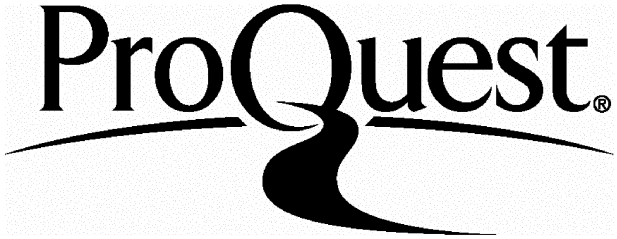
A proteomic approach to the study of inborn errors of metabolism

Kevin Mills

A thesis submitted for the degree of Doctor of Philosophy (Ph.D) in the Faculty of Life Sciences of University College London.

Biochemistry, Endocrinology and Metabolism Unit
Institute of Child Health
University College London Medical School

January 2003



'I do not know what I may seem to the world but as to myself, I seem to have been only like a boy playing on the sea-shore and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me.'

Issac Newton 1642 - 1727

'If we knew what it was we were doing,
it would not be called research, would it?"

Albert Einstein 1879 - 1955

Abstract

This work describes a preliminary study to evaluate the use of proteomics in the study of genetic metabolic disorders. Two metabolic disorders were studied initially, the congenital disorders of glycosylation (CDG) and the α_1-antitrypsin deficiencies, both diseases that result in changes in protein expression and post-translational modifications. Strategies were developed using mass spectrometry for the identification and analysis of proteins, in solution and after 2D-PAGE. Using endoglycosidases and proteases in different permutations, methods were developed to monitor the site-specific glycosylation and to increase the amino acid sequence coverage of the pure standard proteins α_1-antitrypsin and transferrin.

These methods were then applied to proteins purified from patient plasma by affinity chromatography, precipitation or 2D-PAGE. The glycosylation status of the plasma glycoproteins α_1-antitrypsin and transferrin was studied in CDG patients with types 1a, 1b, 1c and 1x enzymatic defects. Analysis of the α_1-antitrypsin underglycosylated species showed that the glycosylation sites were preferentially occupied in the order 46>247>83. In transferrin glycosylation site 413 was also preferentially occupied to site 611 and site occupancy was not a random process as thought previously. These data having implications for the role of glycosylation and its importance in stabilising the 3D structure of proteins during their maturation, and passage through the quality control system of the cell, prior to their delivery to their site of action. Similarly using 2D-PAGE and mass spectrometry the six major isoforms of α_1-antitrypsin were unequivocally identified in control plasma. Using a combination of glycan and peptide analysis, it was possible to confirm the identity of each isoform in the M series of α_1-antitrypsin. α_1-Antitrypsin was then analysed from 3 patients with known amino acid mutations in the α_1-antitrypsin sequence (V213A, E342K and T85M). All three mutations were detected in the α_1-antitrypsin expressed in plasma using both 2D-PAGE and MALDI TOF MS. In conclusion, this work describes the preliminary evaluation and the promise of proteomics as a tool in the study of genetic metabolic disease.

Papers published during the course of this work

Diettrich O, Mills K, Johnson AW, Hasilik A and Winchester BG (1998) Application of Magnetic Chromatography to the Isolation of Lysosomes From Fibroblasts of Patients with Lysosomal Storage Disorders. *FEBS Lett* **441**: 369-72.

Mills K, Johnson AW, Clayton PT, Diettrich OGP.and Winchester BG (2000) A Strategy for the Identification of Site-Specific Glycosylation in Glycoproteins using MALDI TOF MS. *Tetrahedron Assymetry* **11**: 75-93.

Mills PB, Mills K, Johnson AW, Clayton PT and Winchester BG (2001) Analysis by Matrix Assisted Laser Desorption/Ionisation-Time of Flight Mass Spectrometry of the Post-Translational Modifications of α_1-Antitrypsin Isoforms Separated by Two-Dimensional Polyacrylamide Gel Electrophoresis. *Proteomics.* **1**:778-86.

Mills PB, Mills K, Clayton P, Johnson A, Whitehouse D and Winchester B (2001) Congenital Disorders of Glycosylation Type I Leads to Altered Processing of N-linked Glycans, as Well as Underglycosylation. *Biochem J* **359**:249-54.

Mills K, Mills PB, Clayton PT, Johnson AW, Whitehouse DB and Winchester BG (2001) Identification of α_1-Antitrypsin Variants in Plasma with the Use of Proteomic Technology. *Clin Chem* **47**:2012-22.

Mills K, Mills PB, Clayton PT, Mian N, Johnson AW and Winchester BG (2003) The Underglycosylation of Plasma α_1-Antitrypsin in Congenital Disorders of Glycosylation Type I (CDG-I) is Not Random. *Glycobiology* **13**:73-85.

Mills PB, Mills K, Mian N, Winchester BG and Clayton PT (2003) Mass Spectrometric Analysis of Glycans in Elucidating the Pathogenesis of CDG type IIx. *J Inherit Metab Dis* **26**:1-16.

Table of contents.

List of Figures

Chapter 1

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Chapter 1

No tables

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Abbreviations

αC4HA	α-Cyano-4-hydroxycinnamic acid
[NO]-α_1-antitrypsin	nitrosylated α_1-antitrypsin
1D	One dimensional
2D	Two dimensional
ACN	Acetonitrile
ACTH (clip 18-39)	Adrenocorticotrophic hormone fragment 18-39
amu	Atomic mass units
APS	Ammonium persulphate
ATP	Adenosine 5'-triphosphate
β4Gal T	β1-4 Galactosyltransferase
Bis	N,N'-Methylene bisacrylamide
C-18	Octadecylsilane
CDG	Congenital disorders of glycosylation
CE	Capillary electrophoresis
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propane- sulfonate
CHO	Chinese hamster ovary cells
CI	Chemical ionisation
CMP	Cytidine 5'-monophosphate
CTP	Cytidine 5'-triphosphate
Da	Daltons
DAD	Defence against death
DHB	2,5-Dihydroxy benzoic acid
DNA	Deoxyribonucleic acid
DTE	Dithioerythritol
DTT	Dithiothreitol
E1	Ubiquitin activating enzyme
E2	Ubiquitin-conjugating enzyme
E3	Ubiquitin ligase
EDTA	Ethylenediamine tetra-acetic acid
EI	Electron-impact
Endo F3	Endoglycosidase F3
Endo H	Endoglycosidase H
ER	Endoplasmic reticulum
ERM1	Endoplasmic reticulum mannosidase I
ERM2	Endoplasmic reticulum mannosidase II
ESI	Electrospray ionisation
EST	Expressed sequence tag
FAB	Fast atom bombardment
Fuc	Fucose
Gal	Galactose
GalNAc	N-acetylgalactosamine
GC-MS	Gas chromatography mass spectrometry
GDP-mannose	Guanosine 5'-diphosphate
Glc	Glucose

GlcNAc T	N-acetylglucosaminyltransferase
GlcNAc	N-acetylglucosamine
h	Hour(s)
HPLC	High performance liquid chromatography
IEF	Isoelectric focusing
kDa	Kilo Daltons
kV	Kilo Volts
LC-MS	Liquid chromatography mass spectrometry
LLO	Lipid-linked oligosaccharide
LSI-MS	Liquid secondary ionisation mass spectrometry
m/z	mass to charge ratio
mA	milli Amps
MALDI	Matrix-assisted laser desorption ionisation
MALDI TOF MS	Matrix-assisted laser desorption ionisation time of flight mass spectrometry
Man	Mannose
min	Minute(s)
M_r	Molecular weight
MS	Mass spectrometry
OST	Oligosaccharyltransferase
PA	Polyacrylamide
PAGE	Polyacrylamide gel electrophoresis
PD	Plasma desorption ionisation
PDA	Piperazine diacrylamide
pI	Isoelectric point
PNGase F	Peptide N-glycosidase F
PP	Pyrophosphate
PSD	Post-source decay
PTH	Phenylthiohydantoin
QTOF	Quadrupole time of flight
s	Second(s)
SCX	Cation exchange resin
SDS	Sodium dodecyl sulphate
SPE	Solid phase extraction
TEMED	N,N,N',N'-Tetramethylethylenediamine
TFA	Trifluoroacetic acid
THAP	2, 4, 6 Trihydroxyacetophenone
TGN	Trans Golgi network
uv	Ultra-violet
UDP	Uridine 5'-diphosphate
V	Volts
V8 (DE)	V8 protease
VTC	Vesiculo-tubular cluster

Acknowledgements

Where do you start when you have so little space, to thank so many people, for so much help, culminating in this thesis? I list my thanks in the chronological order that each person have helped me.

I would like to thank my first science teacher at primary school, Mr Tim Jones, who started my curiosity with science using tin cans and boiling water to demonstrate the concept of pressure. I will always be indebted to him for his sound advice that we should never to be afraid to ask a question and to always ask why. My special thanks must also go to my chemistry teachers Mr Woodier and Mr Cole who taught me the basics of chemistry that have both helped and interested me to the present day.

In my time at the Institute of Child Health I have been fortunate enough to make many good friends including Minne Casteels, Kishore Iyer, Sarah and Kenny Young, Phil Whitfield, Eamonn O'Driscoll, Mike Dobbie, Joanne Charlwood, Catherine Evans, Ole Diettrich, Hugh Lemonde, Clare Beesely, Simon Welham, Nasi Mian, Ian Merryweather, Dave Stanton, Simon Eaton, Simon Pope, George Slim, Tom de Koning and Elaine White to name but a few. A specific mention must go to Phil, Hugh and Kishore who made working at ICH an absolute joy. Thanks must also go to Dr Andrew Johnson for being a never ending source of amusement during his time at the department.

During the course of this thesis I would like to thank Dr David Muller for his sound guidance, calming influence and always providing wise words during moments of insanity. I would like to thank the unsung hero of the department, my friend and colleague, Dr Nasi Mian, for all his unselfish help he has provided in the writing of this thesis. I would like to thank my supervisors Professors Bryan Winchester and Peter Clayton, for allowing me to do this PhD and especially for their believing in me when others didn't. No amount of thanks can be put into words to express my gratitude towards these two truly great men.

I would like to thank my mother, father and sister for their continued support and encouragement through my schooling and my undergraduate days at UWCC, during which time I wanted for nothing.

Finally I would like to thank my precious wife, Philippa, for her constant support and help during the writing of this thesis.

Chapter 1

Introduction

Chapter I - Introduction

Contents

1.1 Proteomics and the aims of this project.

The principal aim of this work was to develop methods for investigating co- and post-translational modifications, as well as changes in amino acid sequences in proteins using proteomic technology. The present study can be divided into two phases; the development and optimisation of methods for high resolution detection, identification and analysis of proteins and their application to the analysis of plasma proteins from patients with inherited genetic metabolic diseases, such as the α_1-antitrypsin deficiencies and the congenital disorders of glycosylation. The rarity of these diseases and the relatively small amounts of biological sample available from children suffering from these diseases render the use of conventional methods unsuitable.

Today the term 'proteomics' is used to encompass a myriad of protein characterisation techniques, from mass spectrometry coupled with 2-dimensional polyacrylamide gel electrophoresis (2D-PAGE), to anything remotely related to the quantitative measurement of protein content (Dove, 1999). The term 'proteome', which was first introduced by Wasinger and colleagues (1995) and defined as being 'the total set of proteins expressed in a given cell at a given time'. The study of the proteome has consequently been termed proteomics (Wilkins *et al.*, 1996). The explosion in proteomic research over the last 7 years has been due to the urgent need to bridge the gap between the genome sequence and cellular function of the proteins. Recent advances in technology now allow the study of the dynamic protein products of the genome as well as the static DNA blueprint of a cell using much less material. The genetic code allows the prediction of the amino acid sequence of a particular protein but provides little information about the level of expression, which may be regulated at both transcriptional and translational levels, or any co- and post translational modifications that the proteins may undergo which are vital for its function. The analysis of complex mixtures of proteins from pathological samples by proteomic technology can identify subtle changes in post-translational modifications. Conventional methods for the structural analysis of proteins usually require large

amounts of starting material. The advances in immobilised pH-gradients, imaging software, bioinformatics and mass spectrometry allow full automation of some aspects of proteomic technology. Currently, many pharmaceutical companies have the capability of fully automated analysis of hundreds of protein samples per week, by 2D-PAGE and the identification of individual separated proteins by various mass spectrometric techniques.

1.2 Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE).

Although a variety of analytical techniques can be used for the separation of protein mixtures, including multi-system array and chip technologies, 2D-PAGE is still the most popular protein separation technique. It remains unrivalled and unequalled, in its ability to separate complex mixtures of proteins and to unravel multigenic phenomena at the level of whole cell, tissue and even whole organisms (Banks *et al.*, 2000). Currently, no other technique can match 2D-PAGE in terms of resolution, and sensitivity, in the analyses of large mixtures of proteins (Dove, 1999). 2D-PAGE separates proteins by the sequential use at right angles of two different electrophoretic techniques based on two different properties of the proteins:-

(i) First dimension; the separation of the proteins according to their isoelectric point

and

(ii) Second dimension; the electrophoretic separation of the proteins in the presence of sodium dodecyl sulphate (SDS) according to their molecular weights.

The detection of proteins after 2D-PAGE utilises either antibody-detection techniques or chemical stains such as coomassie blue or silver. The proteins on the gel are detected as distinct spots which can be assigned 'X' and 'Y' co-ordinates according to their isoelectric point (pI) and molecular weights (M_r), respectively. These co-ordinates can be used to identify proteins quickly and to compare gels without further

analysis. 2D-PAGE was first developed by O'Farrell (1975) but has subsequently undergone several modifications designed to improve its resolution, reliability and sensitivity.

