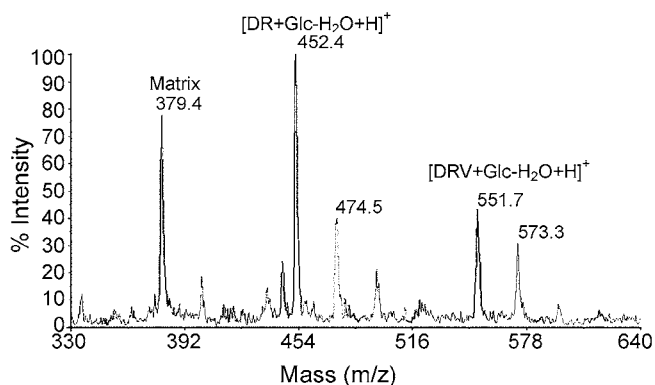

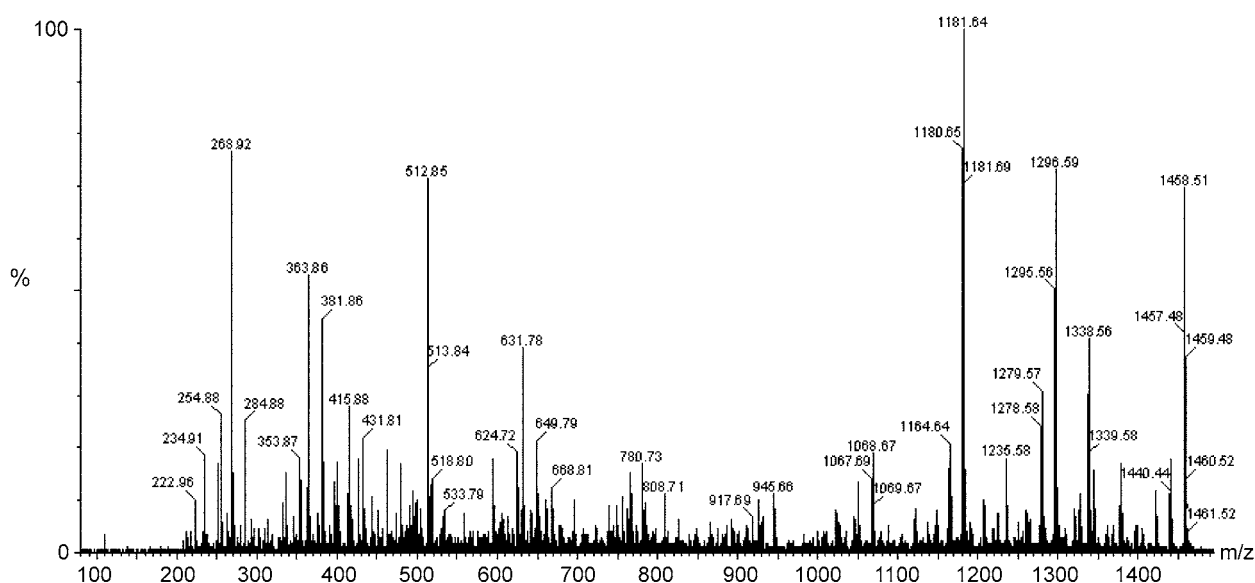**Figure 5**. MALDI-TOF mass spectrum of CPY digest of the isolated angiotensin-glucose adduct.

Figure 6. ESI Q-TOF tandem mass spectrum derived from glycosylated angiotensin ion.

separated by IEF (Fig. 5).[16] [1-2]' and [2-8]' were also observed upon pronase digestion of glycosylated angiotensin. The only residue common to [1-8]', [2-8]', [1-4]', [2-4]', [1-2]' and [1-3]' is R, the second residue. [2-8]' and [2-4]' could not be observed if only D is glycosylated. So there is no question about R being glycosylated to some extent. The results presented so far suggest that both D and R are glycosylated.

