JOURNAL OF MASS SPECTROMETRY, VOL. 33, 358–362 (1998)

Localization of *O*-glycosylation Sites of MUC1 Tandem Repeats by QTOF ESI Mass Spectrometry

Franz-Georg Hanisch,[1] Brian N. Green,[2] Robert Bateman[2] and Jasna Peter-Katalinic[3]
[1] Institut für Biochemie, Universität Köln, Köln, Germany
[2] Micromass UK Ltd., Altrincham, UK
[3] Institut für Medixinische Physik und Biophysik, Universität Münster, Münster, Germany

The potential of electrospray mass spectrometry (ESMS) for the sequencing of glycopeptides was evaluated using quadrupole time-of-flight (QTOF) technology in the MS/MS mode. The location of *O*-glycosylation sites was possible in the positive ion (+) mode by detection of prominent y-and b-fragment ions from the underivatized TAP25-2 [T^{1}APPAHGVT^{9}S^{10}APDT^{14}RPAPGS^{20}T^{21}APPA], an overlapping sequence of MUC1 tandem repeats which had been glycosylated *in vitro* by two GalNAc residues in the positions T9 and T21. The high mass resolution and accuracy of QTOF-(+)ESMS allowed reliable structural assignments. The reduced complexity of the fragment spectra and the higher signal-to-noise ratio render QTOF-(+)ESMS an alternative mass spectrometric approach to the identification of *O*-glycosylation sites when compared with sequencing by post-source decay matrix-assisted laser desorption/ionization MS. Diagnostic ions from the *N*-terminus in the b-series offered direct evidence, which was supported by indirect evidence from the *C*-terminus ions of the y-series. The higher glycosylated $GalNAc_2$-substituted fragments were mainly observed as multiply ionized species. © 1998 John Wiley & Sons, Ltd.

J. Mass Spectrom. **33**, 358–362 (1998)

KEYWORDS: glycopeptide; *O*-glycosylation sites; MUC1; mucin; electrospray mass spectrometry

INTRODUCTION

Post-translational modification by *O*-glycosylation is a ubiquitous feature of secretory proteins[1] which has recently been recognized also in a selected array of nuclear and cytoplasmic proteins.[2,3] In particular, in the mucin-type glycoproteins the protein core serves as a scaffold for abundant *O*-glycosylation with a tremendous structural diversity, as shown for MUC1, the major membrane-associated mucin on human epithelia.[4]

Unlike *N*-glycosylation, the introduction of *O*-linked glycans into a peptide exhibits no strict dependence on specific motifs defining putative glycosylation sites and a reliable prediction of site-specific *O*-glycosylation can only be based on the analytical data from *in vivo*-processed mucins. Such a localization of *O*-glycosylation sites is conventionally performed by Edman degradation techniques identifying the glycosylated serine or threonine residues as blanks or positively by the chromatographic behaviour of phenylthiohydantoin-derivatized glycoamino acids.[5] Recently, a mass spectrometric (MS) approach has been established by applying the reflectron analysis of post-source decay (PSD) peptide fragments arising in a time-of-flight (TOF) analyser during matrix-assisted laser desorption/ionization (MALDI) MS[6] to the sequencing of *in vitro*-[7] or *in vivo*-glycosylated mucin peptides.[8] Various glycoforms of the MUC1 tandem repeat peptide were defined by this technique with regard to their substitution patterns using the abundant y-and b-series ions and their companion ions from the *C*-or *N*-terminus, respectively, for the localization of core-GalNAc —(*N*-acetylgalactosamine) residues. While PSD-MALDI/MS is highly sensitive, the complexity of the fragment spectra and limited mass accuracy gave rise to a search for an alternative mass spectrometric approach which would permit the sequencing of high-mass glycopeptides. We have demonstrated previously the potential of (+)-and (−)ESMS for the analysis of non-derivatized *O*-glycosylated amino acids and peptides.[9,10]. In addition to molecular ions of high abundance, diagnostic fragment ions arising from the sugar portion and the sugar peptide link were formed on analysis of mono-or multiply glycosylated mono-to tripeptides.[9,10] The potential of (+)ESMS on triple quadrupole and ion trap mass spectrometers was demonstrated for sequencing of synthetic *O*-glycopeptides corresponding to the MUC4 tandem repeat.[11]