1.2.1 First dimension: isoelectric focusing (IEF).

The first step in the 2D-PAGE analysis is the separation of the proteins according to their isoelectric points (pI). Proteins carry a negative, positive or zero net charge depending on their amino acid composition and covalent modification (such as phosphorylation, nitrosylation, sulphation and glycosylation), and the pH of the environment. The pI of a protein is the pH at which the protein carries no net charge under those conditions. If the proteins are electrophoresed in a pH-gradient they will migrate until they reach a position in the pH-gradient where their overall net charge is zero i.e. the pH is equal to the pI of the protein (Figure 1.01 a-c). At this point the protein becomes 'focused' at its pI. A protein with a net negative charge will migrate towards the anode and proteins with a positive charge will migrate to the cathode. A protein will stop migrating when the pH is equal to its pI. If a protein starts to diffuse away from its pI, the combined effect of the electric field and pH-gradient will return it to its pI. Isoelectric focusing (IEF) allows the separation of proteins in a mixture according to very small differences in their isoelectric points. The original 2D-PAGE method described by O'Farrell (1975) used carrier ampholytes in tube gels to create and maintain, a pH-gradient. Carrier ampholytes are small amphoteric molecules with high buffering capacity near their isoelectric points and are usually employed as mixtures covering a set pH range. When an electric field is applied across a mixture of carrier ampholytes the most negatively charged move towards the anode and the most positive towards the cathode. In this way it is possible to form a continuous pH-gradient within a gel, which is suitable for the focusing of larger amphoteric molecules such as proteins. However, the cylindrical tube gels and carrier ampholytes used to create the pH-gradient had limitations in resolving power and other technical difficulties. The pH-gradients formed by carrier ampholytes are unstable and have a tendency to drift during longer focusing periods. These drifts are usually more

pronounced at the extremes of pH and make isoelectric focusing of basic proteins difficult at pH's greater than 8. The use of tube gels is also demanding technically because of their fragility. Technical expertise and reproducibility between different laboratories and batch to batch variation within the same laboratories were the main drawbacks to the universal use of 2D-PAGE as a major protein separation technique during the 1970's and 1980's.

In 1982 Bjellqvist introduced important modifications which involved the carrier ampholytes being co-polymerised into the gel matrix itself as it was being cast (Bjellqvist *et al.*, 1982). The creation of this ampholyte pH-gradient decreased markedly the cathodal drift observed using the tube gel system of isoelectric focusing. Further developments included the gel being cast onto a plastic support strip making the system much more robust and easier to handle. These innovations significantly improved reproducibility and performance of the first dimension focusing step (IEF) in 2D-PAGE. A modified method of Bjellqvist and colleagues (1982) is currently the standard technique employed by all proteomic laboratories (Gorg *et al* ., 2000).

Figure 1.01

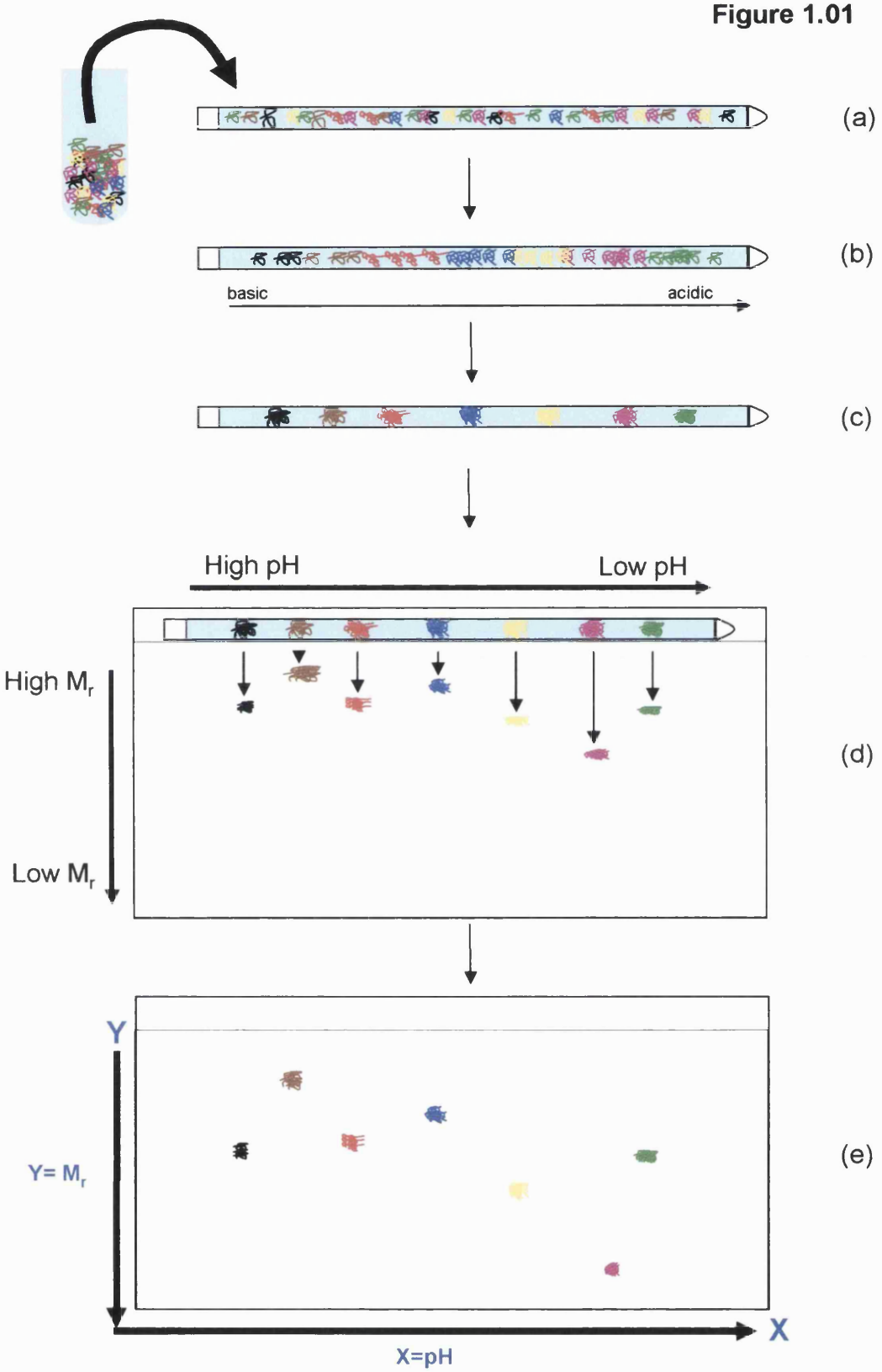


Figure 1.01. Schematic representation of 2D-PAGE protocol.

(a)-(c) Demonstrates the movement of proteins in an electrical field in a pH gradient. a protein will migrate towards the point in the gradient where the pH is equal to its pI.

(d)-(e) Demonstrates the transfer of the proteins from the 1st dimension to the resolving gel and the separation of proteins according to size.

1.2.2 Second dimension: SDS-PAGE

The separation of proteins in the second dimension of 2D-PAGE is based on differences in their electrophoretic mobility due to differences in their size. The second dimension utilises the traditional SDS-PAGE technique with the low % acrylamide IEF strip replacing the stacking gel. SDS is a very effective solubilising agent for a wide range of proteins. The majority of proteins bind SDS at a ratio of 1.4 g SDS / 1 g protein to form negatively charged complexes (Reynolds and Tanford, 1970). This results in masking of the intrinsic charge of the polypeptide chains, so that the net charge per unit mass becomes approximately constant. Although the majority of proteins bind SDS in the expected ratio, it is important to realise that proteins containing non-protein groups e.g. carbohydrates, phosphates, lipids can bind varying amounts of SDS and can result in anomalous mobility during SDS-PAGE (Dunn, 1995a).

During electrophoresis separation depends on the effective molecular radius, which approximates roughly to molecular size, as a result of the sieving effect of the gel matrix. The concentration of the polyacrylamide / cross-linker in the gel determines the sieving effect and hence the effective separation range of the SDS-PAGE.

The current procedure for SDS-PAGE is an amalgamation of methods devised by Laemmli (1970), a denaturing modification of the method of Ornstein (1964) and Davis (1964). Proteins are transferred electrophoretically from the IEF strip into a narrow starting zone prior to entering the main separating gel. This concentrates the proteins and results in very sharp bands or spots. This concentration of the proteins is due to a moving boundary formed between a rapidly moving (leading) and slowly migrating (trailing) ionic species. The re-equilibrated isoelectric focusing strip is placed horizontally onto the surface of the resolving gel (Figure 1.01d) containing the leading constituent (chloride ions), while the upper reservoir of the gel tank contains the buffer phase containing the trailing ion (glycinate) (Dunn, 1995a). As the sample proteins have intermediate mobilities, they are concentrated at the interface between the IEF strip and the resolving gel. The moving boundary is formed in the low % acrylamide IEF strip allowing it to behave like a traditional stacking gel. At the

interface with the resolving gel a higher pH is encountered, which causes further dissociation of glycinate, increasing its mobility so that it moves just behind the chloride ion. Once the protein samples have entered the separating gel, the negatively charged protein-SDS complexes continue to move towards the anode, and because they have the same charge per unit length they travel into the separating gel under the applied electric field with the same mobility. However, as they pass through the separating gel the proteins are resolved on the basis of their size because of the molecular sieving properties of the gel (Figure 1.01e).

1.3 Advances in mass spectrometry and their relevance to the analysis of proteins and peptides.

1.3.1 Development of mass spectrometry.

The first mass spectrometer was described by JJ Thompson (1899), who showed how ions could be deflected in an electric field in a sealed vessel under high vacuum. Since then many different forms of mass spectrometer have been developed. They are all based on three basic principles, the production of ions, the differentiation of those ions according to the mass to charge ratio (m/z) and the detection of ions under vacuum.

Mass spectrometry has proved an extremely valuable tool because it can provide molecular weights of compounds as well as their fragments with both high mass accuracy and sensitivity. However, a limiting factor in mass spectrometry has been the types of molecules amenable to ionisation. The first commercial mass spectrometers were fitted with electron impact (EI) and later chemical ionisation (CI) sources. These ionisation techniques required the molecule to be in the vapour phase before they could be ionised, which in turn required the use of high temperatures. Unfortunately, because of the size, involatility and thermal lability, many classes of compounds such as peptides and proteins could not be analysed initially by mass spectrometry.

1.3.2 Fast atom bombardment ionisation (FAB).

In 1981, Barber and colleagues described the use of noble gas atoms to bombard compounds suspended in a viscous matrix to produce ionisation of relatively large polar and labile molecules (Barber *et al.*, 1981). This type of mass spectral ionisation technique was designated as Fast Atom Bombardment (FAB) and termed a 'soft' ionisation technique because of its ability to ionise polar compounds which were often destroyed using traditional ionisation methods i.e. EI and CI. It also become apparent that FAB allowed the ionisation of relatively large polar molecules and peptides and small proteins up to a mass of 10,000 Da. In a typical FAB analysis the molecule is dissolved in a matrix such as glycerol or thioglycerol and bombarded with a high energy beam of xenon atoms (6-8 KeV). This results in the case of peptides and proteins, typically singly or doubly charged species that are observed in the mass spectrometer with little or no fragmentation of the original molecule (Figure 1.02a). Although, the mechanism of ionisation of molecules in FAB is not fully understood, it has been postulated that either the ions are preformed in the matrix by protonation or complexation with metal ions (Na^+ and K^+) or that the sample is desorbed as a neutral molecule and ionised in the high pressure region above the matrix-vacuum interface (Keough, 1985; Sunner *et al.*, 1986).

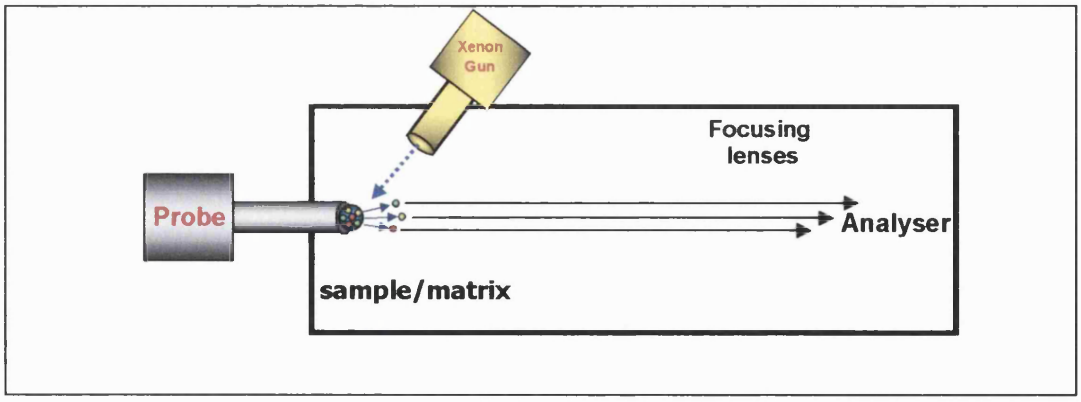


Figure 1.02. Schematic representation of a fast atom bombardment ionisation source.