ESI Tandem Mass Spectrometry. Figure 6 shows a portion of the ESI Q-TOF mass spectrum obtained from the glycosylation reaction mixture. When glycosylated angiotensin, [1-10]', observed at *m/z* 1458.51 was subject to collisional fragmentation, angiotensin without glucose, [1-10], was observed at 1296.59 in addition to [1-10]'. [2-10], the y_9 ion, was also observed at 1181.64. The abundance of these [1-10] and [2-10] species without glucose indicates
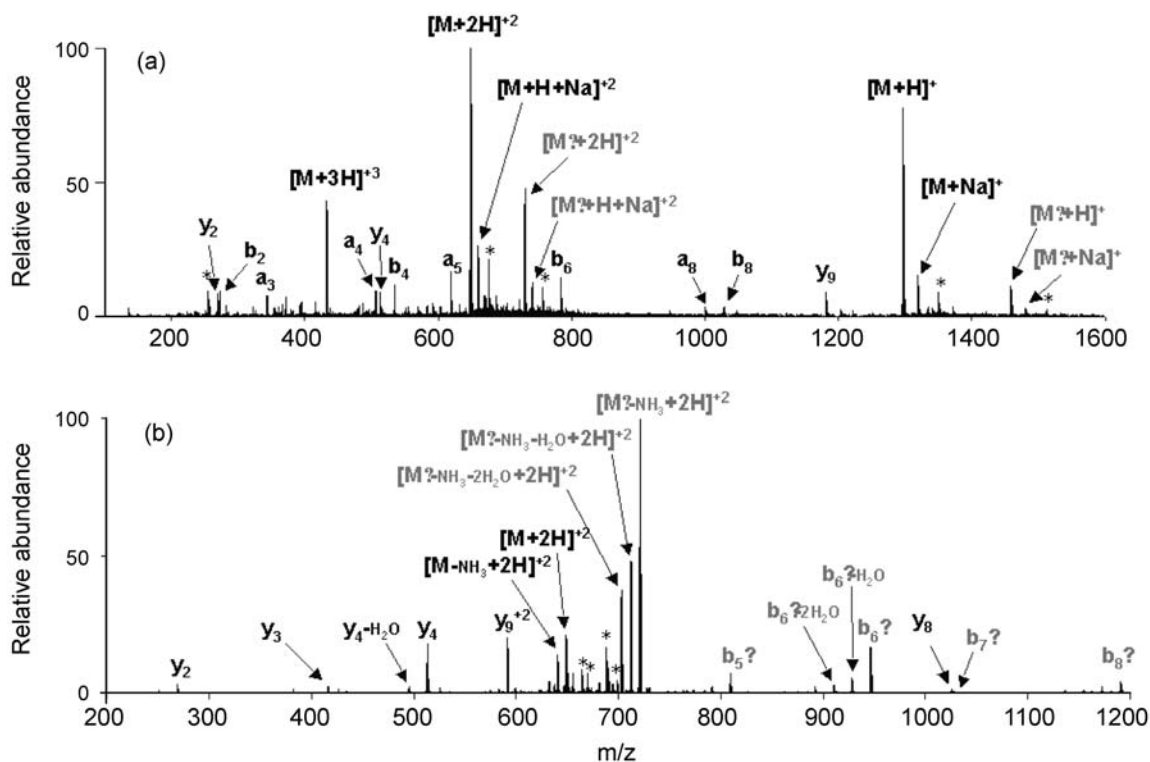

