Pergamon

Tetrahedron: *Asymmetry* 11 (2000) 75–93

TETRAHEDRON:
ASYMMETRY

A strategy for the identification of site-specific glycosylation in glycoproteins using MALDI TOF MS

Kevin Mills, Andrew W. Johnson, Ole Diettrich, Peter T. Clayton and Bryan G. Winchester *

Biochemistry, Endocrinology and Metabolism Unit, Institute of Child Health (University College London) and Great Ormond Street Hospital NHS Trust, 30 Guilford Street, London, WCIN 1EH, UK

Received 2 November 1999; accepted 29 November 1999

Abstract

A strategy for investigation of site-specific glycosylation of glycoproteins has been developed, based on peptide mass fingerprinting using matrix assisted laser desorption ionisation time of flight mass spectrometry (MALDI TOF MS). The glycoprotein is subjected to sequential digestion with a protease and glycan-specific endoglycosidases or with the glycan-specific endoglycosidases followed by the protease. Peptides with characteristic masses are detected for sequences containing glycosylated asparagine residues. By using a panel of three proteases, chymotrypsin, protease V8 and trypsin, and endoglycosidases F3 and H and peptide N-glycanase F, it was possible to monitor the state of glycosylation of all putative *N*-glycosylation sites on three glycoproteins. It was deduced that all potential *N*-glycosylation sites in human serum transferrin (two) and α1-antitrypsin (three) were occupied by non-fucosylated, biantennary, disialylated, complex glycans. In contrast, only four (asparagines 19, 59, 146 and 270) out of the five potential sites were glycosylated in recombinant human β-glucosylceramidase, with the site nearest the C-terminal (asparagine 462) being unoccupied. The glycans at each site consisted of a mixture of non-fucosylated and core α1–6 fucosylated oligomannose glycans (Man_3 $GlcNAc_2$), derived from the enzymic truncation of complex glycans.

1. Introduction

Carbohydrate chains or glycans are found covalently linked to glycoproteins by N-, O- or C-glycosidic linkages. The carbohydrate moiety of a glycoprotein can have many different functions including solubility, stability, resistance to proteolysis, folding, structure, aggregation, transport, recognition and biological activity.[1] The importance of appropriate glycosylation is demonstrated by the severe diseases that result from aberrations in glycosylation.[2] The most common form of glycosylation of mammalian proteins is N-linked glycosylation, in which glycans are attached to the amide N of a side chain of an

* Corresponding author. Tel: 0044-171-242-9789; fax: 0044-171-404-6191; e-mail: b.winchester@ich.ucl.ac.uk

0957-4166/00/$ - see front matter
PII: S0957-4166(99)00545-5

asparagine, which always occurs in a sequence (sequon), AsnXThr/Ser, where X can be any amino acid except proline. Although well over 100 different N-linked glycans have been identified, they fall into three main classes, high mannose, hybrid and complex. An individual glycoprotein may contain different types of glycans at different sequons, or even different glycans at the same site. Furthermore, not all the sequons may be glycosylated.

It is important to know the state of glycosylation of a protein in terms of occupation of potential *N*-glycosylation sites (macroheterogeneity) and type of glycosylation at that site (microheterogeneity) to investigate the function of normal glycosylation or to understand the consequences of aberrant glycosylation. It will also be necessary to define the glycosylation of recombinant glycoproteins for therapeutic purposes.

N-linked glycans can be released from glycoproteins by the action of peptide N-glycanases or by endoglycosidases and analysed by a variety of techniques.[3–7] Although this approach will identify the types and relative proportions of glycans, it will not reveal the distribution of the glycans over the possible glycosylation sites. This can be achieved, though, by proteolytic digestion of the glycoprotein, separation of the resultant mixture of peptides and glycopeptides and release and analysis of the glycans from each glycopeptide.[8–11] Apart from the cost and time of the separation of the glycopeptides, the main problems with this strategy are ensuring that a glycopeptide is resolved from peptides or other glycopetides and the influence of the glycosylation itself on the proteolysis of the glycoprotein. We have developed a strategy for identifying the occupancy and type of N-linked glycan at each potential glycosylation site that does not necessitate separation of glycopeptides after proteolysis. The peptide masses obtained by the theoretical proteolytic digestion of a glycoprotein (MS DIGEST-Protein Prospector, SWISS-PROT) were compared with the experimental sequential digestion of the protein with a protease followed by an endoglycosidase/peptide N-glycanase protocol or vice versa. The mixtures of peptides and glycopeptides can be analysed directly by MALDI TOF MS because the N-linked glycans have either been removed or are small and uncharged, obviating the need for pre-analysis separation. Carbohydrate-containing peptide masses were calculated using 'PAWS' proteomics software (http://proteomics.com/software/paws.htm). It was found that by using an array of digestion protocols using three proteases and specific endoglycosidase/peptide N-glycanases all glycosylation sites could be probed. For example, if the asparagine residue of a glycosylation sequon is glycosylated, cleavage of the glycosylamino linkage by PNGase F will result in conversion of the asparagine to an aspartic acid. Consequently, the Da/e of an ion corresponding to a deglycosylated peptide will be 1 Da greater than that derived from the same unglycosylated peptide, thereby indicating whether the site is glycosylated or not. The type of glycan attached to the asparagine can be deduced from the masses of the peptides obtained with different endoglycosidases because of their precise specificities.[12] Cleavage of high-mannose N-linked glycans by endoglycosidase H will result in peptides with a single *N*-acetylglucosamine attached to the asparagine. The masses of deglycosylated peptide will be 203.1 Da/e greater than that observed from the same unglycosylated peptide. Cleavage of a complex glycan with endoglycosidase F3 will leave a fucosyl *N*-acetylglucosamine or a single *N*-acetylglucosamine attached to the asparagine, depending on whether the glycan is core α-6-fucosylated or not. The mass of such a peptide will be 203.1 (or 349 with fucose) Da greater than that derived from the same unglycosylated peptide.

This strategy has been applied to two serum glycoproteins of considerable diagnostic importance for glycosylation defects, α1-antitrypsin and transferrin and a recombinant therapeutic glycoprotein, β-glucosylceramidase (Cerezyme).

2. Transferrin (Table 1a and Table 2a)

Human serum transferrin consists predominantly of a glycoform in which the two glycosylation sites at asparagine residues, 413 and 611, are occupied by complex, non-core fucosylated, biantennary glycans,[13] which are susceptible to cleavage by both endoglycosidase F3 and PNGase F.

2.1. Proteolytic digestion alone

Ions corresponding to the mass of peptides expected to be obtained on digestion of transferrin were detected when the digestion mixtures for all three proteases were analysed by MALDI TOF MS. The % of the amino acid sequence covered by the peptides detected was 48.7%, 48.3% and 42.4% for trypsin, chymotrypsin and protease V8, respectively. Altogether 469 out of the 679 (69.1%) amino acids in the sequence were detected using the three proteases. Different regions of the sequence were inaccessible to the different proteases, illustrating the importance of using more than one protease. However, no glycopeptides were detected with any of the proteases, indicating that sialylated glycopeptides cannot be analysed directly using either the DHB or the αC4HA matrix at the concentrations used in this study.

2.2. Proteolysis after predigestion with endoglycosidase F3

Transferrin was predigested with endoglycosidase F3 before digestion with the three proteases separately to see if any glycopeptides with truncated glycans could be detected. However, no masses corresponding to peptides with an additional mass of 203.1 were detected and the overall coverage of the amino acid sequence for the three proteases decreased from 69.1% to 57.0% (387 out of the 679 amino acids).

2.3. Proteolysis followed by digestion with endoglycosidase F3

Similarly, the digestion of transferrin with endoglycosidase F3 after proteolysis did not produce any masses that could be designated as glycopeptides. The combined coverage of the amino acid sequence remained at 69.1%.

2.4. Proteolysis after predigestion with PNGase F

The glycosylation site, asparagine 611, occurs in the tryptic peptide containing amino acids 603–623. The theoretical Da/e of this peptide is 4605.9 when a complex biantennary glycan is attached and 2498.1 Da/e if this sequon is not glycosylated. No peptide of 2498.1 Da/e was detected by MALDI TOF analysis after tryptic digestion of transferrin (Fig. 1a), but when the transferrin was predigested with PNGase F and then digested with trypsin a peptide of 2499.1 Da/e was observed (Fig. 1b). This corresponds to the deglycosylated peptide in which asparagine 611 has been converted into aspartic acid by PNGase F, indicating that this site is occupied by a complex glycan.