In the present study, we assessed the capacity of (+)ES-tandem MS (MS/MS) on a quadrupole/time-of-flight (QTOF) hybrid mass spectrometer to sequence a glycosylated 25-mer peptide based on the tandem repeat of human MUC1 mucin. The substitution sites of two GalNAc (*N*-acetylgalactosamine) residues within the sequence TAPPAHGVTSAPDTRPAPGSTAPPA (TAP25) had previously been investigated by Edman degradation combined with liquid secondary ion MS (LSIMS) of proteolytic fragments[12] and by PSD-MALDI/MS.[7]

* Correspondence to: J. Peter-Katalinic, Institut für Medizinische Physik und Biophysik, Universität Münster, Robert-Koch-Str. 31, D-48149 Münster, Germany
E-mail: jkp@uni-muenster.de

CCC 1076–5174/98/040358–05 $17.50

Received 6 October 1997
Accepted 12 January 1998

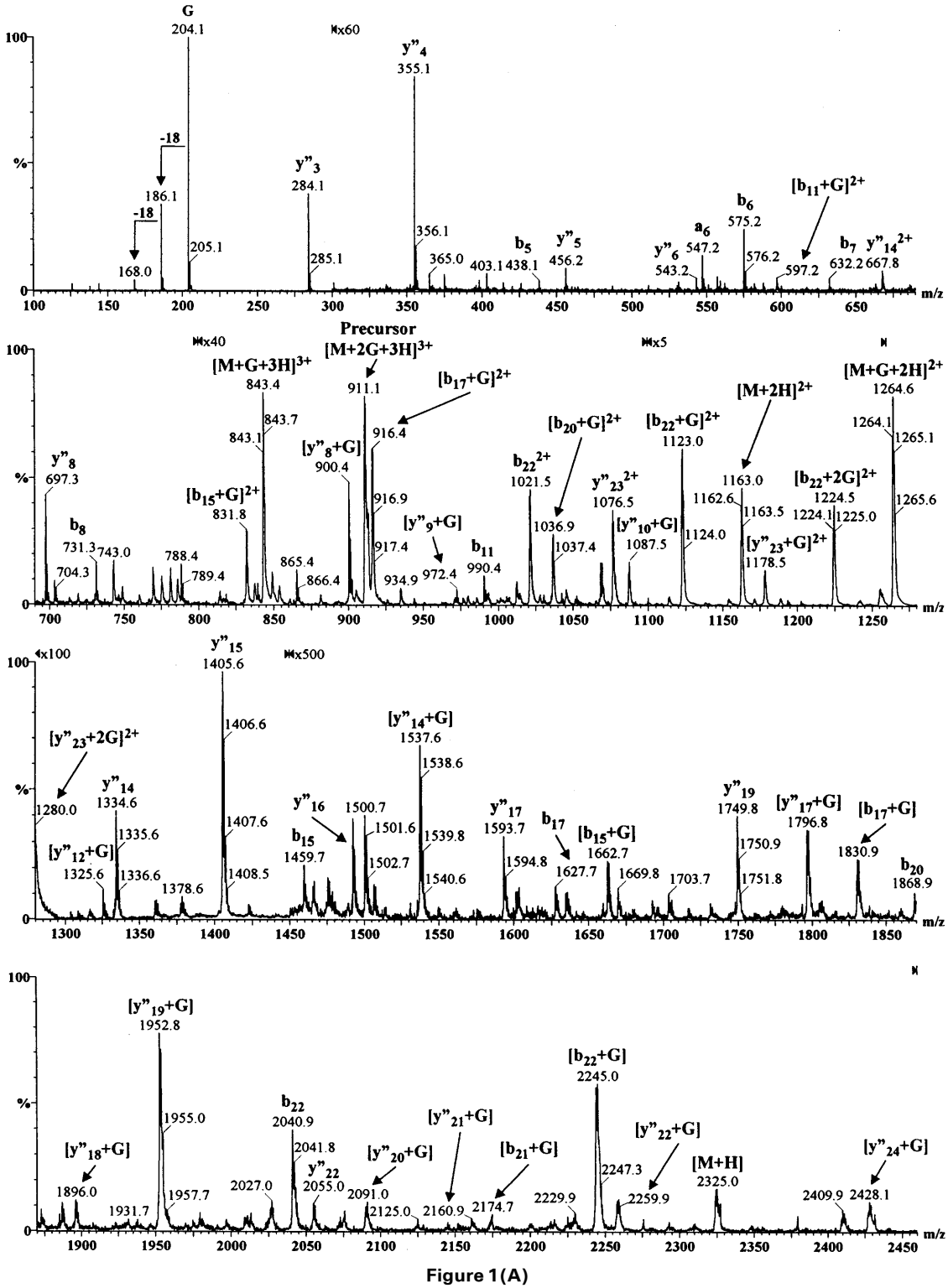

Figure 1(A)

Figure 1. (A) (+)ESMS/MS of the GalNAc-disubstituted TAP25-2 peptide performed in the QTOF mode using *m/z* 911 as a precursor. The fragment ions from the *C*-terminal y_n and *N*-terminal b_m ions are in accordance with the peptide sequence listed in Table 1 and assigned regarding their status of glycosylation. G = GalNAc. Ions of low abundance are not assigned. (B) Enlarged detail from (A) featuring the region of the doubly and triply charged ions in the mass range *m/z* 735–860.

EXPERIMENTAL

The synthetic peptide TAP25, kindly provided by Dr J. Taylor-Papadimitriou, (Imperial Cancer Research Fund, London, UK), had previously been glycosylated *in vitro* using UDP-GalNAc and a preparation of UDP-GalNAc: peptide *N*-acetylgalactosaminyltransferases from human premature skim milk.[12] The sites of GalNAc substitution had been investigated by Edman degradation combined with LSIMS of proteolytic fragments[12] and by PSD-MALDI/MS[7] after separation