Although the advent of FAB ionisation revolutionised the analysis of polar molecules by mass spectrometry and permitted the analysis of proteins by mass spectrometry it has some limitations (Nguyen *et al.*, 1995). Molecules which have extensive hydrogen bonding with the matrix tend to form homogenous solutions with concentration of the analyte, in no particular locus, within the droplet. Hydrophobic analytes, however, tend to concentrate on the surface of the droplet. Signal suppression of hydrophilic peptides is therefore often reported in FAB-MS analysis of a peptide mixture. This limits the interpretation of FAB mass spectra in peptide mapping experiments and in the quantification of analytes present in a sample, unless internal standards are used. Additional problems originate from sample-matrix interaction because of the nucleophillicity of some matrix components towards the proteins. Other chemical reactions observed during FAB-MS include exchange of sulphur atoms by oxygen, ring opening of lactones, hydrolysis and most importantly, the reduction of disulphide bonds. The reduction of the disulphide bonds during ionisation could obscure the correct assignment of disulphide bridges within a protein molecule. Another experimental consideration in the analysis of proteins by FAB-MS is ensuring that the sample mixture is free of salt or buffer. This requires the purification of the mixture by HPLC or other chromatographic techniques prior to the analysis (Nguyen *et al.*, 1995).

Other drawbacks of protein analysis using FAB are low sensitivity (100 pmol) and the inability to generate a good signal intensity with proteins of molecular weight greater than 10,000 Da because of their involatility and thermal instability. The introduction of Liquid Secondary Ionisation Mass Spectrometry (LSI-MS), where the xenon beam is replaced with caesium ions, has shown some improvement for proteins in the 15,000-20,000 Dalton range (Nguyen *et al.*, 1995). However in practice, FAB-MS and LSI-MS are useful for the analysis of proteins with molecular masses of no more than about 5000 Da. Nevertheless, FAB ionisation is still used extensively for the analysis of carbohydrate structures and bile acids (Dell and Morris, 2001; Clayton, 2001).

1.3.3 Plasma desorption ionisation (PD).

The first breakthrough in solving the problems associated with protein mass spectrometry came with the development of field desorption ionisation or plasma desorption mass spectrometry (PD-MS). Beuhler and colleagues (1974) observed that rapid heating of a sample could lead to preferential desorption of surface molecules. At the same time Torgerson and colleagues discovered that when high-energy fission fragments were produced from ^{252}Cf which was used to irradiate thin films of arginine and cysteine, intact molecular ions were observed (Torgerson *et al.*, 1974). Today, plasma desorption is used primarily for molecular mass determination of proteins. This technique provides very little fragmentation, and therefore little structural information can be obtained about the sample. The sensitivity of the technique is impressive with spectra having been obtained on pmol amounts of protein of molecular mass approximately 45,000 Da (Nguyen *et al.*, 1995). PD-MS has been a particularly attractive method for protein analysis because of its operational simplicity and high reliability.

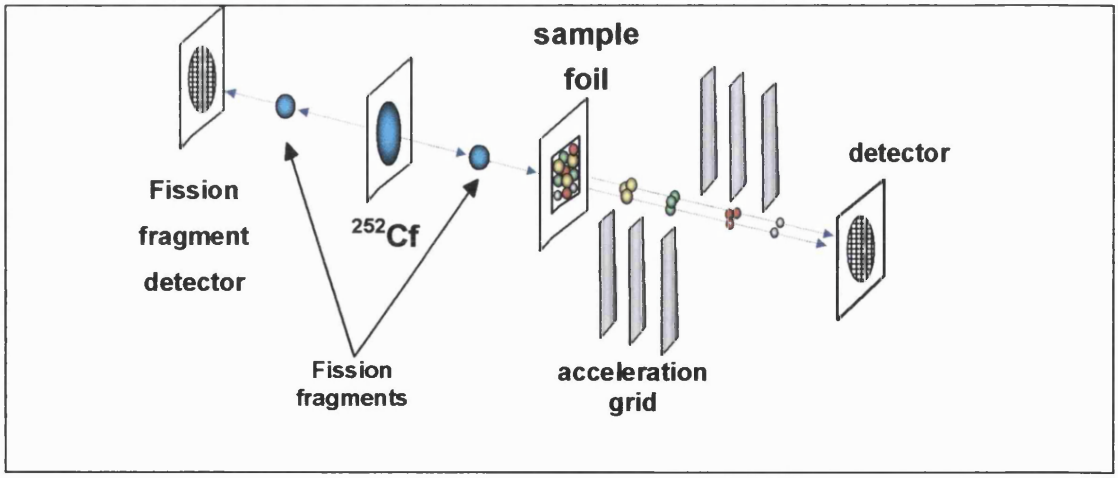


Figure 1.03. Schematic representation of a plasma desorption ionisation source.

Perhaps the most useful application of this technique is its use in peptide mapping by *in situ* enzymatic digestion. This procedure is normally performed on a nitrocellulose-bound sample that has already been used for molecular mass determination. A protease is applied to the bound sample and after an appropriate period of time, the proteolytic digestion is terminated by spin drying the sample target. The target containing the resultant peptide mixtures is then analysed by PD-MS (Tsabopoulos *et al.*, 1988).

PD-MS is based on the spontaneous fission of ^{252}Cf, which produces a pair of nuclear fragments $^{144}Cs^{20+}$ and $^{106}Tc^{22+}$. These fission fragments, with MeV energy, move in opposite directions (Figure 1.03). One of these fragments is oriented to hit the sample foil and ionise the sample molecule, whilst the other fragment is directed onto a fission fragment detector to record the time of the fission. The desorbed protein or peptide (after ionisation by the fission fragment) is then accelerated by a grid maintained at ground potential into a flight tube of the mass spectrometer and into the ion detector. Both detectors produce an electronic pulse when an ion is detected (Nguyen *et al.*, 1995). Plasma desorption ionisation techniques are usually used in conjunction with time of flight mass spectrometers (see section 1.3.5). The masses of the ions are calculated from the time of fission event and the time it takes for an ion to reach the detector. The time an ion takes to traverse the flight tube of the mass spectrometer is dependent on the mass to charge ratio of the ion (m/z). This mode of mass spectrometry, whereby ions are calculated by the time taken to cross a particular path, has been termed 'time of flight-mass spectrometry' (TOF MS). In principle, TOF MS has no upper mass limits, however, there are several factors which limit the mass-range of samples that can be analysed when a TOF MS is coupled to plasma desorption sources (Nguyen *et al.*, 1995).

The energy density generated by the fission fragments of ^{252}Cf when directed onto the sample is extremely high and most of the protein molecules undergo pyrolysis. This problem becomes evident when the molecular mass of the protein exceeds 50,000 Da. Above this mass range very few intact molecular ions can be detected (Nguyen *et al.*, 1995). Other limitations associated with PD-MS include the necessity of a radioactive source and restrictions to the class of compound it can analyse. Poor

results have often been observed with glycoproteins, highly acidic, membrane-bound proteins and carbohydrates.

1.3.4 ElectroSpray Ionisation (ESI).

The next development in the analysis of proteins and peptides by mass spectrometry was the introduction of electrospray ionisation process (ESI). ESI was first described in 1968 by Dole and colleagues in their studies on synthetic and natural polymers of masses in excess of 100,000 Da (Dole *et al.*, 1968). However due to the multiply charged ions formed in the ionisation process the technique was largely ignored. It re-emerged as a major innovation in mass spectrometry when Wong and co-workers were the first to demonstrate the ability of ESI to analyse accurately high molecular mass samples, such as polyethylene glycol (17,000 Da), bearing a net charge of up to +23 (Wong *et al.*, 1988). The unique ability of electrospray to produce multiply charged ions, which was an impediment initially was exploited in the analysis of large proteins in a mass analyser with a limited mass range (m/z <4000) and with an accuracy of greater than 0.1% (Covey *et al.*, 1988; Mann *et al.*, 1989). Typically ESI is interfaced with quadrupole mass spectrometers to enable the higher pressures (10^{-5} Torr) generated in the electrospray ionisation process to be tolerated.

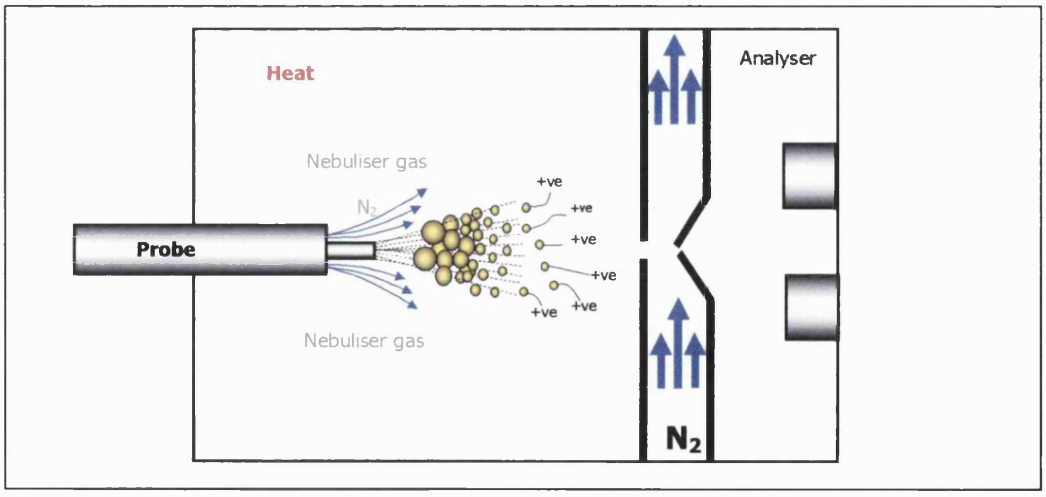


Figure 1.04. Schematic representation of an electrospray ionisation source

Electrospray is produced by applying a high electrical field to a relatively small flow

of liquid, which is sprayed from a capillary tube using a jet of nitrogen gas (Figure 1.04). The electric field causes the liquid surface to be highly charged and a spray of charged liquid droplets forms at the end of the capillary tube. The polarity of the charged droplets can be controlled by the applied polarity of the capillary. This is in contrast to FAB/LSI-MS in which both negative and positive ions are formed. The actual ionisation process is not yet fully understood but the general consensus is the hypothesis put forward by Dole and colleagues and is known as the Raleigh jet or charged residue mechanism (Dole *et al.*, 1967). ESI-MS involves a fine electrically charged spray leaving the tip of a capillary in the form of relatively large charged droplets. This spray is then countered by the application of a hot bath gas (usually N_2). This bath gas serves two purposes. The first is to counteract the large amount of condensable vapour produced from the spray nozzle and the second to increase the evaporation of solvent from the highly charged droplet so desorption of ions can occur. Once the droplet leaves the capillary and enters the bath gas it continues to lose solvent until the charge density of the droplet exceeds its surface tension (approximately 1.1×10^5 elemental charges). At this point, at which the Raleigh constant is exceeded, the droplet explodes (termed a Coulomb explosion) resulting in smaller charged droplets (Abbas and Latham, 1967). This process continues until the droplet is small enough for ion-desorption to occur. The desorbed ions are directed into the mass analyser through a series of focusing lenses. The advantage of multiply charging a protein is that it permits the use of a relatively low mass range (for example 0-2000 Da) for the analysis of very large proteins (for example up to 150,000 Da). Unlike FAB, LSI and PD-MS whereby one peak observed in the spectrum corresponds to one protein, in ESI-MS an envelope of peaks is observed for a single protein analysis. Interpretation of such spectra are relatively easy in homogenous samples with the aid of specialised deconvolution computer software. In order to determine the molecular mass of a protein using ESI-MS/MS the following presumptions must be made: i) the adjacent peak in an envelope of ions in the mass spectrum represents a species differing by one charge and ii) the charge is due to protonation of the molecular ion. The determination of two adjacent peaks in an envelope are then sufficient to determine the molecular mass (Nguyen *et al.*, 1995).

However, the more heavily a protein is glycosylated, the more complicated the mass spectral envelope becomes, due to the heterogeneity of the carbohydrate moieties which results in overlapping charge states.

Electrospray mass spectrometry is gaining popularity in protein chemistry because of its simplicity and versatility. The fact that ESI-MS can be coupled to HPLC allows the routine use of LC-MS. This facilitates the rapid on-line desalting of samples before they are passed into the MS for analysis. Applications such as peptide mapping, which could take hours or days using FAB-MS, can now be done with ESI-MS in a few hours. The popularity of ESI is due to its ability to determine the mass of proteins of M_r >100,000 Da, dynamically and with low pmol sensitivity. ESI-MS, does have some disadvantages, because of its intolerance to buffers and salts which form adducts causing ionic suppression and complicate the resulting spectra.