No peptides corresponding to the glycosylated or deglycosylated tryptic peptide containing asparagine 413 were detected with or without predigestion with PNGase F. However, digestion of deglycosylated transferrin with protease V8 produced two peptides that covered the first glycosylation asparagine, 413, with masses of 1259.5 and 2572.3 Da/e corresponding to amino acids 411–420 and 393–416+1 Da, respectively. Chymotryptic digestion of transferrin, that had been predigested with PNGase F,

Table 1

(a) Summary of results for transferrin

Digestion protocol	Coverage of the sequence (%)			Total coverage of the protein (%)
	Trypsin	chymotrypsin	V8 (DE)	
(1) Protease	331 / 679 48.7%	328 / 679 48.3%	288 / 679 42.4%	469 out of 679 amino acids 69.1%
(1) Endo F_3 (2) Protease	145/ 679 21.4%	148 / 679 21.8%	213 / 679 31.4%	387 out of 679 amino acids 57.0%
(1) Protease (2) Endo F_3	331 / 679 48.7%	328 / 679 48.3%	288 / 679 42.4%	469 out of 679 amino acids 69.1%
(1) PNGase F (2) Protease	407 / 679 59.9%	384 / 679 56.6%	283 / 679 41.7%	604 out of 679 amino acids 88.9%
(1) Protease (2) PNGase F	360 / 679 53.0%	375 / 679 55.2%	301 / 679 44.3%	519 out of 679 amino acids 76.4%

(b) Summary of results for α1-antitrypsin

Digestion protocol	Coverage of the sequence (%)			Total coverage of the protein (%)
	Trypsin	Chymotrypsin	V8 (DE)	
(1) Protease	146 / 394 37.1%	139 / 394 35.3%	186 / 394 47.2%	284 out of 394 amino acids 71.8%
(1) Endo F_3 (2) Protease	236 / 394 59.9%	228 / 394 57.9%	235 / 394 59.6%	353 out of 394 amino acids 89.6%
(1) Protease (2) Endo F_3	161 / 394 40.9%	164 / 394 41.6%	198 / 394 50.3%	330 out of 394 amino acids 83.8%
(1) PNGase F (2) Protease	173 / 394 43.9%	248 / 394 62.9%	125 / 394 31.7%	341 out of 394 amino acids 86.5%
(1) Protease (2) PNGase F	146 / 394 37.1%	139/ 394 35.3%	186 / 394 47.2%	284 out of 394 amino acids 71.8%

(c) Summary of results for Cerezyme®

Digestion protocol	Coverage of the sequence (%)			Total coverage of the protein (%)
	Trypsin	Chymotrypsin	V8 (DE)	
(1) Protease	324 / 497 65.2%	306 / 497 61.6%	259 / 497 52.1%	418 out of 497 amino acids 84.1%
(1) Endo H (2) Protease	157 / 497 31.6%	350 / 497 70.4%	230 / 497 46.3%	444 out of 497 amino acids 89.3%
(1) Protease (2) Endo H	392 / 497 78.9%	333 / 497 67.0%	294 / 497 59.2%	479 out of 497 amino acids 96.4%
(1) PNGase F (2) Protease	277 / 497 55.7%	335 / 497 67.4%	285 / 497 57.3%	440 out of 497 amino acids 88.5%
(1) Protease (2) PNGase F	338 / 497 68.0%	336 / 497 67.6%	258 / 497 51.9%	435 out of 497 amino acids 87.5%

Table 2

(a) Summary for transferrin analysis

	Digestion Sequence			
Protease	(1) PNGase F (2) Protease	(1) Protease (2) PNGase F	(1) Endo F3 (2) Protease	(1) Protease (2) Endo F3
Trypsin	$X_{413}\checkmark_{611}$	$X_{413}\checkmark_{611}$	$X_{413}X_{611}$	$X_{413}X_{611}$
Chymotrypsin	$\checkmark_{413}\checkmark_{611}$	$X_{413}\checkmark_{611}$	$X_{413}X_{611}$	$X_{413}X_{611}$
V8(DE)	$\checkmark_{413}X_{611}$	$\checkmark_{413}X_{611}$	$X_{413}X_{611}$	$X_{413}X_{611}$
Summary	All sites monitored.			

Transferrin glycosylation sites are 413 and 611.

(b) Summary for α1antitrypsin analysis

	Digestion Sequence			
Protease	(1) PNGase F (2) Protease	(1) Protease (2) PNGase F	(1) Endo F3 (2) Protease	(1) Protease (2) Endo F3
Trypsin	$X_{46}\checkmark_{83}X_{247}$	$X_{46}X_{83}X_{247}$	$X_{46}X_{83}\checkmark_{247}$	$X_{46}X_{83}\checkmark_{247}$
Chymotrypsin	$\checkmark_{46}\checkmark_{83}\checkmark_{247}$	$X_{46}X_{83}X_{247}$	$\checkmark_{46}\checkmark_{83}\checkmark_{247}$	$\checkmark_{46}X_{83}\checkmark_{247}$
V8(DE)	$X_{46}\checkmark_{83}X_{247}$	$X_{46}X_{83}X_{247}$	$X_{46}X_{83}X_{247}$	$X_{46}\checkmark_{83}X_{247}$
Summary	All sites monitored.			

α1-antitrypsin glycosylation sites are 46, 83 and 247.

(c) Summary for Cerezyme analysis

	Digestion Sequence					
Protease	(1) PNGase F (2) Protease	(1) Protease (2) PNGase F	(1) Endo H (2) Protease [peptide+GlcNac]	(1) Endo H (2) Protease [peptide+GlcNac + fucose]	(1) Protease (2) Endo H [peptide+GlcNac]	(1) Protease (2) Endo H [peptide+GlcNac + fucose]
Trypsin	$X_{19}\checkmark_{59}\checkmark_{146}\checkmark_{270}$	$X_{19}X_{59}X_{146}\checkmark_{270}$	$X_{19}X_{59}\checkmark_{146}\checkmark_{270}$	$X_{19}X_{59}X_{146}\checkmark_{270}$	$X_{19}\checkmark_{59}\checkmark_{146}\checkmark_{270}$	$X_{19}\checkmark_{59}X_{146}X_{270}$
Chymotrypsin	$X_{19}\checkmark_{59}\checkmark_{146}\checkmark_{270}$	$X_{19}X_{59}\checkmark_{146}X_{270}$	$X_{19}X_{59}\checkmark_{146}\checkmark_{270}$	$X_{19}X_{59}X_{146}X_{270}$	$X_{19}X_{59}\checkmark_{146}X_{270}$	$\checkmark_{19}X_{59}\checkmark_{146}X_{270}$
V8(DE)	$X_{19}\checkmark_{59}\checkmark_{146}\checkmark_{270}$	$X_{19}X_{59}X_{146}X_{270}$	$\checkmark_{19}X_{59}X_{146}X_{270}$	$X_{19}X_{59}X_{146}X_{270}$	$\checkmark_{19}X_{59}\checkmark_{146}X_{270}$	$X_{19}\checkmark_{59}X_{146}\checkmark_{270}$
Summary	All sites monitored.					

Protease	Glycopeptide Detected (-fucose)	Glycopeptide Detected (+fucose)
Trypsin	$\checkmark_{19}X_{59}X_{146}X_{270}$	$X_{19}X_{59}X_{146}\checkmark_{270}$
Chymotrypsin	$X_{19}X_{59}X_{146}X_{270}$	$X_{19}X_{59}X_{146}X_{270}$
V8(DE)	$X_{19}X_{59}X_{146}X_{270}$	$X_{19}X_{59}X_{146}X_{270}$

Cerezyme glycosylation sites are 19, 59, 146 and 270.

Key:

$\checkmark_n$	-indicates the detection of a peptide which covered a sequon
X_n	- indicates the detection of a peptide which covered a sequon was not observed
Where **n** indicates the position of the asparagine in the glycosylation sequon in the amino acid sequence	

produced masses of 1747.7 and 1725.7 Da/e, which correspond to the amino acid sequences 413–427 and 608–622+1 Da, containing the two glycosylation sites 413 and 611.