by reversed-phase high-performance liquid chromatography of the diglycosylated species TAP25-2. According to these analyses, TAP25-2 is glycosylated at Thr9 and Thr21.[7]

Positive ion ESMS was performed on a QTOF mass spectrometer (Micromass, Manchester, UK). The ions are produced in an atmospheric pressure ionization (API)/ESI ion source and are transported to the mass spectrometer through a hexapole lens for optimal transmission. The source temperature was 30 °C and a drying gas flow rate of 40 l h^{-1}. A potential of 1–1.4 kV applied to the Nanoflow tip in the ion source combined with a nitrogen back-pressure of 5–6 psi produced a sample flow rate of 10–30 nl min^{-1} into the analyser. The cone voltage was 40 V. A quadrupole analyser was used to select the precursor ion for fragmentation in the hexapole collision cell. The collision gas was Ar at a pressure of $\sim 5 \times 10^{-3}$ mbar and a collision energy of 30 V. Product ions were analysed using an orthogonal time-of-flight analyser, fitted with a reflector, a microchannel plate detector and a time-to-digital converter.

The acquisition and deconvolution of data were performed on a Mass Lynx Windows NT PC data system. The mass accuracy of all measurements was within 0.2 *m/z* unit.

The sequencing of the glycosylated TAP25 was performed on ~2–5 pmol of material, dissolved in 1:1 (v/v) methanol–water containing 0.1% concentrated formic acid. For nanospraying, a borosilicate micropipette of 1.5 µl total volume was used.

RESULTS AND DISCUSSION

In MS1, (+)ESI mass spectra of the native glycopeptide cationized molecules were detected as the singly, doubly and triply charged ion species (spectra not shown), giving deconvoluted values of 2730.3 ($[M + H]^+$), 2752.3 ($[M + Na]^+$), 2768.2 ($[M + K]^+$), 2774.3 ($[M + 2Na - H]^+$) and 2790.2 ($[M + Na + K - H]^+$), which is in accord with the calculated molecular mass of TAP25-2 (2730.8).