1.3.5 Matrix-Assisted Laser Desorption Ionisation (MALDI)

The development of modern laser technology has provided a means of directing a large amount of energy into a sample, leading to the desorption of intact molecules rather than the thermal decomposition as is sometimes observed in PD-MS. MALDI was introduced in 1988 by Tanaka, Karas and Hillenkamp (Tanaka *et al.*, 1988; Karas and Hillenkamp, 1988), working independently of each other. The MALDI technique uses a pulsed UV or infra red laser beam to ionise samples co-crystallised with the matrix and to desorb them from a stainless steel target (Zenobi and Knochenmuss, 1998). Sample preparation includes mixing the sample with a suitable matrix in a molar ratio excess of the matrix (1:1000 - 1:10,000). The sample / matrix mixture is deposited onto the target which is left to dry slowly leaving a crystalline deposit containing the sample. The matrix serves to reduce sample damage from the laser by absorbing the incident beam. A fast laser pulse (~4 ns) irradiates the target producing a burst of ions, which are then accelerated in an electric field (receiving constant kinetic energy) and travel to the detector. The lighter ions can travel faster than the heavier ions and reach the detector first (Figure 1.05). An ion of mass m when accelerated by a potential V_{acc} volts has a kinetic energy of $eV_{acc}=(mv^2)/2$, where e is

the electronic charge (TofSpec E & SE Operator Manual; Vestal and Juhasz, 1998). Therefore, for any given mass the velocity is equal to $v = \sqrt{(2eV_{acc}/m)}$. The velocity of an ion when it is accelerated through a fixed potential is inversely proportional to the square root of its mass. The time of flight of an ion is proportional to the square root of its mass (equation 1.) (TofSpec E & SE Operator Manual; Vestal and Juhasz, 1998).

Equation 1. $\sqrt{m} = At + B$

m = mass of the ion, A and B are instrument constants, t = time of flight of the ion

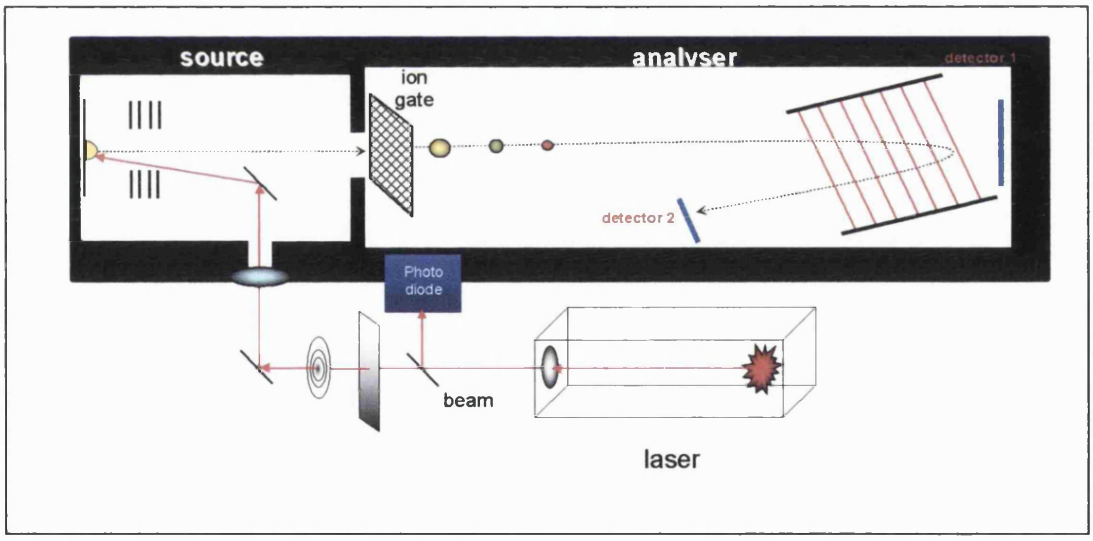


Figure 1.05. Schematic representation of MALDI - time of flight mass spectrometer.

However due to problems inherent in the ion formation in the MALDI process, the resolution of the first TOF analysers was poor in comparison to quadrupole and sector

instruments. This is because not all ions start with the same initial velocity and therefore do not reach the detector in tight packets. This leads to peak broadening and an accompanying loss of resolution (Figure 1.06). To overcome these problems delayed extraction and / or a reflectron, also described as ion mirrors, were added to TOF instruments (Kaufmann, 1995).

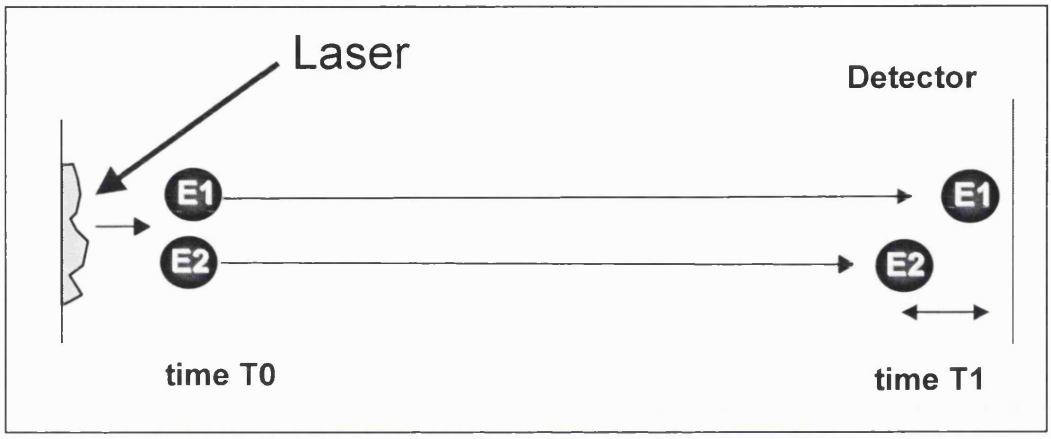


Figure 1.06. The ionisation process in MALDI.

Ions E1 and E2, with the same mass, do not emerge from the source at the same velocity and reach the detector over a time period of T1.

1.3.5.1 The reflectron or ion mirror.

The reflectron is a series of lenses (~12-20) which are capable of creating an increasing ion-retarding field. The mass spectrometer can be operated with or without the reflectron, the latter is sometimes referred to as linear TOF analysis. When the reflectron is in operation the ions are directed into the retarding field. Ions of higher velocity will penetrate deeper into the reflectron than an ion travelling with a lower velocity, before being redirected to another detector, which is usually situated at an angle to the primary flight tube. When the reflectron is in operation, the mass spectrometer is able to compensate for the initial uneven energy distributions and

improve resolution as represented in Figure 1.07 (Kaufmann, 1995).

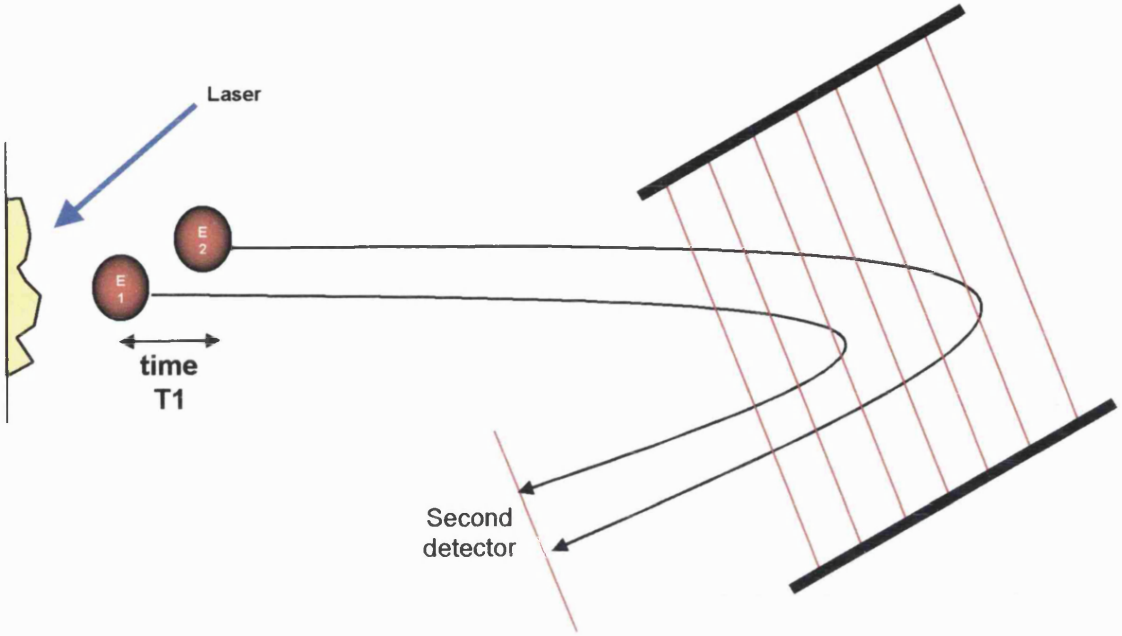


Figure 1.07 Principle of the reflectron.

Schematic representation of the flight path of the ions E1 and E2, which have different energies but the same mass.

This can be explained simply that by controlling the distance an ion penetrates the reflectron, the longer it will take to traverse the reflectron and reach the detector. How the reflectron can be utilised to compensate for the initial energy distribution and therefore improve resolution is shown in Figure 1.07 (Kaufmann, 1995).

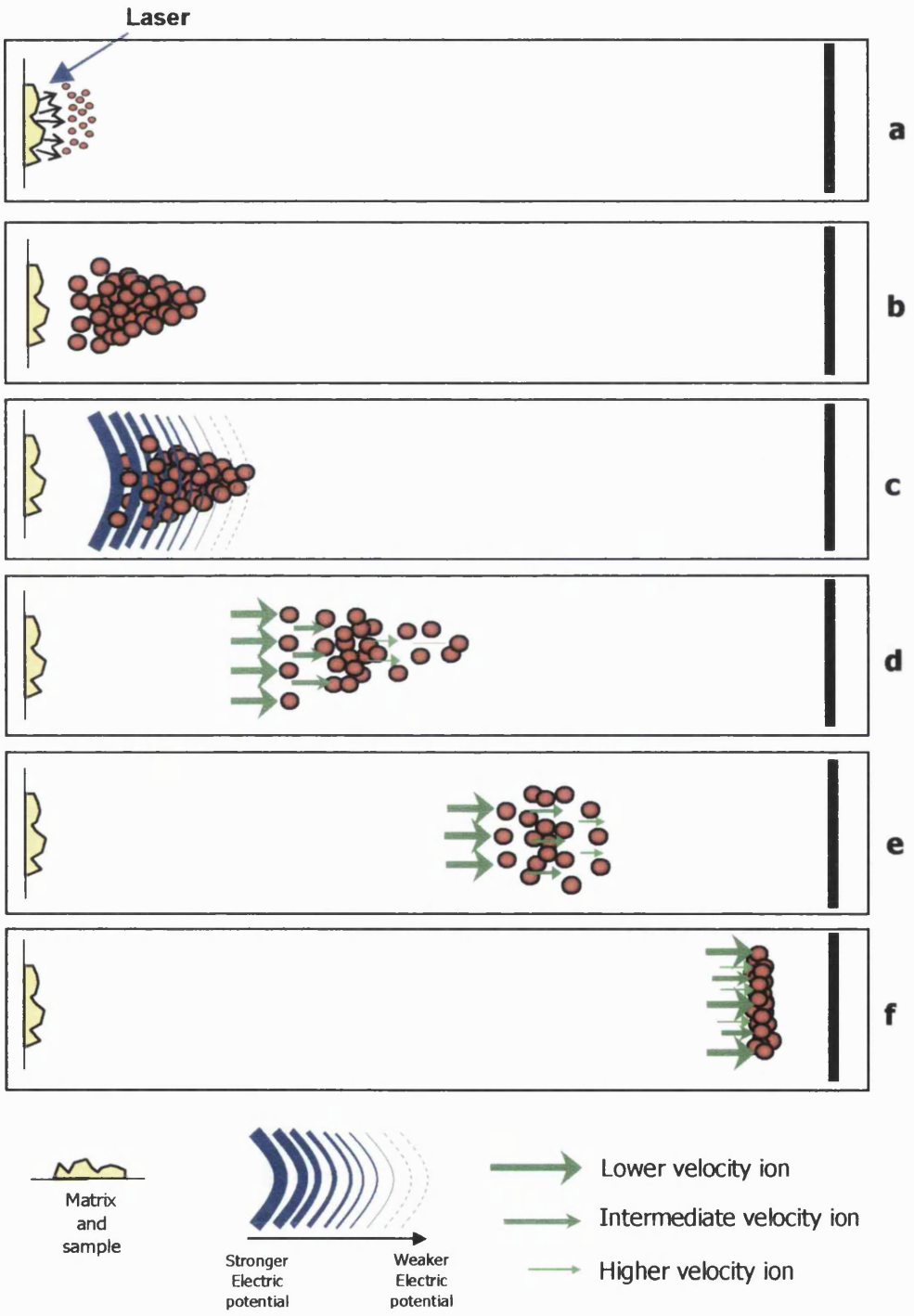
1.3.5.2 Delayed extraction or time-lag focusing.

Delayed extraction or time-lag focusing was first described by Wiley and McClaren (1955) as a means for compensating for any small changes in the initial velocities of ions after ionisation. Time-lag focusing like the reflectron, also improves mass resolution by reducing peak broadening. After the initial ionisation (Figure 1.08a), the ions produced are delayed in a near zero electric field between their formation and

their extraction from the source into the analyser (Figure 1.08b). During this time those ions with greater initial energy velocities will disperse further than those ions of lesser energy velocities (Figure 1.08c). At a specific time an electric field or extraction potential is applied to the ions to extract them out of the source and into the analyser region of the mass spectrometer (Figure 1.08c). During the application of the applied extraction potential those ions that had dispersed or travelled further away from the source in the course of the drift time period (time-lag), will receive less energy than those ions nearer the source and of lesser velocity. The variable extraction potential imparted to each ion, depending on its initial dispersion position, allows the 'velocity focusing' of each ion (Figure 1.08d and e). Consequently they reach the detector at the same time, thereby improving resolution (Wiley and McClaren, 1955) (Figure 1.08f).