Thus, it was possible to show by deglycosylation with PNGase F prior to proteolysis that both glycosylation sites of transferrin are occupied by complex glycans. Further, the overall coverage of the amino acid sequence was increased from 69.1% to 88.9% (604/679 amino acids).

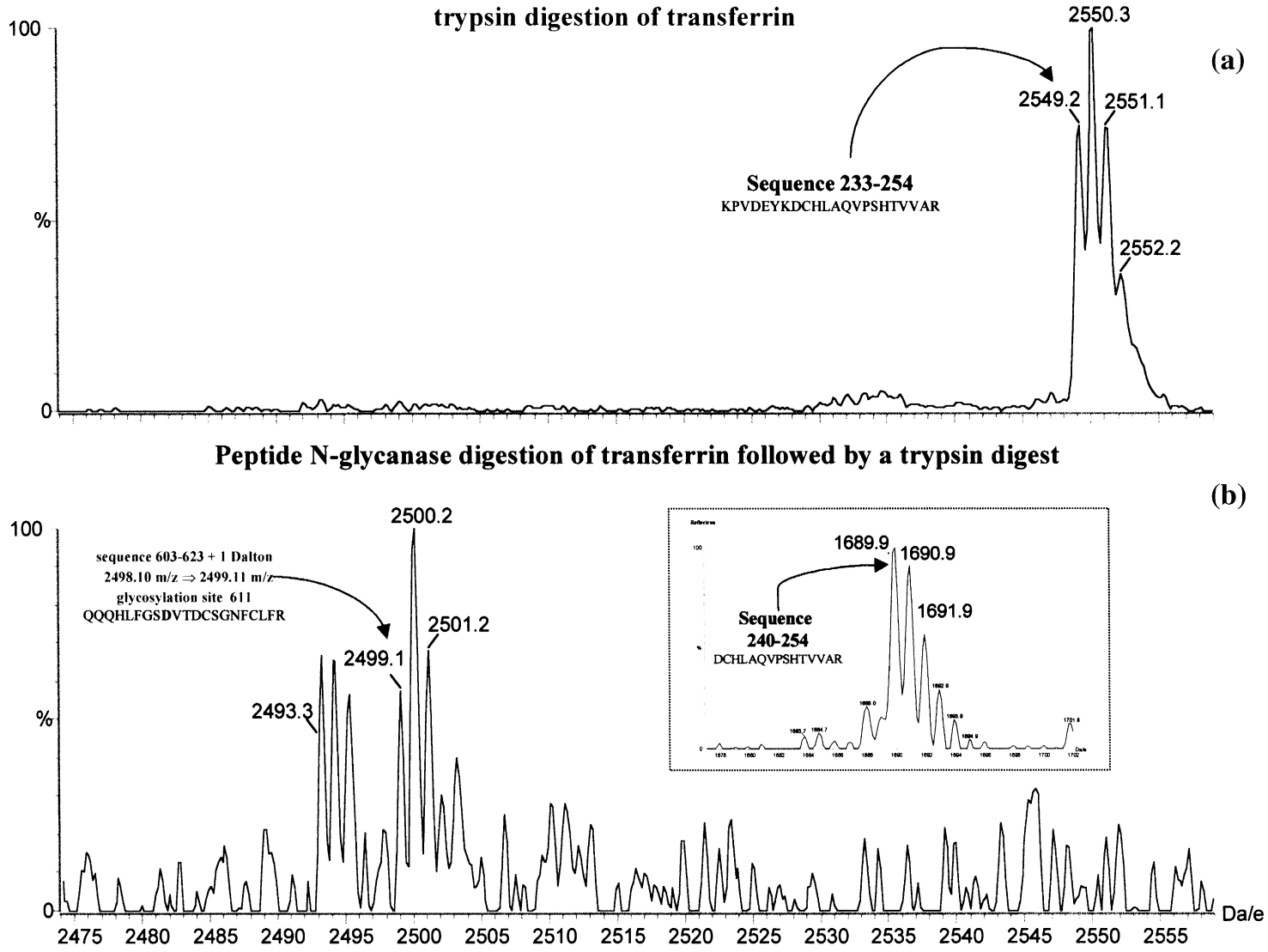

Fig. 1. Mass spectra of a tryptic peptide containing glycosylation site asparagine 611 of human transferrin before (a) and after (b) enzymatic removal of the glycans by PNGase F

2.5. Proteolysis followed by PNGase F digestion

The use of PNGase F after proteolysis of the protein also resulted in peptides being detected that covered both sequons of the molecule. Both tryptic and chymotryptic digestion of transferrin, prior to treatment with PNGase F, produced peptides in which asparagine 611 had been converted to aspartic acid. Conversely, the use of protease V8 produced peptides covering the first sequon only. Digestion with the three proteases followed by PNGase F treatment, increased the coverage of the amino acid sequence to 76.4% (519/679 amino acids).

3. α1-Antitrypsin (Table 1b and Table 2b)

Human α1-antitrypsin has three glycosylation sites at asparagine residues 46, 83 and 247, which are occupied by biantennary disialylated complex glycans,[14] which are not core α1–6 fucosylated.

3.1. Proteolytic digestion alone

No glycopeptides were detected in the mass spectrum after α1-antitrypsin was digested with any of the three proteases using DHB or αC4HA as the matrix. The peptides detected in the mass spectrum after digestion with trypsin, chymotrypsin and protease V8 covered 37.1%, 35.3% and 47.2% of the amino acid sequence, respectively.

3.2. Proteolysis after predigestion with endoglycosidase F3

This digestion protocol resulted in the detection of a glycopeptide with a mass of 1958.9 Da/e corresponding to the amino acid sequence 244–259 + one *N*-acetylglucosaminyl unit (Fig. 2b). It was not detected after digestion with trypsin alone (Fig. 2a). Such truncated glycopeptides were not detected for the other two glycosylation sites. Glycopeptides with an *N*-acetylglucosaminyl residue attached to asparagine corresponding to all three glycosylation sites were detected when α1-antitrypsin was predigested with endoglycosidase F3 prior to chymotryptic digestion. No truncated glycopeptides were detected after successive digestion with endoglycosidase F3 and protease V8. The truncation of the glycans, prior to proteolysis with the three proteases, resulted in an increase in coverage of the amino acid sequence to 89.6% (353/394 amino acids) (Table 1b and Fig. 3b).

3.3. Digestion with endoglycosidase F3 after proteolysis

This digestion also resulted in the detection of truncated glycopeptides corresponding to all three sequons with one or the other of the three proteases. Tryptic digestion of the α1-antitrypsin resulted in truncated glycopeptides being detected that covered asparagine 247 only. Chymotryptic digestion produced truncated glycopeptides that covered the asparagines 46 and 247 but not 83, whilst digestion with protease V8 produced truncated glycopeptides only covering asparagine 83. The combined coverage of the amino acid sequence after successive digestion with a protease and endoglycosidase F3 was 83.8% (330/394 amino acids) (Table 1b).