The ion at *m/z* 911.1 corresponding to $[M + 3H]^{3+}$ was chosen as a precursor for the MS/MS analysis of peptide fragments. The results of this analysis are presented in Fig. 1 and the fragment ions observed are listed in Table 1. The presence of GalNAc was indicated by the abundant ion at *m/z* 204.1 ($[GalNAc]^+$) and the loss of one GalNAc at *m/z* 1264.6 ($[M - GalNAc + 2H]^{2+}$). These ions indicate a relative easy cleavage of the α-glycosidic bonds to the peptide at the Ser/Thr sites resulting in loss of GalNAc residues from TAP25-2 under MS/MS conditions.

The evidence for the localization of a GalNAc residue on Thr9 was based on the presence of the singly charged, monoglycosylated b_9 ion at *m/z* 1036.9 and supported by a series of singly and doubly charged b-type ions (Table 1). Distinction of the potential glycosylation sites on Thr9 and Ser10 was supported by the y_{16}^+ of very low abundance. The position of the second GalNAc residue on Ser20 or Thr21 could be proposed from the observation of bisglycosylated b_{22}^{2+} (*m/z* 1224.5). Although the cleavage of the peptide bond does take place at Thr21, there is a high probability of a quantitative neutral loss of one GalNAc unit resulting in the *m/z* 2174.7 (b_{21}^+) and 1087.0 (b_{21}^{2+}) ions. Following the observation of the peptide bond cleavage ion at Ala22, indicating the presence of two GalNAcs in this stretch (b_{22}^{2+} at *m/z* 1224.5), the assumption of one of the previous attachment sites must be considered. Additional indirect evidence was provided also by b_{23}^{3+} (*m/z* 848.7). Interestingly, the bisglycosylated y_n and b_m fragment ions appeared primarily as doubly or triply