With the incorporation of the reflectron into TOF MS analysers and delayed extraction to MALDI sources, the resolution achieved by MALDI TOF-mass spectrometers can now exceed that of triple quadrupole instruments. Sensitivity was generally the major advantage of MALDI over electrospray ionisation, with detection of peptides at the attomol level (Vorm *et al.*, 1994), with a good signal to noise spectrum being obtained with a few laser pulses from a single spot on the sample target. Several successive spectra can be obtained from a single spot, thus rendering the technique a relatively non-destructive technique. In practice, the actual sample amount consumed is so minute that samples of interest maybe recovered from the matrix for further analysis. MALDI has a higher mass range than any other current technique available at this time, with an upper limit of detection being of 350,000 Da at low pmol concentrations (Kaufmann, 1995). However, the distinct advantage of MALDI over other ionisation techniques is in its extreme tolerance to relatively high concentrations of salts and buffers. This permits protein analysis on crude extracts with little, or in some cases, no clean up of the sample.

Figure 1.08. Schematic representation of the process of time-lag focusing (delayed extraction).



1.4 Application of mass spectrometry to the analysis of proteins.

1.4.1 Molecular mass determination

The molecular mass of a protein is an important parameter in its biochemical characterisation. SDS-PAGE is a universal technique for determining the molecular mass of a protein, with an accuracy ranging from a few percent for a typical unglycosylated globular protein to about 30% for a heavily glycosylated protein (Nguyen, 1995). This makes the accurate determination of molecular mass by this method unreliable. With the recent advances in ionisation techniques it is obvious why mass spectrometry of a protein has become such a useful tool for protein molecular mass measurement. Techniques such as MALDI and ESI allow the determination of molecular weight of a protein at the low pmol level for masses in excess of 100,000 Da and with much greater precision.

ESI-MS can determine the mass of the protein with an accuracy exceeding 0.1%, which surpasses the best results obtained by SDS-PAGE. Although ESI can measure molecular masses above 150,000 Da (Siuzdak, 1994), the analyses become complicated to interpret as the mass exceeds this figure. Recent advantages in MALDI TOF instruments fitted with time-lag focusing have brought molecular mass determination to nearly comparable to that of ESI. However, although continuous flow MALDI analysis is being developed, it is not yet commercially available. Currently only ESI-MS methods for molecular mass determination are automated, with a high throughput of samples and the determination of the mass of a protein in minutes.

1.4.2 Peptide mapping

As the genomes of various organisms are sequenced, the ever increasing number of databases containing protein sequence information has become a valuable aid to

identification of proteins by mass spectrometry. A key component of strategies for defining molecular pathways is the development of a rapid and sensitive method for generating an 'address' with which a protein can be found in a database. Mass spectrometry has revolutionised this area.

Traditionally, short stretches of amino acid sequences were used to search databases to locate protein or gene sequences, or to identify proteins of similar sequence. The short amino acid sequence constituted a unique address for a protein and generally led to successful searches. The amino acid sequence information was usually obtained by N-terminal sequencing using Edman sequencers, a slow time-consuming process. In many cases the N-terminal of the protein was found to be blocked or modified, either naturally or artefactually, preventing the initial coupling reaction of the Edman degradation and prematurely terminating the sequencing.

In 1981, peptide mapping techniques using mass spectrometry were introduced by Henzel and colleagues (1993) to determine the masses of peptides resulting from the proteolytic and chemical cleavage of a protein, to verify gene sequences. It was quickly found that by combining molecular weight determination of peptides obtained from a proteolytic digest, with the information in a database (Geneva, Switzerland, http://www.expasy.ch), a powerful and rapid approach for protein identification was possible. By comparing experimental peptide masses from proteins after chemical cleavage or proteolytic digest, with the predicted theoretical masses (Protein Prospector, University of San Francisco (http://falcon.ludwig.ucl.ac.uk/mshome3.2.htm), it has been shown that a peptide mass map produces a highly informative fingerprint (Clauser *et al.*, 1999).

This technique is not only a rapid and sensitive method for the identification of known proteins but has also been useful for identifying frame shift /deletion /insertion mutations, co- and post-translational and chemical modifications and for confirming the gene sequence. This is achieved by detecting changes in mass resulting from modifications to the peptide fragment.

1.4.3 Post-translational modifications

1.4.3.1 Disulphide bond assignment

Many proteins contain disulphide bonds between sulphur atoms of cysteine residues. Characterisation of these bonds is very important, especially in recombinant protein technology, since the formation of disulphide bonds and protein folding are not under direct genetic control. Disulphide bridges play a key role in the determination of the tertiary structure of a protein. The native, correctly folded state of a protein is stabilised by these disulphide bonds and is important in the 3D-structure of a protein. Therefore a protein with the correct amino acid sequence but incorrectly formed disulphide bonds may not assume its active conformation. It is important to note that proteins with the wrong conformation can be distinct antigenically from the natural protein. Thus, it is important not only to establish the total cysteine /cystine content of a protein but also to determine which cysteines residues are involved in disulphide bridges.

Total cysteine content of a protein or peptide can be determined quickly and accurately using mass spectrometry. The current technique involves the complete reduction of all cysteines involved in disulphide bonds within the protein (using DTT or DTE) and the determination of the molecular mass by mass spectrometry. The reduced protein is derivatised with iodoacetamide, which acetylates the R-S-H group in a cysteine residue and the molecular mass determined by mass spectrometry. For every derivatised cysteine residue in the protein there will be an increment in the mass to charge ratio of 57, corresponding to the mass of the acetamide-derivatised cysteine residue. By subtracting the mass of the reduced protein from the mass of the derivatised protein and then dividing by the mass of acetamido group (57 m/z) (Nguyen *et al.*, 1995), it is possible to ascertain the number of cysteines present in the protein. Which cysteine bonds are involved in disulphide bridges, can be deduced by the derivatisation of the cysteine sulphur groups with iodoacetamide, without the prior reduction of the disulphide bridges using DTT or DTE. This way only the cysteines that are not involved in disulphide bridges are derivatised. The number of cysteine

residues involved in disulphide bridges can be determined by subtracting the mass of the derivatised but not reduced protein, from the mass of the fully reduced and derivatised molecule, and then dividing by the mass of the acetamide group.

1.4.3.2 Phosphorylation

Another common post-translational modification of proteins is phosphorylation of serine and threonine residues. Phosphorylation of a peptide or protein can be readily identified in mass mapping studies by a characteristic mass gain of 80 Da due to the presence of a phosphate group on the peptide fragment after cleavage or digestion. A mass discrepancy of 80 Da, from the expected peptide mass, would indicate the phosphorylation of the peptide (Zhang *et al.*, 1995). The peptide can then be analysed further by using additional digests to find the phosphorylation site. Alternatively the peptide can be sequenced to identify a mass increase of 80 Da above the expected mass of an amino acid. A rapid way of determining whether or not a peptide is phosphorylated is to use tandem-MS to perform a parent ion scan of 80 m/z. In this way only phosphorylated peptides within a mixture are detected.

1.4.3.3 Glycosylation

Glycosylation is the most common post-translational covalent modification observed in eukaryotic proteins. It can have a profound influence on the properties of the glycoprotein. These include biological activity, immunogenicity, clearance rate, solubility, thermal stability and proteolytic resistance (see section 1.6 for an in depth description of glycosylation). The study of glycosylation has become an important field of research, especially in understanding human disease and in the therapeutic use of recombinant protein technology, where rapid and accurate monitoring of glycosylation patterns is vital in quality control of the product. Recombinant glycoproteins generally exist as a set of glycosylated variants exhibiting heterogeneity with respect to both the proportion of potential glycosylation sites that are occupied i.e. macroheterogeneity, and the oligosaccharide structures present at each glycosylation site i.e. microheterogeneity (Varki, 1993).

The precise analysis of site-specific glycosylation and the structural analysis of carbohydrates are still major challenges for conventional analytical methods. A number of techniques for structural analysis of glycoproteins are available including, lectin-binding, various forms of chromatography, flurophore-assisted carbohydrate electrophoresis (Mechref and Novotny, 2002) and the use of highly specific endo- and exoglycosidase enzyme assays (Guile *et al.*, 1996). However, these methods can be slow, laborious or require relatively large amounts of sample. Hence, mass spectrometry is becoming a major technique for the analysis of protein glycosylation. It is capable of providing complete structural analysis of the branched nature of the glycans, linkage of one monosaccharide unit to its neighbour and the number of possible structural sugar isomers that may be involved. More significantly mass spectrometry requires substantially less glycoprotein for analysis than that of other procedures. A flow chart illustrating how mass spectrometry may be used in the analysis of glycoproteins and carbohydrates is shown in Figure 1.09.

One of the quickest and simplest ways of gaining a significant amount of information on the carbohydrate content of a protein can be obtained from an

Figure 1.09

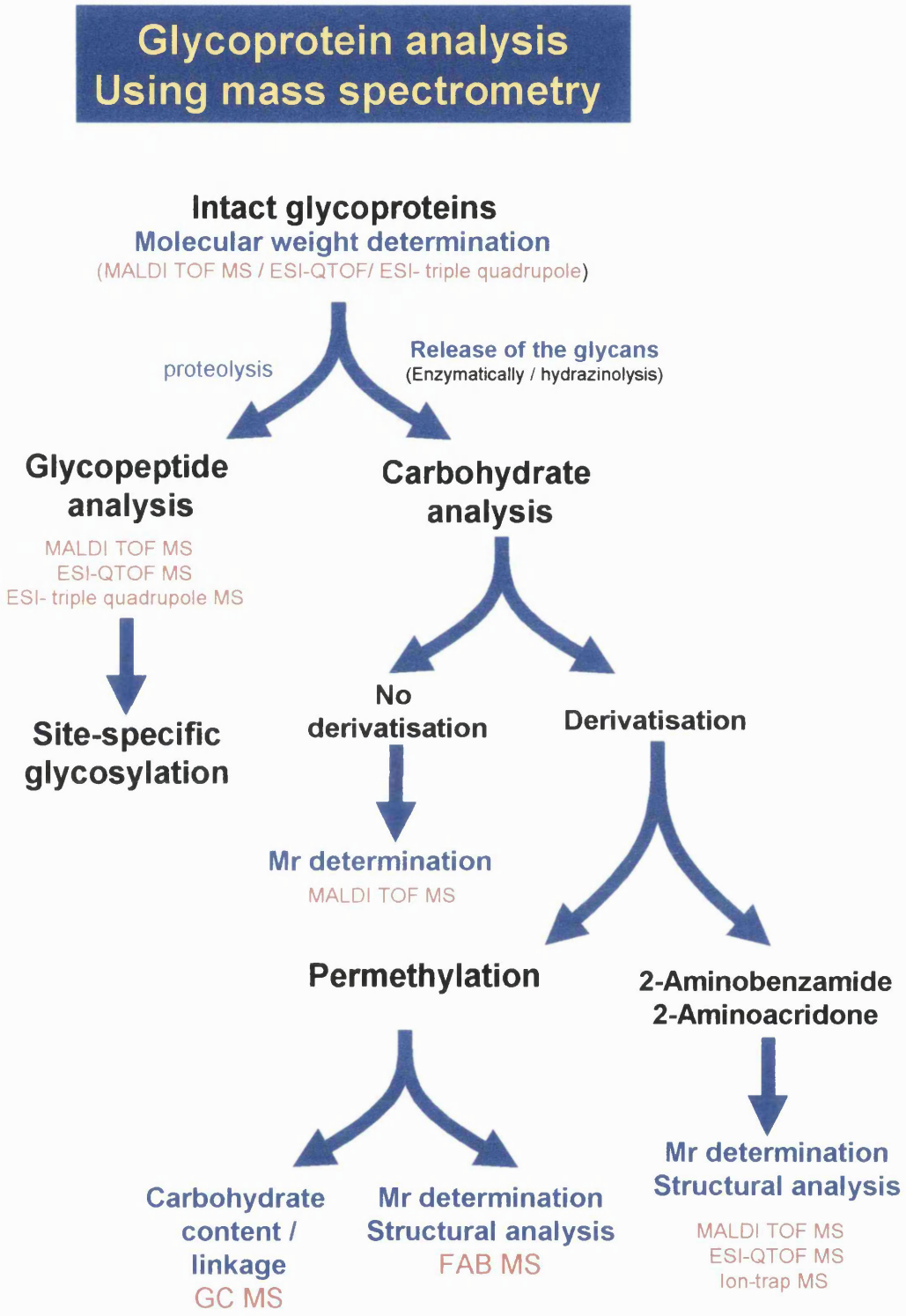


Figure 1.09 **Flow chart showing how mass spectrometry can be used to study glycosylation.**

accurate measurement of its molecular mass. Mass spectrometry can provide molecular mass measurements of glycoproteins with an accuracy of approximately 0.1%. MALDI and ESI-MS can be used to record the spectra of intact glycoproteins but the capability of these instruments to resolve individual glycoforms depends not only on the mass of the protein of interest but also on the degree of glycosylation. The resolution of TOF instruments will in general only permit the identification of small glycoproteins with a limited number of glycans attached (Harvey 2002). Greater resolution of individual glycoforms has been achieved with ESI instruments for glycoproteins with a high level of homogeneity (Yamashita 1993). Studies on the congenital disorders of glycosylation have shown how MALDI TOF MS and ESI-MS can be used to detect the presence or absence of glycosylation and how they can be used to study the disease process (see Chapter 6) (Wada *et al.*, 1992; Yamashita *et al.*, 1993). Both groups demonstrated that purified serum transferrin from normal controls had a mass of Mr 79,600 Da, whereas transferrin from a CDG Type-I patient exhibited two extra ions of molecular masses 77,400 and 75,200 Da. These correspond to species lacking one and two complex biantennary oligosaccharides, respectively.