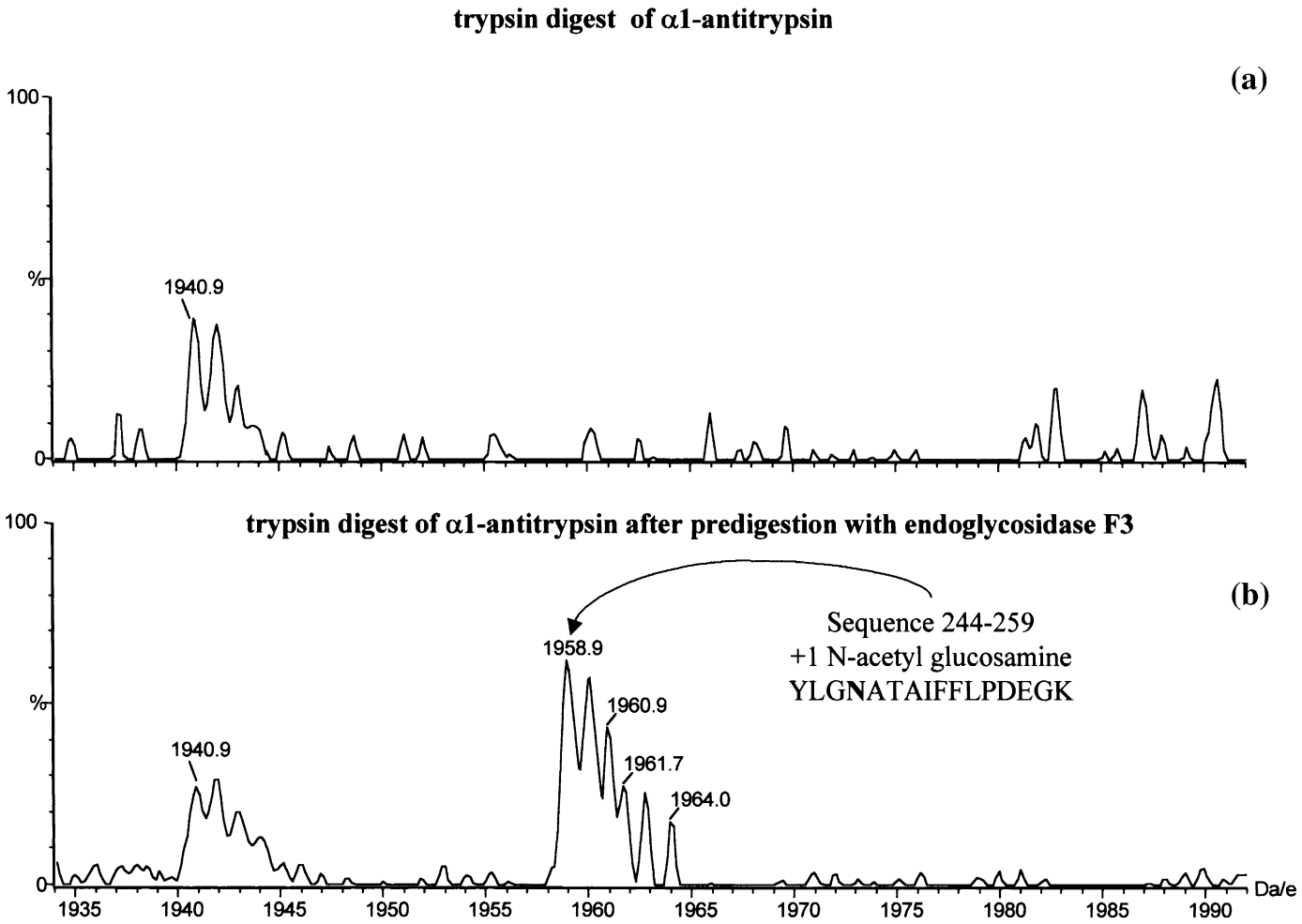

Fig. 2. Mass spectra of a tryptic digestion of α1-antitrypsin before (a) and after (b) enzymatic truncation of the glycans by *endo* F3. The peak at 1958.9 Da/e corresponds to the peptide sequence 244–259 + one *N*-acetylglucosamine residue (+203.1 Da)

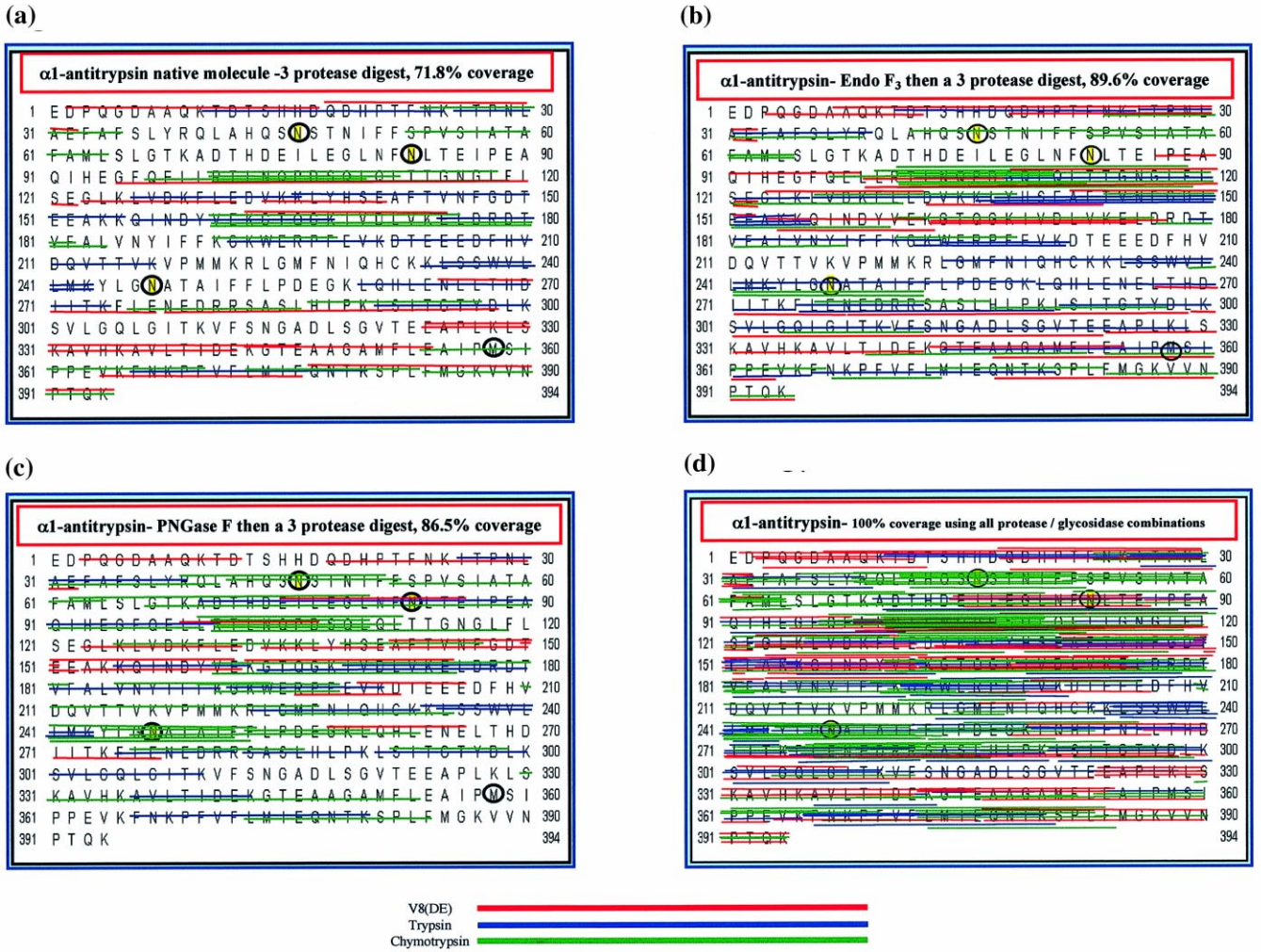

Fig. 3. Distribution of peptides over the amino acid sequence of α1-antitrypsin after digestion by different protocols

3.4. Predigestion with PNGase F followed by proteolytic digestion

The enzymatic release of the intact glycans using PNGase F, prior to proteolysis with the three proteases, allowed the detection of peptides covering all three sequons of the protein. Tryptic digestion of the deglycosylated protein allowed the detection of the peptide that included asparagine 83 but peptides containing asparagines 46 and 247 were not observed. Again, digestion with chymotrypsin resulted in multiple peptides covering all three sequons in the molecule. The use of protease V8 produced a peptide that covered asparagine 83. By deglycosylating the protein with PNGase F prior to proteolysis, the combined coverage of the amino acid sequence with the three proteases increased to 86.5% (341/394) (Table 1b and Fig. 3c).

3.5. Proteolysis followed by PNGase F digestion

No additional peptides were detected by deglycosylation with PNGase F after proteolysis with any of the three proteases. Consequently, no information about the glycosylation sites was obtained nor any gain in the coverage of the amino acid sequence (Table 1b).