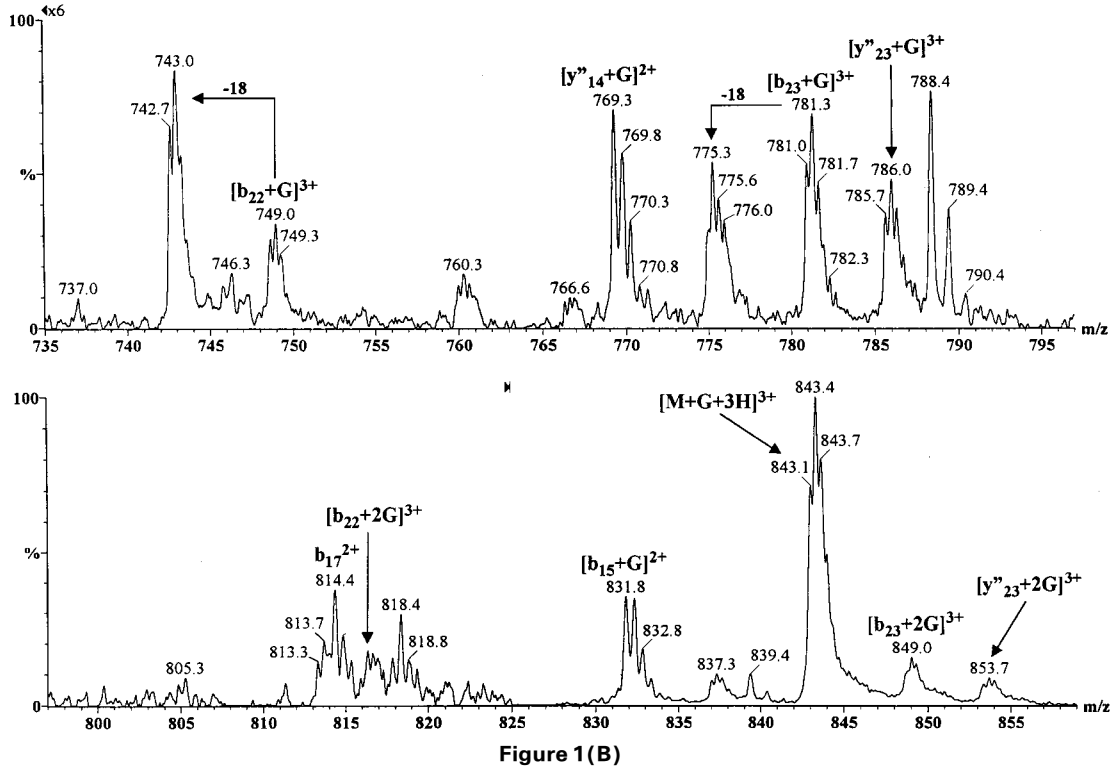

Figure 1(B)

Table 1. Product ion spectrum obtained by CID of *m/z* 911 for GalNAc-disubstituted TAP25-2 peptide[a]

	Amino acid	Masses of *C*-terminal fragments				
y_n	sequence	GalNAc 0		1		2
1	A					
2	P	187.1 $(y)^+$				
3	P	284.1 $(y)^+$				
4	A	355.1 $(y)^+$				
5	T	456.2 $(y)^+$				
6	S	543.2 $(y)^+$				
7	G					
8	P	697.3 $(y)^+$		900.4 $(y)^+$		
9	A			972.4 $(y)^+$		
10	P	865.4 $(y)^+$				
11	R			1224.5 $(y)^+$		
12	T	1122.9 $(y)^+$		1325.6 $(y)^+$		
13	D					
14	P	1334.0 $(y)^+$	667.8 $(y)^{2+}$	1537.6 $(y)^+$	769.3 $(y)^{2+}$	
15	A	1405.6 $(y)^+$				
16	S	1492.4 $(y)^+$				
17	T	1593.7 $(y)^+$		1796.8 $(y)^+$		
18	V			1896.0 $(y)^+$		
19	G	1749.8 $(y)^+$		1952.8 $(y)^+$		1076.5 $(y)^{2+}$
20	H	1886.8 $(y)^+$		2090.6 $(y)^+$		
21	A			2160.8 $(y)^+$		788.4 $(y)^{3+}$
22	P	2055.0 $(y)^+$		2259.2 $(y)^+$		
23	P	1076.5 $(y)^+$			785.7 $(y)^{3+}$	853.4 $(y)^{3+}$
24	A			2428.1 $(y)^+$		
25	T	2323.8 $(y)^+$				

	Amino acid	Masses of *N*-terminal fragments				
b_m	sequence	GalNAc 0		1		2
25						
	A					
24	P					
23	P		1069.5 $(b)^{2+}$		781.0 $(b)^{3+}$	848.7 $(b)^{3+}$
22	A	2040.9 $(b)^+$	1021.0 $(b)^{2+}$	2243.9 $(b)^+$	1122.5 $(b)^{2+}$	1224.5 $(b)^{2+}$
21	T			2174.7 $(b)^+$	1087.0 $(b)^{2+}$	
20	S		934.9 $(b)^{2+}$	1036.5 $(b)^{2+}$		
19	G					
18	P					
17	A	1627.7 $(b)^+$	814.4 $(b)^{2+}$	1830.9 $(b)^+$	916.4 $(b)^{2+}$	
16	P					
15	R	1459.7 $(b)^+$		1662.7 $(b)^+$	831.8 $(b)^{2+}$	
14	T					
13	D	588.2 $(a)^{2+}$		1406.7 $(b)^+$		
12	P	1087.0 $(b)^+$				
11	A	990.4 $(b)^+$		597.2 $(b)^{2+}$		
10	S			1122.9 $(b)^+$		
9	T	831.8 $(b)^+$		1036.9 $(b)^+$		
8	V	731.3 $(b)^+$	703.3 $(a)^+$			
7	G	632.2 $(b)^+$				
6	H	575.2 $(b)^+$	547.2 $(a)^+$			
5	A	438.2 $(b)^+$				
4	P					
3	P					
2	A					
1	T					

[a] The *m/z* values of the fragment ions from the *C*-terminus (x_n, y_n and z_n ions) and the *N*-terminus of TAP 25-2 (a_m, b_m and c_m ions) are listed in accordance with the peptide sequence. Ions of low abundance are not assigned. The masses are accompanied by assignments of the fragment ions given in parentheses and their corresponding charge states.

charged species. The assignment of charges in multiply charged ions was possible owing to the high resolution in MS2.