Although determination of the macroheterogeneity (site occupation) can provide much information about the glycoprotein, it is also important to study the diversity of the glycan structures (microheterogeneity) at each glycosylation site. To accomplish this, the glycan chains have to be removed from the protein either enzymatically or chemically. The N-linked glycans can be released enzymatically using several commercially available endoglycosidases. The precise specificities of the different enzymes permit the identification of the class of glycan attached to the asparagine residue i.e. complex, high mannose or hybrid. The endoglycosidase most widely used is peptide N-glycosidase F (PNGase F) which releases N-linked glycans except those containing fucose $\alpha 1 \rightarrow 3$ linked to the reducing terminal GlcNAc, as commonly encountered in plant glycoproteins. This enzyme cleaves the glycosylamino linkage between the asparagine of the glycosylation sequon and the glycan moiety, releasing the intact glycan. Using this strategy it is not only possible to mass map the molecular ions of the glycan moieties that have been released but it is also possible to

investigate the site-specific glycosylation of proteins (Mills *et al.*, 2000). The enzymatic removal of a glycan converts the amide group of the asparagine residue to a carboxylic acid resulting in the formation of an aspartic acid residue. This results in an increase of 1 Da in the mass of the deglycosylated peptide which can be detected easily using mass spectrometry.

Although several endoglycosidases are available commercially that can release N-linked glycans from proteins, there is only one endoglycosidase that can release O-linked glycans. However, O-glycosidase (O-Glycanase) will only release unsubstituted, Ser/Thr –linked Gal-GalNAc from glycoproteins. More complex O-linked structures can only be removed when this enzyme is used in conjunction with additional enzymes. Chemically it is possible to non-selectively remove intact O- and N-linked glycan structures from the polypeptide using hydrazine. However, the use of hydrazine can be potentially hazardous and results in the denaturing of the protein. Hence, the precise site of glycosylation on the protein cannot be determined (Harvey, 2001).

Analysis of the glycan by MS, once it has been released from the protein, may or may not necessitate chemical modification or derivatisation. This will depend on the type of ionisation process used, the amount of sample available and the analysis required (Harvey, 2001).

More than 90% of all naturally occurring complex oligosaccharides, are made up of only four glycosyl residues, with incremental masses of 146 (fucose), 162 (hexose e.g. mannose, glucose, galactose), 203 (N-acetyl hexosamine) and 291 Da (sialic acid), respectively. Thus, after deglycosylation and purification of the carbohydrates, recording of the molecular mass of even a large carbohydrate moiety can yield the glucosyl composition. This molecular weight information can facilitate predicting the putative composition of the oligosaccharide structure. MALDI is arguably the most sensitive of the 'soft ionization' techniques and is ideally suited to carbohydrate analysis of 'valuable' clinical samples. Currently, MALDI TOF MS is the pre-eminent technique for screening molecular ions because it permits a simple and rapid analysis of unmodified glycans that have been released from a glycoprotein. FAB has also been used extensively to determine the molecular weight of glycan structures,

although the analysis requires derivatisation and is not as sensitive as MALDI TOF MS (Dell and Morris, 2001).

Even though the molecular weight can provide much information on the structure of a glycan it is often necessary to determine the branching, linkage, configuration and identification of the same-mass sugar isoforms of glycans. This can be investigated by several mass spectral techniques. The individual content of the different hexose units that make up the glycan and their linkages to one another can be determined using gas chromatography MS (GC MS) and FAB MS/MS analyses on multi-sector or tandem mass spectrometers (Dell 1987; Sutton-Smith *et al.*, 2000). However, these analyses usually require relatively large amounts of starting material and the permethylation of the glycans to increase their volatility and the sensitivity of the analysis. It has also been shown that derivatisation of the non-reducing end of the sugar moiety using 2-aminoacridone or 2-aminobenzamide can also facilitate the analysis of glycan structures using MALDI TOF MS, QTOF MS or ion trap mass spectrometry (Harvey, 1999).

Another powerful ionisation technique for the structural analysis of carbohydrates is electrospray ionisation mass spectrometry. The advantage of ESI is that it can be coupled to separation methods such as LC, HPLC and CE for use in conjunction with a range of mass analysers including QTOF and ion-trap mass spectrometers (Gennaro *et al.*, 2002; Chai *et al.*, 2001; Zamfir and Peter-Katalanic 2001).

1.4.4 Sequencing of peptides.

Analysis of the primary structure of proteins has improved due to technological advances in protein purification, techniques for handling small samples, peptide sequencing and amino acid analysis. Automated Edman degradation is routinely carried out to sequence the first 20-30 residues of a peptide or protein using less than 100 pmol. The chemistry involves stepwise release of amino acids from the exposed N-terminus of the peptide or protein. The overall sensitivity of this technique is limited by the ability to detect by UV-absorbance the phenylthiohydantoin (PTH) derivative of the amino acid that is released in each cycle. This becomes increasingly difficult as the yields of PTH-amino acids fall and the background rises with each degradation cycle due to prior incomplete cleavages. There are three other main problems associated with Edman degradation;

(i) A relatively large amount of sample is required along with extensive work up so that the peptide is in pure format before analysis.

(ii) It is time-consuming.

(iii) The sequence reaction proceeds from N-terminus of the peptide. If the first amino acid in the sequence is blocked for example by acetylation then the reaction will not commence. Any modified amino acids will appear as a blank cycle in the sequence.

The introduction of FAB-MS in 1981, and the initial attempts at sequencing using mass spectrometry (Barber *et al.*, 1981), showed the potential for overcoming the difficulties with the Edman technique. Today, with MALDI TOF, ion-trap and tandem mass analysers, sequencing can be performed in seconds, with no extensive clean up necessary. Peptides have been fully sequenced at the low fmol level (Schevchenko *et al.*, 1996).

Unlike the Edman procedure, sequencing by mass spectrometry does not start from one terminus or another. The mass spectrometer generates a series of fragmentation ions by shattering the parent ion into many species of various sizes, the masses of

which can be measured. Each of these fragments differs from the next by one amino acid in the simplest of cases, and each amino acid in the sequence is identified from the mass difference between successive peaks. In practice the fragment pattern observed will be a mass spectrum consisting of many peaks which can correspond to the mass of one amino acid to the molecular ion of the peptide under analysis. All mass spectrometers provide sequence data on this basis but differ in the way they produce the fragmentation of the peptide.

In peptides the fragmentation generally occurs at the peptide backbone, however cleavage at peripheral bonds can occur, particularly in magnetic sector instruments. To understand peptide fragmentation, a nomenclature was proposed by Roepstorff and Fohlman (1984, Figure 1.10).

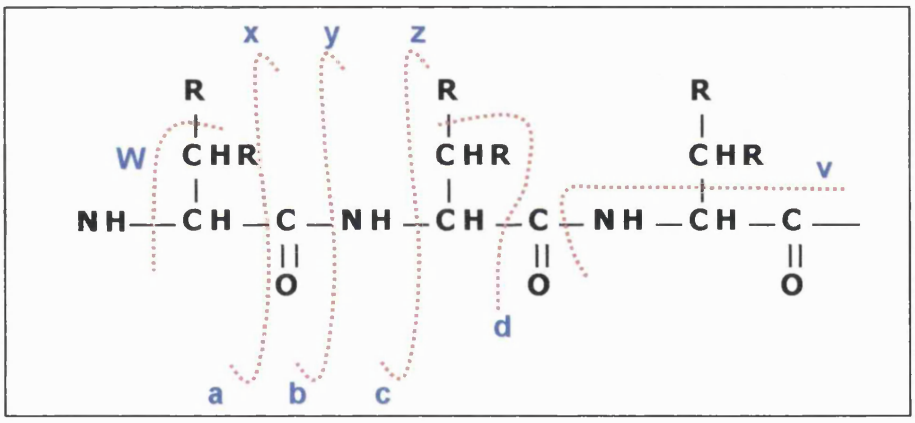


Figure 1.10 Nomenclature for peptide fragmentation.

Fragment ions of the type a_n, b_n and c_n are generated if the charge is retained on the N terminus of the peptide. The x_n, y_n and z_n ions are formed if the charge is retained at the C terminus. Fragment ions corresponding to side chains are also seen in some spectra and are denoted d_n (N-terminus), v_n (C-terminus), w_n (C-terminus). The d_n ions are very informative as they provide a means to differentiate between the isomeric amino acids leucine and isoleucine. If leucine were the amino acid at the terminus, an isopropyl radical would be eliminated, whereas, a methyl or ethyl radical will be eliminated with isoleucine (Roepstorff and Fohlman, 1984).

1.4.4.1 Sequencing using MALDI TOF MS

Although the sequence data obtained using different types of mass spectrometry are fundamentally the same, the mechanism by which the peptide fragmentation occurs depends on the type of mass spectrometer used.

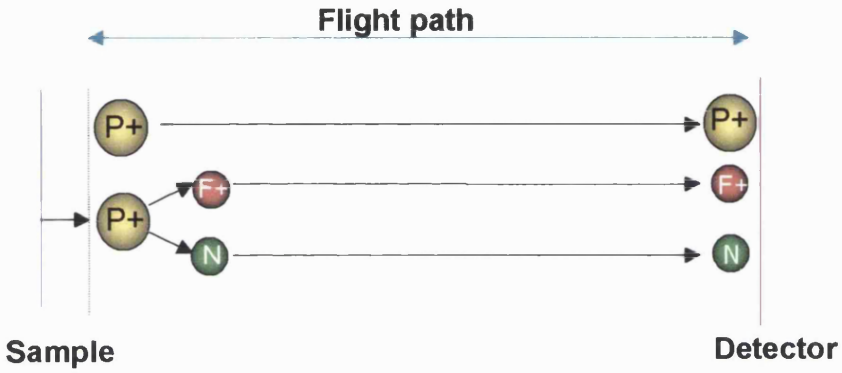


Figure 1.11. Post source decay of ions observed during MALDI TOF MS.
In linear TOF mode an intact parent ion (P+) reaches the detector at same time as a charged fragment (F+) and thus registers as one ion. The neutral fragment (N), also reaches the detector at the same time as the parent (P+) and fragment ion (F+), and together they register as one ion.

The fact that in linear time of flight machines no prompt fragmentations were observed, led to the assumption of high ion stability in MALDI TOF MS (Kaufmann, 1995). However, a closer look at this problem revealed that abundant fragmentation occurs based on unimolecular and bimolecular reactions. Since practically all these fragmentations occur after the ions have left the acceleration region (during flight) they go undetected in linear TOF machines because an intact molecular ion (P^+) will arrive at the detector at the same time as an ensemble of its decay products (N and F^+) (Figure 1.11, Kaufmann, 1995).

If, however, a retarding field is placed in front of the detector, the extent of the fragmentation can easily be determined because the arrival of the ions (F^+) and neutral fragments (N) at the detector will be staggered (Figure 1.12).

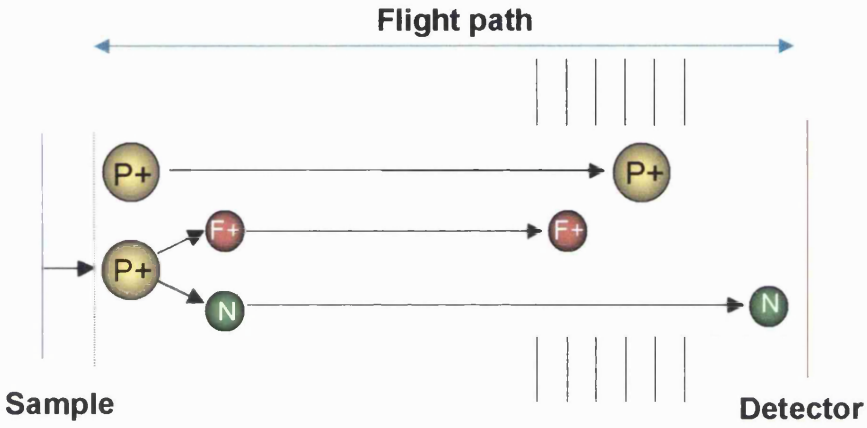


Figure 1.12. A MALDI TOF MS fitted with retarding ion lenses.
The use of the retarding lenses allows the differentiation of the ions P+, F+ and the neutral fragment N.

If a TOF instrument is fitted with a reflectron and a second detector, the neutral fragment will remain undetected as it is unaffected by the field produced by the reflectron. The parent and fragment ions will be turned around in the reflectron and detected at the second detector (Figure 1.13). The lighter fragment ions will reach the detector sooner than the heavier ions, as a result of the lighter ions not being able to penetrate as deep into the reflecting fields as the heavier ions and thus leaving the reflectron sooner (Figure 1.13). The fragment ions travel with the same velocity as the precursor ions, but with less kinetic energy, therefore they cannot penetrate the retarding field of the reflectron as deeply as these precursors and consequently leave

earlier (Kaufmann, 1995).