4. Human β-glucosylceramidase (Table 1c and Table 2c)

Human recombinant β-glucosylceramidase (Cerezyme, EC 3.2.1.45) has five potential glycosylation sites at asparagine residues 19, 59, 146, 270 and 462. N-linked complex glycans have been truncated enzymically to terminate in the core structure Man_3 $GlcNAc_2$ + or −Fuc, making them susceptible to cleavage by endoglycosidase H, as well as PNGase F. The recombinant enzyme differs from the modified placental enzyme (Ceredase) in that the arginine residue at position 495 has been replaced by histidine and the high mannose glycan attached to asparagine 19 has been replaced by the truncated glycan, Man_3 $GlcNAc_2$-Fuc, which is also reported to be attached to asparagine 146.[15]

4.1. Proteolytic digestion alone

Intact glycopeptides containing Man_3 $GlcNAc_2$- and Man_3 $GlcNAc_2$-Fuc attached to asparagines 19 and 270, respectively, were observed after digestion of β-glucosylceramidase with trypsin. The glycopeptides were observed using both DHB and αC4HA as the matrix but the signals were much weaker (decreased signal to noise ratio) with αC4HA. Glycopeptides were not detected with the other two proteases. Several peptides, covering the potential glycosylation site, 462, were detected and all contained an unmodified asparagine, indicating that it was not glycosylated. Combining the results for the three proteases it was possible to monitor 418 out of the 497 amino acids in the recombinant enzyme by peptide mass analysis, significantly this excluded two of the reported glycosylation sites 59 and 146.

4.2. Detection of the substitution of arginine 495 to histidine in Cerezyme

Comparison of the tryptic digestions of the placental (Ceredase) and recombinant (Cerezyme) enzymes confirmed that arginine 495 in the placental enzyme is replaced by a histidine in the recombinant form. It was not possible to analyse Ceredase directly because a large concentration of albumin is added during the manufacturing process. However, Ceredase and Cerezyme were purified by SDS-PAGE (to separate the Ceredase from the albumin) and subjected to in-gel tryptic digestion. Fig. 4 shows the mass spectra

obtained after the in-gel digestion of both proteins and analysis of the peptides by MALDI TOF MS. A peptide of 2522.3 Da/e covering amino acids 474–495 is observed with Ceredase (Fig. 4a) but is absent in Cerezyme (Fig. 4b). Conversely, a peptide of 2659.3 Da/e is present in Cerezyme (Fig. 4b) but is absent in Ceredase (Fig. 4a). This peptide corresponds to amino acids 474–496 in the sequence of Cerezyme, in which arginine 495 has been replaced by histidine, abolishing the tryptic cleavage site, but with the addition of arginine 496, which is the new cleavage point. The net gain in mass of the peptide observed in the Cerezyme is 137.0 Da, which agrees with the theoretical value of 137.1 Da.

4.3. Proteolysis after predigestion with endoglycosidase H

This digestion protocol revealed glycopeptides with *N*-acetylglucosamine attached to asparagine residues 59, 146 and 270 (Table 2c). Peptides, corresponding to amino acids 263–277 with *N*-acetylglucosamine (1834.9 Da/e) or fucosyl *N*-acetylglucosamine (1981.0 Da/e) attached to asparagine 270, were detected in the tryptic digestion (Fig. 5). The truncation of the glycans of the protein prior to proteolysis increased the coverage of the amino acid sequence from 84.1% to 89.3% (444/497 amino acids). Interestingly, the amino acid substitution of arginine 495 to histidine was not detected by this protocol.

4.4. Digestion with endoglycosidase H after proteolysis

Information about all four glycosylation sites was obtained with this digestion protocol (Table 2c). Peptides containing asparagines with *N*-acetylglucosamine and peptides containing fucosyl *N*-acetylglucosamine attached were observed for all four glycosylation sites, showing that the microheterogeneity observed for asparagine 270 occurred at all four glycosylation sites. The combined coverage of the amino acid sequence increased to 96.4% (479/497 amino acids).

4.5. Predigestion with PNGase F followed by proteolytic digestion

The use of PNGase F to remove the glycans prior to proteolysis confirmed that asparagines 59, 146 and 270 are glycosylated. The peptides covering these sites all contained aspartic acid residues and were obtained with all three proteases (Table 2c). No peptides including asparagine 19 were detected with any of the proteases. This digestion protocol increased coverage of the amino acid sequence to 88.5% (440/497 amino acids).

4.6. Proteolysis followed by PNGase F digestion

The use of PNGase F after proteolysis with the three proteases only allowed detection of peptides covering two of the occupied asparagines 146 and 270 (Table 2c). The coverage of the amino acid sequence was 87.5% (435/497 amino acids), which is very similar to that obtained when the protein was deglycosylated before proteolysis. However, the peptide containg glycosylation site 59 was not detected using this protocol.

In summary, the array of digestion protocols used showed that four (16, 59, 146 and 270) out of five of the potential glycosylation sites in Cerezyme are occupied and that the fifth site, 462, is not glycosylated. The glycans at all four sites occur both as core α1–6 fucosylated and non-fucosylated forms. Truncated oligomannose glycans, Man_3 $GlcNAc_2$- and Man_3 $GlcNAc_2$-Fuc, were

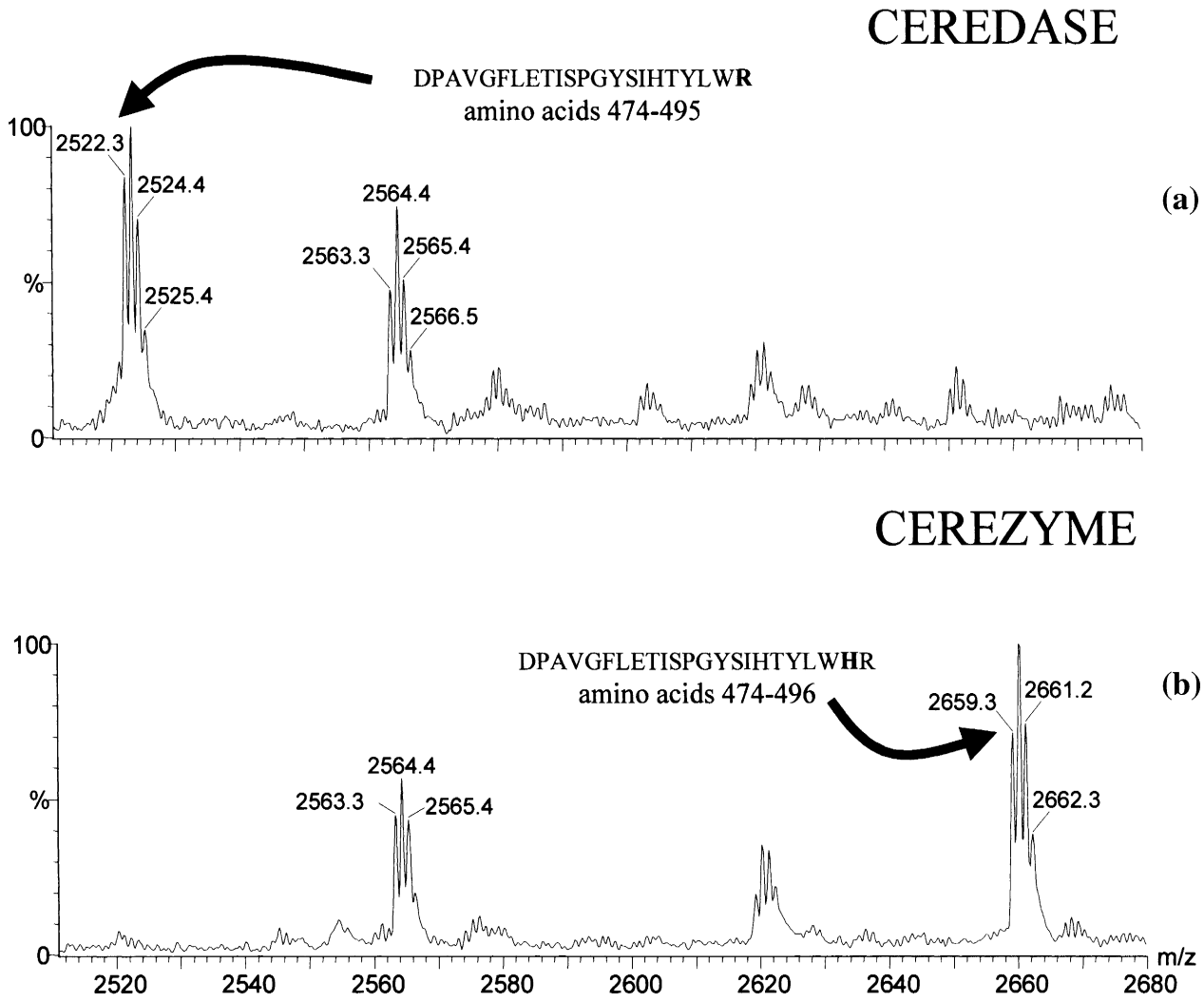

Fig. 4. Mass spectra of the in-gel tryptic digestion of purified (a) Ceredase and (b) Cerezyme