Figure 7. (a) ESI mass spectrum obtained from glycosylation reaction mixture. (b) Tandem mass spectrum of doubly protonated glycosylated angiotensin ion.
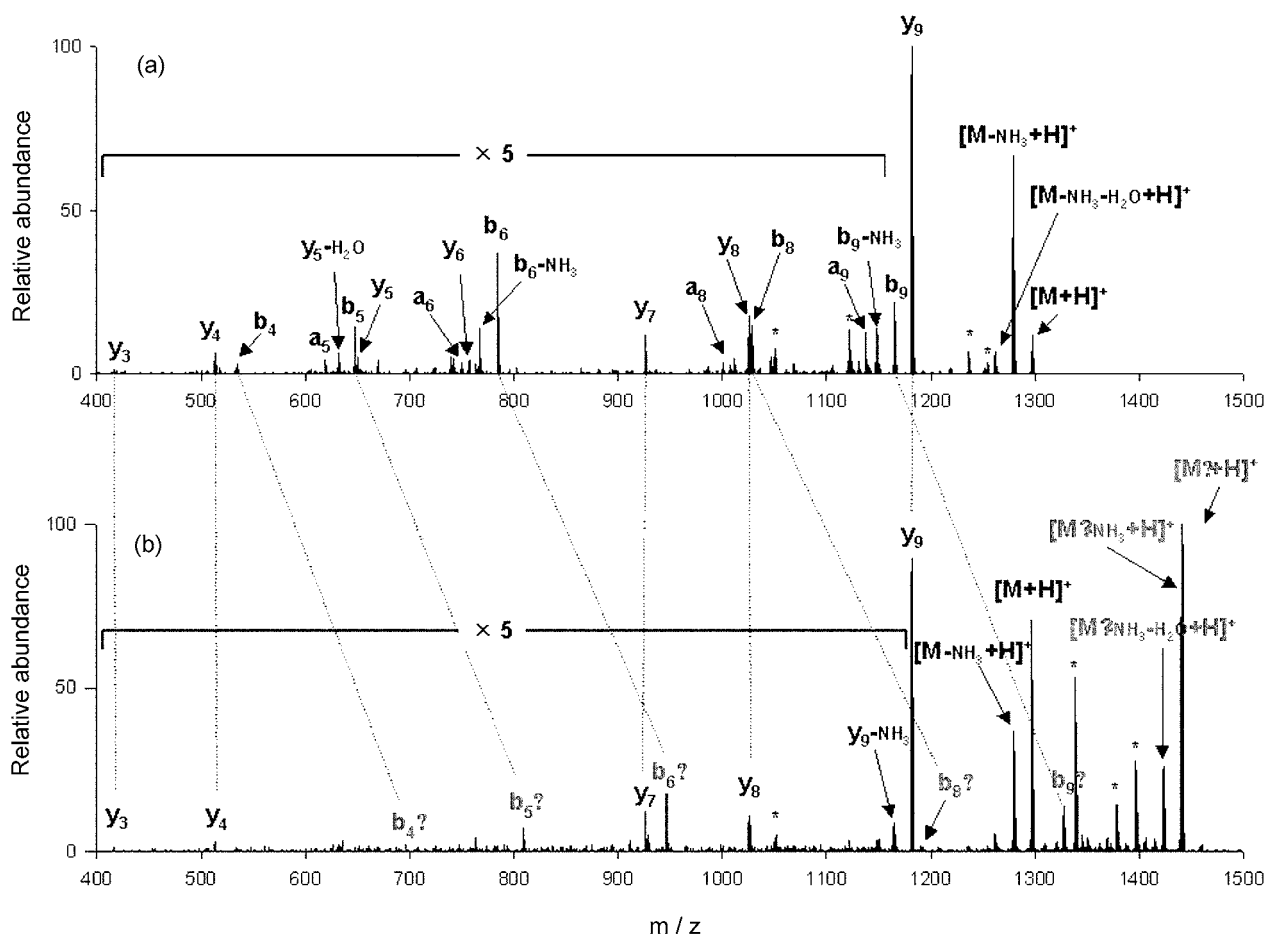
Figure 8. (a) MS/MS spectrum of singly protonated aniotensin. (b) MS/MS spectrum of singly protonated glycosylated angiotensin.

that glucose is quite readily detached upon collision with argon. Dissociation of the glucose will make identification of the glycosylation site by direct observation of the amino acid-glucose adduct by the MS/MS experiment difficult. Fragments retaining glucose were searched, and two such fragments were found. The 945.66 peak in Figure 6 corresponds to [2-7]' and supports glycosylation of R. Moreover, a small peak of glycosylated arginine was observed at 337.09 (expected m/z 337.16). Glycosylated aspartic acid was not observed. Glycosylated arginine has a free amino group which could facilitate ionization by protonation. However, glycosylated aspartic acid has the amino group modified by glucose. Therefore, the presence of glycosylated arginine and the lack of glycosylated aspartic acid should not be taken as an indication of R being more readily glycosylated than the terminal amino group in D.

Figure 7a shows ESI mass spectrum obtained from the glycosylation reaction mixture using the LCQ method. Singly and doubly protonated angiotensin ions were observed at m/z 1296.6 ($[M+H]^+$) and 648.9 ($[M+2H]^{+2}$), respectively, along with singly ($[M'+H]^+$, 1458.7) and doubly protonated glycosylated angiotensin ($[M'+2H]^{+2}$, 729.9). The peaks corresponding to the glycosylated angiotensin were observed with lower intensity compared to those

of unmodified angiotensin. This is consistent with the MALDI result. Figure 7b displays the low energy fragmentation pathways for the doubly protonated glygosylated peptide, $[M'+2H]^{+2}$. Glycosylated b_5' and b_6' were observed at m/z 809.3 and 946.3, respectively. This observation indicates that glucose is attached to one of the five residues from the N-terminus. Unmodified y_8 (VYIHPFHL) ion was also observed at m/z 1025.4. Since this ion was derived from glycosylated angiotensin ion, one could assume that either D or R is the glycosylation site. By the same token, observation of the y_9^{+2} ion could be taken as a sign of D being glycosylated. One should keep in mind though that y_9^{+2} could be observed if glucose were detached from R upon collision as evidenced by occurrence of $[M+2H]^{+2}$ in Figure 7b.