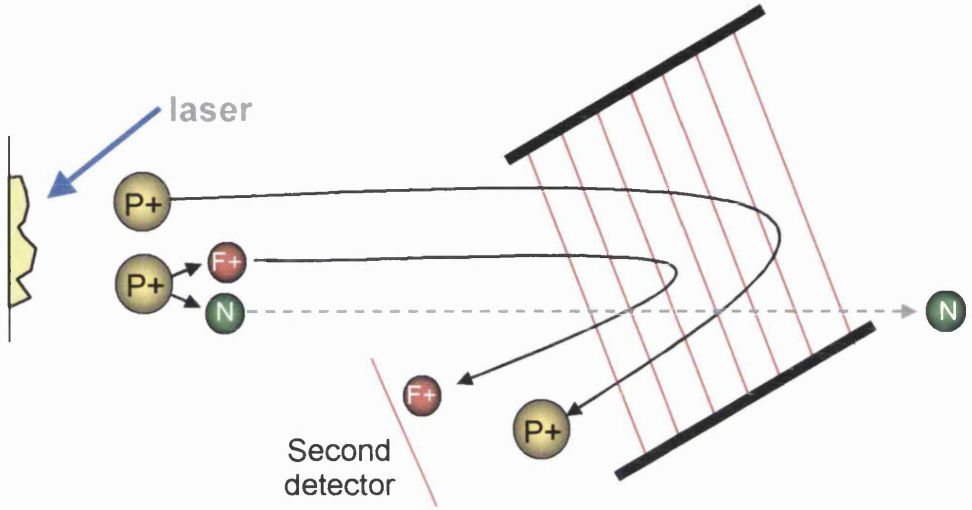


Figure 1.13. TOF instrument fitted with a reflectron.
Charged fragment ions (F+) turn around earlier in the reflector than the heavier parent ion (P+) and reach the second detector more quickly.

This phenomenon was termed post source decay (PSD) by Spengler *et al.*, (1992) and is the basis for the detection of the fragmentation seen by MALDI TOF MS. The fragmentation is due to the multiple collisions of analyte ions with the matrix ions during the early plume expansion and ion acceleration, followed by collisions with residual gas molecules in the drift region of the flight tube. Some MALDI sources also have collision cells for this purpose, situated in between the source and the flight tube. These are used to give d_n ion fragmentation data. Thus post-source decay or unimolecular decomposition is exploited in conjunction with a reflectron to gain peptide fragmentation spectra. In practice the peptide is selected for sequencing from an enzymatic digest or mixture, using a beam-blanking device called a 'Bradbury-Nielson' ion gate (Figure 1.05, see section 1.3.5). This device has a high enough electric field to prevent any ions from reaching the reflectron detector and is situated between the source and the reflectron. The ion gate can be switched on and off to allow only mass ions of interest (determined by the flight time) into the reflectron for

analysis. During the fragment analysis, the operating voltages of the reflectron have to be reduced stepwise to cover the full mass range of possible ions and 10-12 spectral segments have to be recorded sequentially. These segments are then 'stitched together' using computer software to give the complete sequence spectra.

Unfortunately, fragmentation of peptides during PSD is often an unpredictable and uncontrollable phenomenon because it often depends on the chemistry of the peptide. Some peptides are inherently stable to PSD and thus give little or no fragmentation data. Conversely, some peptides may undergo such extensive fragmentation that the spectra obtained are impossible or extremely complicated to interpret. Although the MALDI TOF MS is the most sensitive of mass spectrometers in the detection of peptides and proteins, relatively large amounts of sample are required for sequencing (at least 10 pmol). However, it is the unpredictable nature of PSD fragmentation that prevents MALDI TOF MS being the first choice in sequencing by mass spectrometry.

1.4.4.2 Sequencing using quadrupole-time of flight mass spectrometers (QTOF)

The quadrupole-time of flight mass spectrometer (QTOF) is a hybrid mass spectrometer because it combines two mass spectrometry technologies to create the most powerful, commercially available, mass spectrometer for proteomic analysis. The QTOF is the amalgamation of quadrupole / collision cell technology coupled with an orthogonally situated time of flight analyser (Figure 1.14).

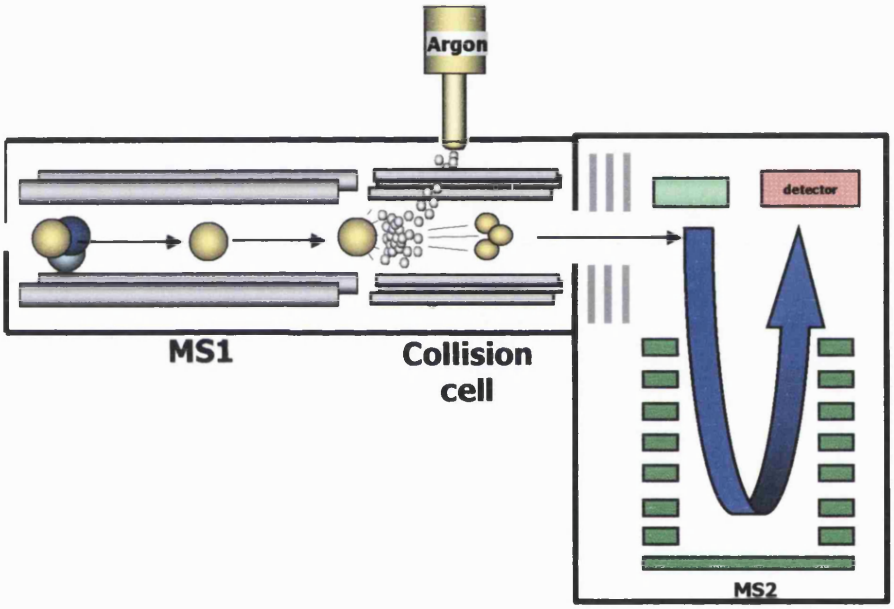


Figure 1.14. Schematic representation of a hybrid quadrupole time of flight mass spectrometer.
MS1 is the first set of quadrupoles used to select the ion for fragmentation in the collision cell, while MS2 is the TOF analyser used to record the peptide fragments.

The ESI source and quadrupole mass filters facilitate the creation of multiply charged ionic species, which are good for intact protein molecular weight determination and fragmentation studies, with a concomitant high ion transmission rate into the TOF sector of the instrument. The hexapole collision cell, situated between the quadrupole

mass filter and the TOF analyser, allows collision-induced fragmentation ions to be created with much greater control of the fragmentation than is observed in PSD analyses with MALDI TOF MS. The TOF analyser provides superior sensitivity and mass resolution in the instrument (Chernushevich *et al.*, 2001).

The QTOF mass spectrometer has a major advantage over other types of mass spectrometers because it acquires sequence information rapidly and simultaneously for all peptides present in a proteolytic digest. The high resolution data gained from the proteolytic digest can be used to search mass mapping data bases. If a definitive match is not achieved then expressed sequence tag searching can be employed. It is this ability to gain automated sequence data, with excellent sensitivity and mass resolution, which gives the QTOF its superior capability over MALDI TOF MS.

The QTOF acquires sequence data from a peptide using the same principles employed in triple quadrupole instruments in which the peptide of choice is directed into the collision cell by a first set of quadrupoles (MS1) and the daughter ions are focused onto the detector using another set of quadrupoles (MS2). In QTOF MS, the MS2 quadrupoles and detector are replaced by a TOF analyser which is used to record the fragment ions leaving the collision cell. The fragment ions leaving the collision cell are introduced into the orthogonally situated flight tube of the TOF by a rapidly pulsing electrical field. This electric field serves to repulse the ion out of its flight trajectory and into a new flight path within the TOF analyser. Sensitivity of 15 fmol during scanning mode and 100 fmol for high quality sequencing data, and its high mass resolution capabilities are the advantages of QTOF MS. Perhaps the only disadvantage of a QTOF instrument is its cost, which is typically 2-3 times that of a triple quadrupole or MALDI TOF MS.

1.4.4.3 Sequencing of peptides using carboxypeptidases

Although, the sequencing of peptides by mass spectrometry usually exploits fragmentation of the peptide to yield sequence information, it is not the only way in which sequence information can be obtained. Carboxypeptidases are exoproteinases that will sequentially cleave amino acids from the C-terminus of a peptide. By

incubating the peptide to be sequenced with carboxypeptidase and monitoring the subsequent masses of the products produced by mass spectrometry as a function of time, it is possible to deduce the sequence of the peptide (Nguyen *et al.*, 1995). Aliquots can be removed from the reaction mixture over a period of time, quenched and stored for later analysis. Using PD-MS, the digestion can be carried out directly on the nitrocellulose support. However, using ESI-MS the digestion mixture can be continuously monitored by direct infusion (a method also used to study enzyme kinetics by mass spectrometry). Using these methods carboxypeptidase P has been used to fully sequence superoxide dismutase (Nguyen *et al.*, 1995). However the use of carboxypeptidases in sequencing of peptides is not without its problems, mainly because the digestion can often stop or become extremely slow at certain amino acids (tryptophan, aspartic acid). The use of carboxypeptidases has only been used in the sequencing of well characterised proteins. An unknown protein would present a much greater challenge and currently the method is usually carried out to confirm sequence data.

1.5 Protein glycosylation.

1.5.1 The different forms of glycosylation.

Glycosylation is an important, highly conserved, covalent post-translational modification which confers subtle structural and functional characteristics on the protein. O- and N-glycosylation are the most common form but C-linked glycosylation has also been reported (Spiro, 2002). The O-glycosylation of proteins is based on O-glycosidic linkage between GalNAc and the hydroxyl group of serine or threonine residues. The GalNAc can be elongated by the stepwise addition of other glycosyl residues and can result in branched structures. This type of glycosylation is common in both prokaryotes and eukaryotes (Spiro, 2002). Another type of O-glycosylation involves the addition of a single GlcNAc residue on serine or threonine residues on nuclear and cytoplasmic proteins (Hart, 1997).

N-glycosylation, which is unique to eukaryotes, is a co-translational modification that is carried out in the lumen of the endoplasmic reticulum (ER) of the cell following translocation of the nascent polypeptide across the membrane of the ER. A tetradeca oligosaccharide unit ($Glc_3Man_9GlcNAc_2$) is transferred *en bloc* from a lipid carrier onto an asparagine in an Asn-X-Ser/Thr sequon in the nascent polypeptide, catalysed by an ER membrane-bound, multi-subunit complex called oligosaccharyltransferase (OST) (Aebi and Hennet, 2001; Helenius and Aebi, 2001). The common precursor oligosaccharide is assembled on the lipid carrier, dolichol pyrophosphate in the membrane of the ER. The newly glycosylated protein is subjected to quality control in the ER before it is transported to the *cis* Golgi apparatus. As the glycoprotein passes through the compartments of the Golgi apparatus the N-linked glycans are processed to the wide range of structures seen on mature glycoproteins (Roth, 2002; Helenius and Aebi, 2001). The steps in the biosynthesis of N-linked glycans are described in more detail in section 1.5.2.

1.5.2 N-glycosylation.

1.5.2.1 Assembly of the common lipid-linked oligosaccharide precursor (LLO).

Under normal physiological conditions the LLO is synthesised in a highly ordered step-wise manner on dolichol phosphate (Figure 1.15), which is anchored in the lipid bilayer of the ER membrane. The dolichol phosphate is produced by phosphorylation of the ER membrane-bound dolichol by a CTP-dependent protein kinase. The dolichol is also used directly for the synthesis of dolichol-P-mannose and dolichol-P-glucose, which also act as glycosyl-donors in the synthesis of LLO (Varki *et al.*, 1999a).

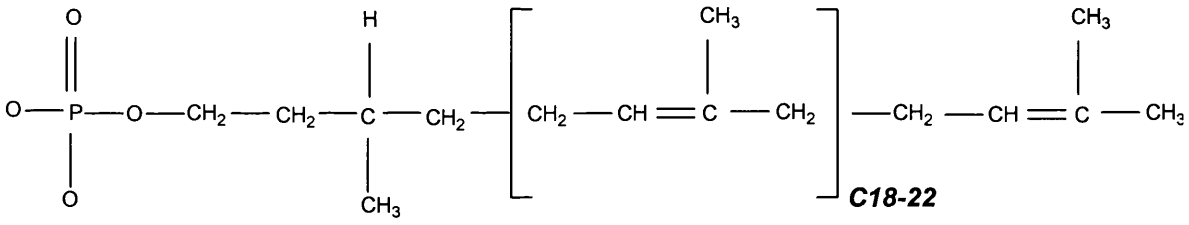


Figure 1.15. The structure of dolichol phosphate.

Dolichol is a vital component of cellular membranes and is synthesised *via* the mevalonic acid pathway as are cholesterol and ubiquinone (Figure 1.16). After transfer of the common oligosaccharide precursor from the dolichol-PP, and mannose and glucose from dolichol-P-mannose and dolichol-P-glucose, respectively, the dolichol-pyrophosphate and dolichol-phosphate are converted back to dolichol by

specific pyrophosphatases and phosphatases. The dolichol can then be recycled for LLO synthesis. Any disturbances in the flux of dolichol can result in impaired synthesis of the LLO precursor and consequently protein N-glycosylation (Kornfeld and Kornfeld, 1985).