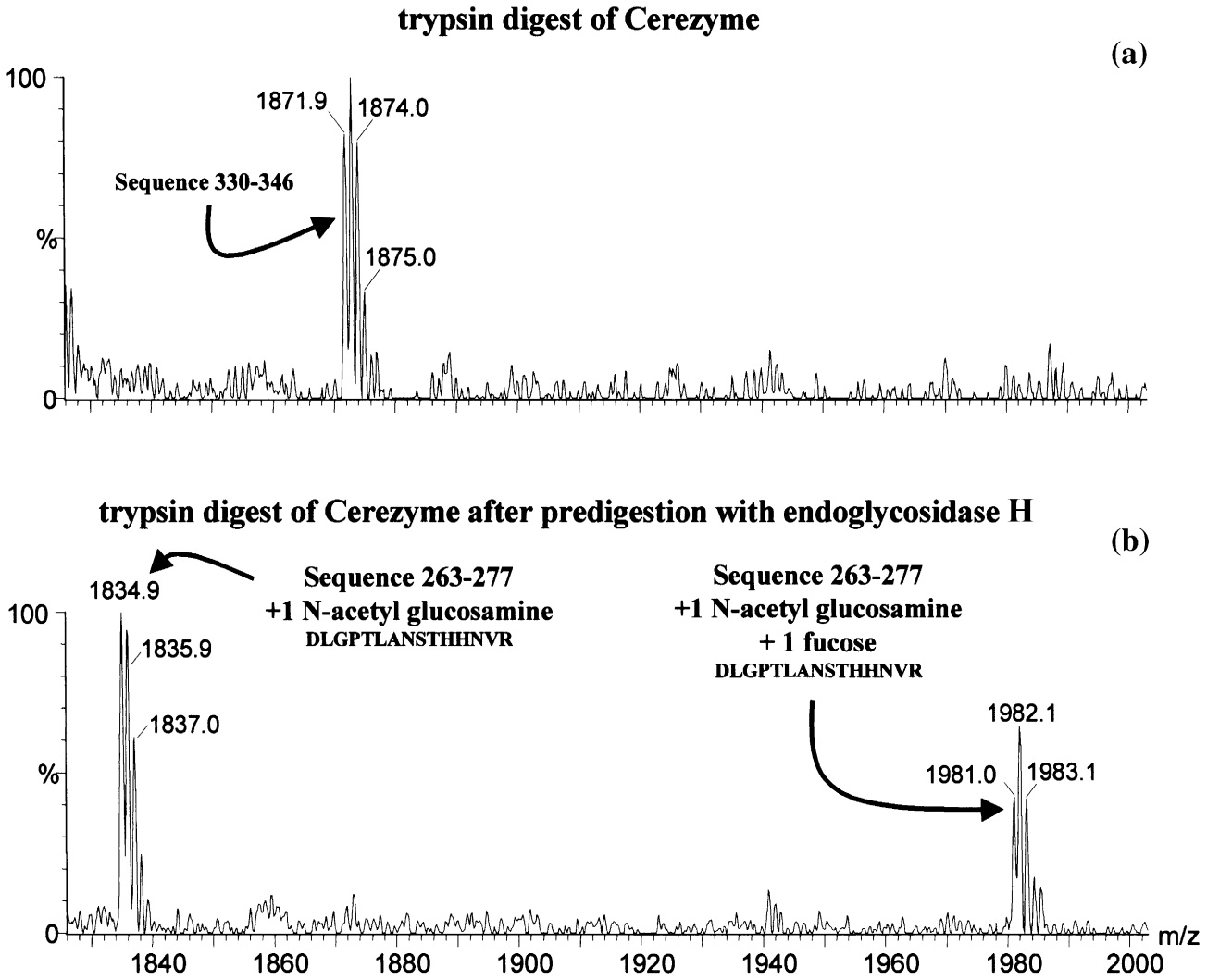

Fig. 5. Tryptic digestion of Cerezyme (a) without and (b) with prior digestion with endoglycosidase H

detected directly at asparagines 19 and 270, respectively, and it is probable that a mixture of both structures occurs at all four glycosylation sites.

5. Discussion

The aim of this work was to develop a rapid and sensitive technique for identifying site-specific glycosylation and amino acid substitutions in glycoproteins. Our strategy of using several digestion protocols has allowed us to formulate guidelines for increasing the probability of observing specific peptides and glycopeptides. However, it was not always possible to predict which peptides would be observed on analysis of proteolytic digests by MALDI TOF MS. The use of MALDI TOF MS for the analysis of proteolytic digests is appropriate because the optimal mass range for these instruments, in terms of sensitivity, mass accuracy and mass resolution, is in the 750–2500 mass range. The use of a panel of proteases and deglycosylating enzymes produces peptides that fall into this optimal range, and it was possible to monitor all putative glycosylation sites and amino acid substitutions in the three proteins studied. Over 99% of the amino acid sequence of human β-glucosylceramidase and α1-antitrypsin could be observed using different permutations of the proteases and deglycosylating enzymes (Fig. 3). The coverage for transferrin, using all combinations, is only 95.1% because there are two sequences, amino acids 123–138 and 155–173, that are either resistant to digestion by any of the combinations of enzymes or unsuited to analysis by MALDI TOF MS. It is not clear why some peptides are not susceptible to MALDI TOF MS analysis but it is likely to be due to either poor ionisation or signal suppression by co-analytes.

Although DHB was the better matrix for analysis of intact neutral glycopeptides as seen in the analysis of Cerezyme, the glycopeptides tended to be too large (>3000 Da/e) for optimum analysis using MALDI TOF MS operating in reflectron mode. No intact sialic acid-containing glycopeptides from transferrin and α1-antitrypsin were detected using either matrix. Removal of the glycans decreased the mass of the glycopeptides to within the optimal range of MALDI TOF MS and allowed the use of the more sensitive matrix αC4HA. The use of fucose as a co-matrix with αC4HA not only increased the sensitivity of the analysis of peptides and truncated glycopeptides but it also concentrated the sample into the centre of the target, eliminating the need to 'search' the target for signal.[16]

The proportion of the amino acid sequence detected after proteolysis alone varied with the three proteases and for the different glycoproteins (Table 2a–c). The complete sequence was never observed without deglycosylation and the only glycopeptides detected directly by proteolysis alone were some of the neutral truncated glycans on asparagines 19 and 270 of β-glucosylceramidase. When the peptides produced by proteolysis are deglycosylated (Table 2a–c), there is little or no difference in coverage compared to proteolysis alone, although in some cases deglycosylated peptides become apparent (Table 1a–c). When the proteins are deglycosylated prior to proteolysis there is, apart from the predigestion of transferrin with endoglycosidase F_3, an increase in the sequence coverage, especially when the results for the three proteases are combined (Table 1 and Figs. 3 and 6). The intensity of the proteolytic peptides also increases because predigestion with PNGase F results in higher signal to noise values for the proteolytic peptides. This effect not only increases the assay sensitivity but also may have implications for glycan analysis, because PNGase F is often used to release the glycans for analysis after proteolysis. Our observations indicate that enzymatic release of glycans by PNGase F is more efficient on the intact denatured proteins. It has been demonstrated that PNGase F can release glycans from short peptides but no information on the rate of release was given.[17] In contrast endoglycosidases F3 and H tend to cleave the glycans more efficiently after proteolysis, indicating a preference for peptides over polypeptides.