Figure 8 compares the MS/MS spectra of singly protonated angiontensin (Fig. 8a) and its glycosylated counterpart (Fig. 8b). It is evident from the comparison that all y-type ions (*i.e.*, y_9) from both have identical masses while the b-type ions (*i.e.*, b_6') of the glycosylated peptide are 162.1 Da higher in mass than those of unmodified peptide. Especially, the observation of y_9 implies that the N-terminal amino group is glycosylated. Figure 8b also shows $[M+H]^+$ derived from $[M'+H]^+$, upon collisional dissociation of the glucose. Under the experimental condition used for glycosylation, the

Soo-Jin Park et al.

guanidinium group of the arginine is likely protonated due to its high basicity while the N-terminal amino group is unprotonated. The pKa value of N-terminal α-ammonium group is *ca.* 7.8 while that of guanidinium group is *ca.* 16.5 [17]. Therefore, the N-terminal amino group is more likely to undergo glycosylation while the guanidinium group is protonated and less reactive toward glycosylation.

We note that the analysis of the tandem mass spectra would have been difficult if the sequence information obtained from the MALDI experiments was not available. With the sequence information the combined MALDI-TOF and ESI tandem mass spectrometric data could be used to confirm the involvement of N-terminal D and R in the glycosylation reaction. It would be of interest to compare the reactivity of the N-terminal amino group and the guanidinium group using a peptide where R is toward the C-terminus. This example based on a simple model peptide illustrates the value of having complementary methods, namely combinatorial sequencing using MALDI-TOF data and observation of peptide fragments by ESI tandem mass spectrometry in investigating the modification site of a peptide.

References

1. Morris, H. R.; Paxton, T.; Dell, A.; Langhorne, J.; Berg, M.; Bordoli, R. S.; Hoyes, J.; Bateman, R. H. *Rapid Comm. Mass Spectrom.* **1996**, *10*, 889-896.
2. Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M. *Rapid Comm. Mass Spectrom.* **1997**, *11*, 1015-1024.
3. Figeys, D.; Lock, C.; Taylor, L.; Aebersold, R. *Rapid Comm. Mass Spectrom.* **1998**, *12*, 1435-1444.
4. Borchers, C.; Peter, J. F.; Hall, M. C.; Kunkel, T. A.; Tomer, K. B. *Anal. Chem.* **2000**, *72*, 1163-1168.
5. Wattenberg, A.; Orgen, A. J.; Schneider, K. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 772-783.
6. Knierman, M. D.; Coligan, J. E.; Parker, K. C. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 1007-1010.
7. Patterson, D. H.; Tarr, G. E.; Regnier, F. E.; Martin, S. A. *Anal. Chem.* **1995**, *67*, 3971-3978.
8. Thiede, B.; Wittmann-Liebold, B.; Bienert, M.; Krause, E. *FEBS Letters* **1995**, *357*, 65-69.
9. Schrä, M. *Chimica* **2001**, *55*, 16-18.
10. Gobom, J.; Mirgorodskaya, E.; Nordhoff, E.; Hojrup, P.; Roepstorff, P. *Anal. Chem.* **1999**, *71*, 919-927.
11. Sakurai, T.; Matsuo, T.; Matsuda, H.; Hatakuse, I. *Biomed. Mass Spectrom.* **1984**, *11*, 396-399.
12. Ishikawa, K.; Niwa, Y. *Biomed. Environ. Mass Spectrom.* **1986**, *13*, 373-380.
13. Hamm, C. W.; Wilson, W. E.; Harvan, D. *J. CABIOS* **1986**, *2*, 115-118.
14. Zidarov, D.; Thibault, P.; Evans, M. J.; Bertrand, M. *J. Biomed. Environ. Mass Spectrom.* **1990**, *19*, 13-16.
15. Kim, H.-J.; Leszyk, J.; Taub, I. A. *J. Food Agric. Chem.* **1997**, *45*, 2158-2165.
16. Park, D.-H. *M.S. Thesis*, Seoul National University: Feb 2003.
17. Fukamizo, T.; Juffer, A. H.; Vogel, H. J.; Honda, Y.; Tremblay, H.; Boucher, I.; Neugebauer, W. A.; Brzezinski, R. *J. Biol. Chem.* **2000**, *275*, 25633-25640.