The y_n and b_m fragment ions were in general the most abundant in the spectrum, whereas z_n, x_n and also a_m and c_m ions were only rarely detected. Similarly, the companion ions of singly charged y_n-fragments ($y_n - 17$, $y_n - 18$) or b_m-fragments ($b_m - 17$, $b_m - 18$) were rare and of low abundance. Facile loss of GalNAc moieties gave rise to the nearly complete series of non-glycosylated y_n and b_m fragment ion species, which are also listed in Table 1. Internal fragmentation of y_n ions at X-Pro, as observed in PSD-MALDI/MS, was insignificant. In general, the (+)ESI mass spectrum of TAP25-2 seems to be much less complex in its appearance than the corresponding PSD-MALDI mass spectrum.[7] In comparison with MS sequencing methods for localization of *O*-glycosylation sites, the bisglycosylated MUC1 peptide 25-mer can be used as an excellent model compound owing to its structural features, namely the substitution of only two out of six possible glycosylation sites, and to the high content of proline in the peptide chain. The present nanospray QTOF-ESMS compared with PSD-MALDI/TOF-MS[7] features less structure-irrelevant companion ions and a clear distinction of higher and lower glycosylated fragment ions based on the charge number observed. Owing to the high mass accuracy and resolution, the assignment of ions is straightforward. The considerable cleavage of α-glycosidic bonds of GalNAc moieties to the Ser/Thr linkage sites should be the object of further investigations.

Acknowledgements

This study was supported by the Deutsche Forschungsgemeinschaft, grant Ha2092/4-7.

REFERENCES

1. D. M. Carlson, *J. Biol. Chem.* **243**, 616 (1968).
2. G. W. Hart, R. S. Haltiwanger, G. D. Holt and W. G. Kelly, *Annu. Rev. Biochem.* **58**, 841 (1989).
3. S. Goletz, F.-G. Hanisch and U. Karsten, *J. Cell Sci.* **110**, 1585 (1997).
4. F.-G. Hanisch, G. Uhlenbruck, J. Peter-Katalinic, H. Egge, J. Dabrowski and U. Dabrowski, *J. Biol. Chem.* **264**, 872 (1989).
5. A. Pisano, D. R. Jadine, N. H. Packer, J. W. Redmond, K. L. Williams, A. A. Gooley, V. Farnsworth, W. Carson and P. K. Cartier, in *Techniques in Glycobiology*, edited by R. Townsend and A. Hotchkiss, p. 299. Marcel Dekker, New York (1997).
6. R. Kaufmann, D. Kirsch and B. Spengler, *Int. J. Mass Spectrom. Ion Processes* **131**, 355 (1994).
7. S. Goletz, B. Thiede, F.-G. Hanisch, M. Schultz, J. Peter-Katalinic, S. Müller, O. Seitz and U. Karsten, *Glycobiology* **7**, 881 (1997).
8. S. Müller, S. Goletz, N. Packer, A. A. Gooley, A. M. Lawson and F.-G. Hanisch, *J. Biol. Chem.* **272**, 24780 (1997).
9. J. Peter-Katalinic, K. Williger, H. Egge, B. Green, F.-G. Hanisch and D. Schindler, *J. Carbohydr. Chem.* **13**, 447 (1994).
10. J. Peter-Katalinic, A. Ashcroft, B. Green, F.-G. Hanisch, Y. Nakahara, H. Ilijima and T. Osawa, *Org. Mass Spectrom.* **29**, 747 (1994).
11. K. Alving, R. Körner, H. Paulsen and J. Peter-Katalinic, *J. Mass Spectrom.*, submitted.
12. T. R. E. Stadie, W. Chai, A. M. Lawson, P. G. H. Byfield and F.-G. Hanisch, *Eur. J. Biochem.* **229**, 140 (1995).