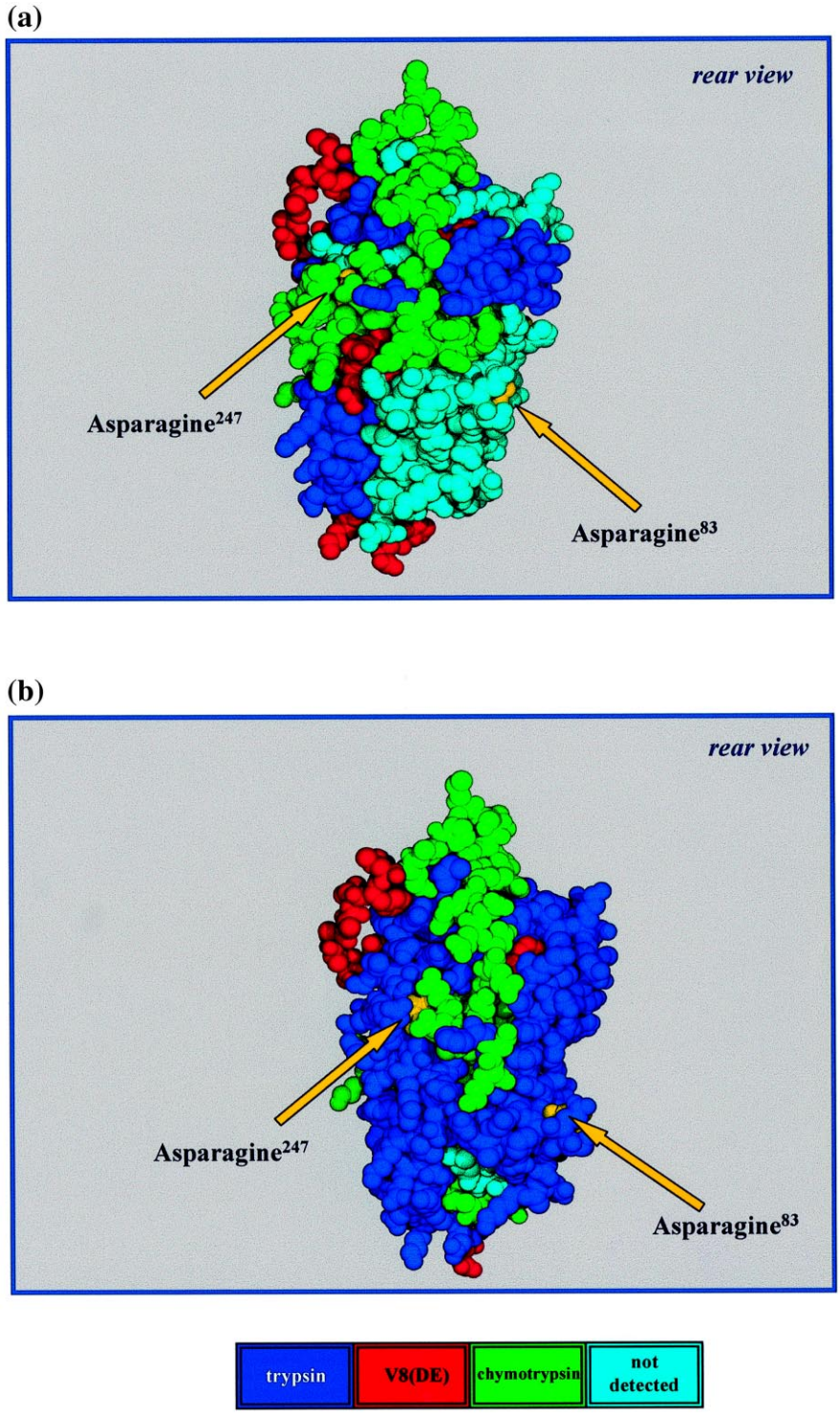

Fig. 6. Three-dimensional representation of the distribution of peptides detected after digestion of α1-antitrypsin by different protocols, showing the increase in proteolytic sequence coverage (a) before and (b) after truncation of the glycans with endoglycosidase F3

Glycosylation can also affect proteolytic sites remote from the glycosylation sites. For example, a peptide, corresponding to amino acids 233–254 with one missed cleavage site, was detected in the tryptic digestion of transferrin (Fig. 1a) but not after predigestion with PNGase F. However, a peptide 1689.9 Da/e (Fig. 1b), corresponding to the expected peptide amino acid sequence of 240–254, was consistently found after predigestion with PNGase F. This new peptide corresponds to the sequence generated by cleavage at the original missed cleavage site. Presumably this is due to removal of steric hindrance by the glycan or to an alteration in the conformation of the protein. This effect produces smaller peptides (Fig. 3c) and a more even distribution of peptides over the protein sequence (Fig. 3b) compared to proteolysis alone or proteolysis followed by deglycosylation. This information should be taken into account in peptide mapping studies using mass spectrometry. Apart from the action of endoglycosidase F3 on transferrin, the prior deglycosylation of the three proteins made the glycosylation sites amenable to analysis.

These results indicate that intact glycans on glycoproteins provide some protection against proteolysis around the sequons and at other sites (Fig. 3a). The truncated glycans generated by the endoglycosidases still seem to affect proteolysis as incomplete digestion is obtained in their presence (Fig. 3b). These observations support the theory that one of the functions of glycosylation is protection against protease activity in vivo.[1] The presence of glycans tends to direct proteolysis to particular regions of the molecule (Fig. 6). Prior deglycosylation can be used to increase the proteolytic coverage but several proteases may be needed to achieve full coverage. This is essential for mutation studies because single amino acid changes can be picked up by both mass difference (with the excellent mass resolution of MALDI TOF MS in this mass range) and changes in the pattern of proteolytic digestion (Fig. 4a, b).

The sites of glycosylation are clearly observed after deglycosylation by mass increments of 1 (PNGase F), 203.1 (endoglycosidases F3 and H) or 349.1 (endoglycosidase H with fucosylation) in the putative peptide covering the sequon. However, when a site is not occupied (see Section 4) then the predicted mass of the unmodified amino acid sequence of the peptide is observed.

In summary this work provides a rapid and straightforward method for the detection of glycosylation sites and the confirmation of changes in the amino acid sequence of glycoproteins using small amounts of purified protein. The use of a panel of enzymes is essential for studies where the majority of the protein sequence needs to be monitored. Once an informative digestion protocol has been established for a specific glycoprotein it can be applied routinely for quality control or diagnosis. Currently we are applying this strategy to the analysis of mutant glycoproteins in paediatric genetic diseases.

6. Experimental

6.1. Reagents

All chemical reagents used were research grade and obtained from Sigma–Aldrich Company (Poole, Dorset, UK) unless stated otherwise. Repelcote VS was obtained from BDH (Poole, Dorset, UK). Human serum transferrin and α1-antitrypsin were obtained from Sigma–Aldrich Company (Poole, Dorset, UK). Ceredase (human placental β-glucosylceramidase) and Cerezyme (recombinant human β-glucosylceramidase) were a gift from the Genzyme Corporation (Cambridge, MA, USA). Sequence grade endoproteinase Glu-C (protease V8), chymotrypsin and trypsin were obtained from Promega Ltd (Southampton, UK). Peptide N-glycanase F (PNGase F) (EC 3.5.1.52), endoglycosidase H and endoglycosidase F3 (EC 3.2.1.96) were obtained from Oxford GlycoSciences Ltd (Oxford, UK).

6.2. Protein derivatisation

All proteins were reduced and alkylated before proteolysis. 5–100 μg of protein was dissolved in an appropriate volume of 10 mM Tris–HCl buffer, pH 7.0, containing 0.1% (w/v) SDS and 0.15% (w/v) dithioerythritol, to give a protein concentration of 1 μg/μl. The solutions were incubated at 90°C for 60 s to reduce the proteins and left to cool to room temperature. 5 μl of 0.5 M iodoacetamide (an excess) was added to the mixture, which was incubated at 37°C for 30 min in the dark to carboaminomethylate the cysteine residues. Excess reagent and low molecular weight products were removed by ultrafiltration through Microcon YM-30 filters (Millipore, Watford, UK) and the protein solution was reduced by rotary evaporation.

6.3. Proteolytic digestion

The derivatised protein was reconstituted in H_2O to 1 μg/μl and 5 μl was added to 190 μl of the appropriate reaction buffer: for trypsin, 50 mM ammonium bicarbonate buffer, pH 8.5; for chymotrypsin, 100 mM Tris–HCl, pH 7.8; and for protease V8(DE), 50 mM sodium phosphate buffer, pH 7.8. 5 μl of the proteases trypsin (0.05 μg/μl), chymotrypsin (0.05 μg/μl) or protease V8 (0.1 μg/μl) was added to the reaction mixture. After incubation at 37°C for 16 h, the reactions were stopped by the addition of a cocktail of serine protease inhibitors (Sigma) or by heating at 90°C for 30 s. The digestion mixtures were desalted on silanised microcolumns, containing 5 mg of C18 stationary phase (Jones Chromatography, UK), which had been primed by the stepwise addition of 1 ml of 90% acetonitrile containing 0.1% TFA, 1 ml of 10% acetonitrile containing 0.1% TFA and 1 ml of 0.1% aqueous TFA. The digestion mixture was applied to the column, under gravity, and washed with 800 μl of 0.1% aqueous TFA. The desalted peptides/glycopeptides were eluted from the column with three volumes of 75 μl of 50% acetonitrile containing 0.1% TFA. The samples were concentrated and residual acetonitrile and TFA removed under vacuum before reconstitution in 0.1% aqueous TFA for mass spectrometry or in buffer for digestion by an endoglycosidase/peptide N-glycanase.

6.4. Digestions with endoglycosidases F3 and H and peptide N-glycanase F

All detergents were omitted from the endoglycosidase reactions because of the adverse affects observed on any subsequent proteolytic reactions, clean up of digestion products and MALDI TOF MS analysis. For digestion with endoglycosidase F3, 5–100 μg of the derivatised protein was dissolved in 7 μl of H_2O and added to 2 μl 0.4 M sodium acetate buffer, pH 4.5. After thorough mixing 2 mU of the enzyme in 1 μl of the buffer was added and the mixture was incubated for 48 h at 37°C. For digestion with endoglycosidase H, 5–100 μg of the derivatised protein was dissolved in 25 μl of H_2O and added to 25 μl of 200 mM sodium citrate phosphate, pH 5.5. After thorough mixing 10 mU of the enzyme in 10 μl of the buffer was added and the mixture was incubated for 48 h at 37°C. For digestion with peptide N-glycanase F, 5–100 μg of the derivatised protein was dissolved in 25 μl of 20 mM sodium phosphate buffer, pH 7.5, containing 250 mM EDTA. After thorough mixing 2.5 mU of the enzyme in 5 μl of 20 mM Tris–HCl buffer, pH 7.5, was added and the mixture was incubated for 48 h at 37°C. When the protein was deglycosylated after proteolysis, the proteolytic digestion mixtures were desalted and excess proteases removed by chromatography on C18 microcolumns as described above. When the deglycosylated protein was subjected to subsequent proteolytic digestion, the buffers used for the endoglycosidase digestion and the released glycans were removed from deglycosylated proteins by ultrafiltration through Microcon YM-30 filters (Millipore, Watford, UK). The residual endoglycosidase

protein was not removed prior to proteolysis because the target proteins were always present in large excess. As a control, each endoglycosidase was digested separately by each protease and the resultant peptide mixtures analysed by MALDI TOF MS. This allowed any minor peptides derived from the endoglycosidases to be excluded from the analysis of the target proteins.

6.5. MALDI TOF MS

Mass spectrometry was carried out on a matrix assisted laser desorption ionisation, time of flight instrument, fitted with a reflectron and a 337 nm UV laser (TOF Spec E, MicroMass, Manchester, UK). Analysis settings were performed at a source voltage of 20 kV, extraction 19 950 V, focus 16.5 kV and reflectron voltage of 25 kV. Spectra were acquired by averaging over a period of ten scans of highest signal. Data were acquired in reflectron mode, operating over a mass range of 6000 Da/e with matrix suppression set at 650 Da. Data analysis was carried out using MassLynx data analysis software, Protein Prospector database software at the University of San Francisco (http://falcon.ludwig.ucl.ac.uk/mshome3.2.htm) and PAWS proteomic analysis software (http://proteometrics.com/software/paws.htm). Protein structural information and database searching were performed using the SWISS-PROT database (Geneva, Switzerland) (http://www.expasy.ch/cgi-bin/sprot-search-ful).

Markedly increased sensitivity and lower background signals were achieved using recrystallised αC4HA as the matrix. Commercial αC4HA was purified by recrystallisation from an ammoniacal solution by the stepwise addition of concentrated hydrochloric acid. Brilliant yellow crystals were obtained after filtration and freeze drying. The matrix solution consisted of equal volumes of aqueous 50 mM fucose and acetonitrile:ethanol [50:50 v/v] containing 10 mg of αC4HA. 1.5 μl of a desalted digestion mixture reconstituted in 0.1% TFA was added to 1.5 μl of matrix solution and mixed thoroughly by aspiration. 1.2 μl of the sample/matrix mixture was applied to the MALDI TOF target to form a distinct droplet. The target was allowed to dry in air at room temperature for 30 min before being placed into the mass spectrometer. With DHB as the matrix, 1.5 μl of a desalted digestion mixture reconstituted in 0.1% TFA was added to 1.5 μl of a solution of DHB (1 mg/ml) in acetonitrile. After mixing thoroughly, 1.2 μl of the sample/matrix solution was applied to the MALDI TOF target, which was allowed to dry at room temperature. 0.5 μl of ethanol was added to the target, prior to mass spectrometry, to obtain a more homogeneous distribution of crystals over the target surface.[18] All peptides and glycopeptides were reconstituted in 0.1% TFA at a concentration of 550 fmol/μl before analysis by MALDI TOF MS.

Acknowledgements

The support of the Wellcome Trust and The Jules Thorne Charitable Trust is gratefully acknowledged.

References

1. Varki, A. *Glycobiology* **1993**, *3*, 97.
2. *New Comprehensive Biochemistry, Glycoproteins and Disease*; Montreuil, J.; Vliegenthart, J. F. G.; Schachter, H., Eds.; Elsevier Science: Amsterdam, 1996; Vol. 30.
3. Lejeune, P. J.; Mallet, B.; Farnarier, C.; Kaplanski, S. *Biochim. Biophys. Acta* **1989**, *990*, 122.
4. Reason, A. J.; Dell, A.; Romero, P. A.; Herscovics, A. *Glycobiology* **1991**, *1*, 387.
5. Packer, N. H.; Lawson, M. A.; Jardine, D. R.; Sanchez, J. C.; Gooley, A. A. *Electrophoresis* **1998**, *19*, 981.
6. Charlwood, J.; Clayton, P.; Keir, G.; Mian, N.; Winchester, B. *Glycobiology* **1998**, *8*, 351.
7. Jachymek, W.; Niedziela, T.; Petersson, C.; Lugowski, C.; Czaja, J.; Kenne, L. *Biochemistry* **1999**, *38*, 11788.

8. Stahl, B.; Klabunde, T.; Witzel, H.; Krebs, B.; Steup, M.; Karas, M.; Hillenkamp, F. *Eur. J. Biochem.* **1994**, *220*, 321.
9. Rahbek-Nielsen, H.; Roepstorff, P.; Reischl, H.; Wozny, M.; Koll, H.; Haselbeck, A. *J. Mass Spectrom.* **1997**, *32*, 948.
10. Miele, R. G.; Castellino, F. J.; Bretthauer, R. K. *Biotechnol. Appl. Biochem.* **1997**, *26*, 79.
11. Schmitt, S.; Glebe, D.; Alving, K.; Tolle, T. K.; Linder, M.; Geyer, H.; Linder, D.; Peter-Katalinic, J.; Gerlich, W. H.; Geyer, R. *J. Biol. Chem.* **1999**, *274*, 11945.
12. Tarentino, A. L.; Plummer, T. H. *Glycobiology* **1994**, *4*, 771.
13. Seta, N.; Gayno, S.; Jezequel-Cuer, M.; Toueg, M. L.; Erlinger, S.; Durand, G. *J. Hepatol*. **1997**, *26*, 265.
14. Hodges, L. C.; Chan, S. K. *Biochemistry* **1982**, *21*, 2805.
15. Friedman, B.-A.; Vaddi, K.; Preston, C.; Mahon, E.; Cataldo, J. R.; McPherson, J. M. *Blood* **1999**, *93*, 2807.
16. Mills, K.; Clayton, P.; Diettrich, O.; Winchester, B. G.; Johnson, A. W. *Rapid Commun. Mass Spectrom.*, submitted for publication
17. Fan, J. Q.; Lee, Y. C. *J. Biol Chem.* **1997**, *272*, 27058.
18. Harvey, D. J. *Rapid Commun. Mass Spectrom.* **1993**, *7*, 614.