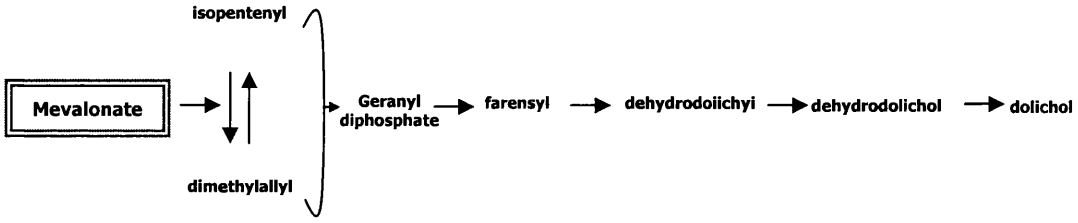


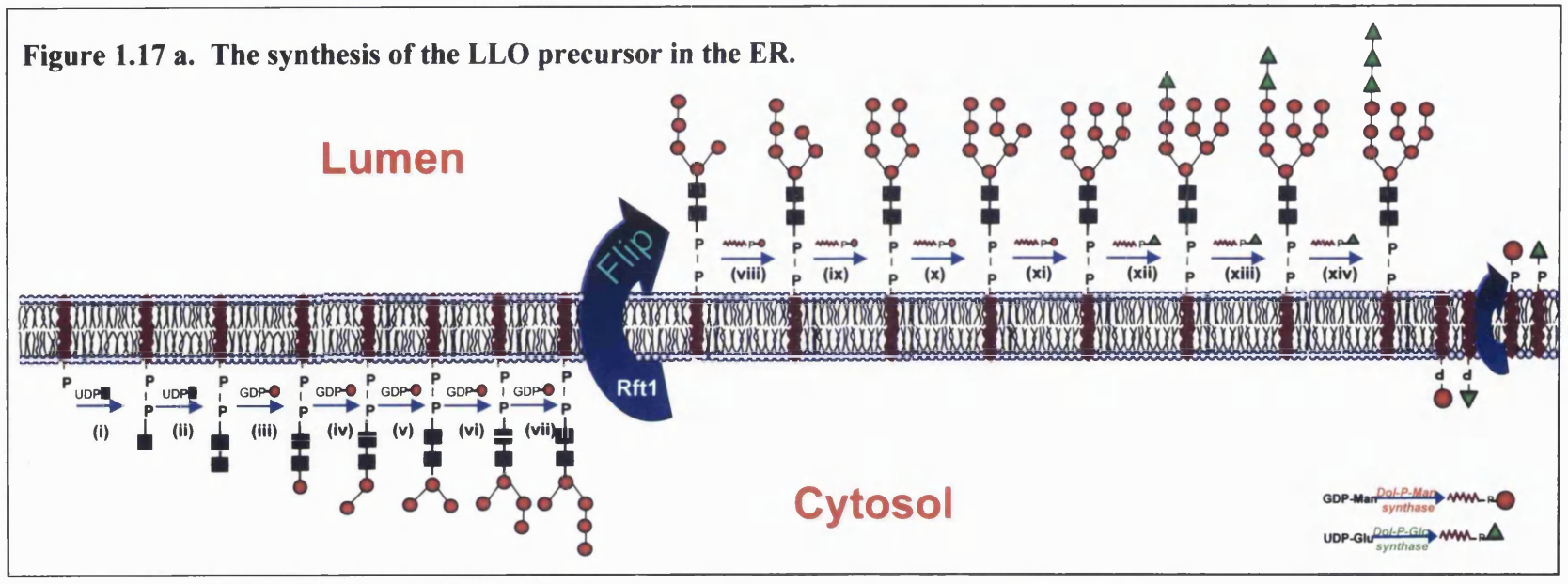
Figure 1.16 Synthesis of dolichol via the mevalonate pathway.

1.5.2.2 Glycosylation of dolichol phosphate in the synthesis of LLO

The first step in the synthesis of the LLO is the transfer of GlcNAc-1-phosphate from UDP-GlcNAc as donor substrate to dolichol-phosphate by GlcNAc-1-phosphotransferase to form dolichol pyrophosphate N-acetyl glucosamine. This is followed by the addition of another GlcNAc residue from UDP-GlcNAc as a donor to make $GlcNAc_2$-PP-dolichol catalysed by GlcNAc-transferase (Figure 1.17a, steps (i) and (ii)). Five mannose residues are then added in a highly ordered manner to the LLO chain using GDP-mannose as the donor to produce a heptasaccharide chain, $Man_5GlcNAc_2$-PP-dolichol (Helenius and Aebi, 2001). This synthesis is carried out on the cystolic side of the membrane and is shown in Figure 1.17a (steps i-vii).

The LLO intermediate is then translocated or 'flipped' onto the luminal side of the ER membrane. Recently, Helenius *et al.*, described a yeast gene RFT1 which encodes a protein involved in the translocation of the $Man_5GlcNAc_2$-PP-dolichol precursor into the lumen of the ER (Helenius *et al.*, 2002).

Figure 1.17 a. The synthesis of the LLO precursor in the ER.



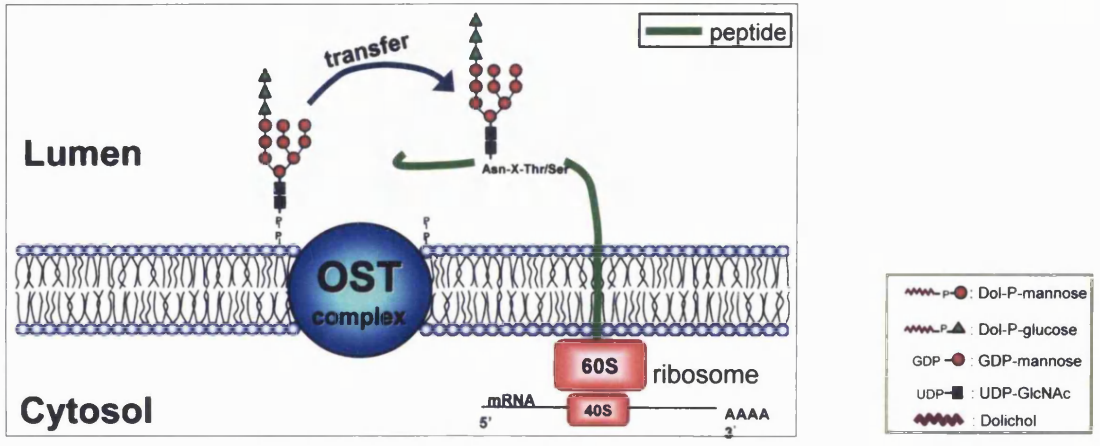


Figure 1.17b. Transfer of the $Glu_3Man_9GlcNAc_2$ oligosaccharide to the nascent polypeptide by OST.

Once the $Man_5GlcNAc_2$-PP-dolichol is facing the lumen of the ER, four mannose residues are added onto the α 1-6-linked mannose branch of the $Man_5GlcNAc_2$-PP-dolichol, again in a highly ordered manner. Dolichol-P-mannose is the glycosyl donor and highly specific mannosyltransferases catalyse the formation of the $Man_9GlcNAc_2$-PP-dolichol intermediate (Helenius and Aebi, 2001) (Figure 1.13a steps viii-xi). Finally three glucose residues are added sequentially onto the peripheral mannose residue on the α1-3 branch of the $Man_9GlcNAc_2$-PP-dolichol, by three distinct ER-membrane bound glucosyl transferases using dolichol-P-glucose as the donor (Figure 1.13a, steps xii to xiv). This completes the synthesis of the complete LLO ($Glc_3Man_9GlcNAc_2$-PP-dolichol, Helenius and Aebi, 2001).

1.5.2.3 Transfer of the oligosaccharide precursor to the protein.

The $Glc_3Man_9GlcNAc_2$ oligosaccharide is transferred *en bloc* from the LLO precursor to an asparagine residue in an Asn-X-Ser/Thr motif in the nascent polypeptide, catalysed by the ER-membrane-bound oligosaccharyltransferase complex (OST). The OST complex in yeast is known to consist of at least nine subunits (Knauer and Lehle, 1999) whereas only four subunits have been identified so far in mammalian OST (Kelleher *et al.*, 1992; Kelleher and Gilmore, 1997). They include, ribophorins I and II, OST 48 and a fourth subunit called DAD1 (Harnik-Ort *et al.*, 1987; Crimaudo *et al.*, 1987; Silberstein *et al.*, 1992; Kelleher *et al.*, 1992; Fu *et al.*, 1997; Kelleher and Gilmore, 1997). The DAD1 subunit (defence against death) was first described as an anti-apoptotic molecule (Nakashima *et al.*, 1993). All proteins destined to be N-glycosylated have to traverse the ER lumen whilst they are being translated and where they may present themselves to the OST complex. In most cases, the polypeptides carry a signal sequence on the N-terminus which targets the protein to the ER. The signal sequence comprises 12-30 predominantly hydrophobic amino acids e.g. MKWVTFLLLLFISGSAFSRGVF and is recognised by a cytoplasmic signal recognition particle (SRP) and to which it binds tightly. This SRP then docks onto the ER-membrane-bound signal recognition peptide receptor, which promotes its translocation across the membrane into the lumen of the ER (Austen and Westwood,

1991). Once the polypeptide is in the lumen of the ER the signal sequence is usually removed by a signal sequence peptidase, although some proteins do retain their signal sequence in the mature form.

An important step in the N-glycosylation of a protein is the recognition of a specific amino acid sequence denoting the glycosylation motif or sequon (Kaplan *et al.*, 1987). This sequence Asn-X-Ser / Thr, where X can be any amino acid except proline, is recognised by the OST complex as a possible site for the covalent attachment of the $GlcNAc_2Man_9Glc_3$ oligosaccharide to the asparagine residue (see Figure 1.17b). The hydroxyl groups present in the amino acid side chains actually participate in the catalytic reaction of the OST complex. Imperiali and Hendrickson (1995) have shown that the OST catalysed glycosylation reaction requires a loop conformation similar to a β-turn to facilitate hydrogen bonding between the asparagine side chain, the peptide backbone and the hydroxyl group of the serine or threonine residue. Although the presence of this sequence is required for N-glycosylation, the sequence may not necessarily undergo N-glycosylation. Apweiler and colleagues (1999) analysed the 75,000 proteins contained in the Swiss-Prot protein database (http://www.expasy.ch/sprot/). Almost two thirds of these proteins contain possible N-glycosylation sites at an average of 3.1 per protein. However, only 10.6% of the total proteins have been confirmed as being N-glycosylated (Apweiler *et al.*, 1999), indicating that not all potential glycosylation sequons are glycosylated.

Several *in vitro* studies using site directed mutagenesis have investigated the factors that influence the frequency of N-glycosylation. Bause (1983) demonstrated that Asn-X-Thr sequons have an approximately 40 fold greater chance of being glycosylated than the same peptide with an Asn-X-Ser motif. Shakin-Eshleman and colleagues (1996) and Mellquist and colleagues (1998) have shown that the X and Y positions in the glycosylation motif (Asn-X-Ser / Thr-Y) are both important in influencing whether a sequon is glycosylated. Threonine and serine in the X or Y positions are the most favourable amino acids for glycosylation of a sequon.

1.5.2.4 Initial processing of N-linked glycans and the folding of glycoproteins.

After the transfer of the oligosaccharide structure to the polypeptide, the terminal glucose residue of the protein-linked $Glc_3Man_9GlcNAc_2$ is immediately removed by the ER-membrane-bound enzyme, α-glucosidase I, to produce $Glc_2Man_9GlcNAc_2$ (Parodi, 2000). The rapid removal of the terminal α1,2 linked glucose residue from the $Glc_3Man_9GlcNAc_2$ oligosaccharide serves two purposes;

(i) The $Glc_2Man_9GlcNAc_2$-Asn structure is not a substrate for the hydrolase activity of OST complex preventing the removal of the oligosaccharide from the polypeptide.
(ii) It exposes the next α1,3 glucosyl residue for the action of glucosidase II.

Another ER-soluble enzyme, glucosidase II, rapidly removes the next glucose unit, leaving a monoglucosylated oligosaccharide, $GlcMan_9GlcNAc_2$-Asn (Parodi, 2000). The monoglucosylated derivative then allows the interaction of the glycoprotein with a series of lectin chaperones, calnexin and calreticulin, which are involved in monitoring and assisting the correct folding of the protein molecule (Parodi, 2000; Ellgaard and Helenius, 2001).

The ER membrane-bound calnexin and soluble calreticulin recognise both the $GlcMan_9GlcNAc_2$ structure and hydrophobic amino acid residues of the polypeptide backbone and facilitate the correct folding of the molecule. Another chaperone protein, ERp57, also interacts with calnexin during this process and is involved in the formation of disulphide bonds (Ellgaard and Helenius, 2001). On the removal of the α1,3 linked glucose residue from the protein-linked oligosaccharide by the enzyme glucosidase II, the glycoprotein is released from the calnexin or calreticulin complex. At this point if the protein has achieved its correct conformation it is exported to the Golgi complex for further processing. However, if the protein remains misfolded the enzyme UDP-glucose: glycoprotein glucosyltransferase re-glucosylates the $GlcNAc_2Man_9$ structure to produce a $GlcMan_9GlcNAc_2$ oligosaccharide (Parodi,

2000). The reglucosylated glycoprotein then undergoes another cycle of interaction with calnexin / calreticulin. The UDP-glucose: glycoprotein glucosyltransferases act as a folding sensor, detecting the misfolded glycoproteins and re-glucosylates them for interaction with calnexin and calreticulin. Figure 1.18 depicts the folding pathway of a typical diglycosylated protein and its interactions with the folding chaperones and various α-glucosidases.

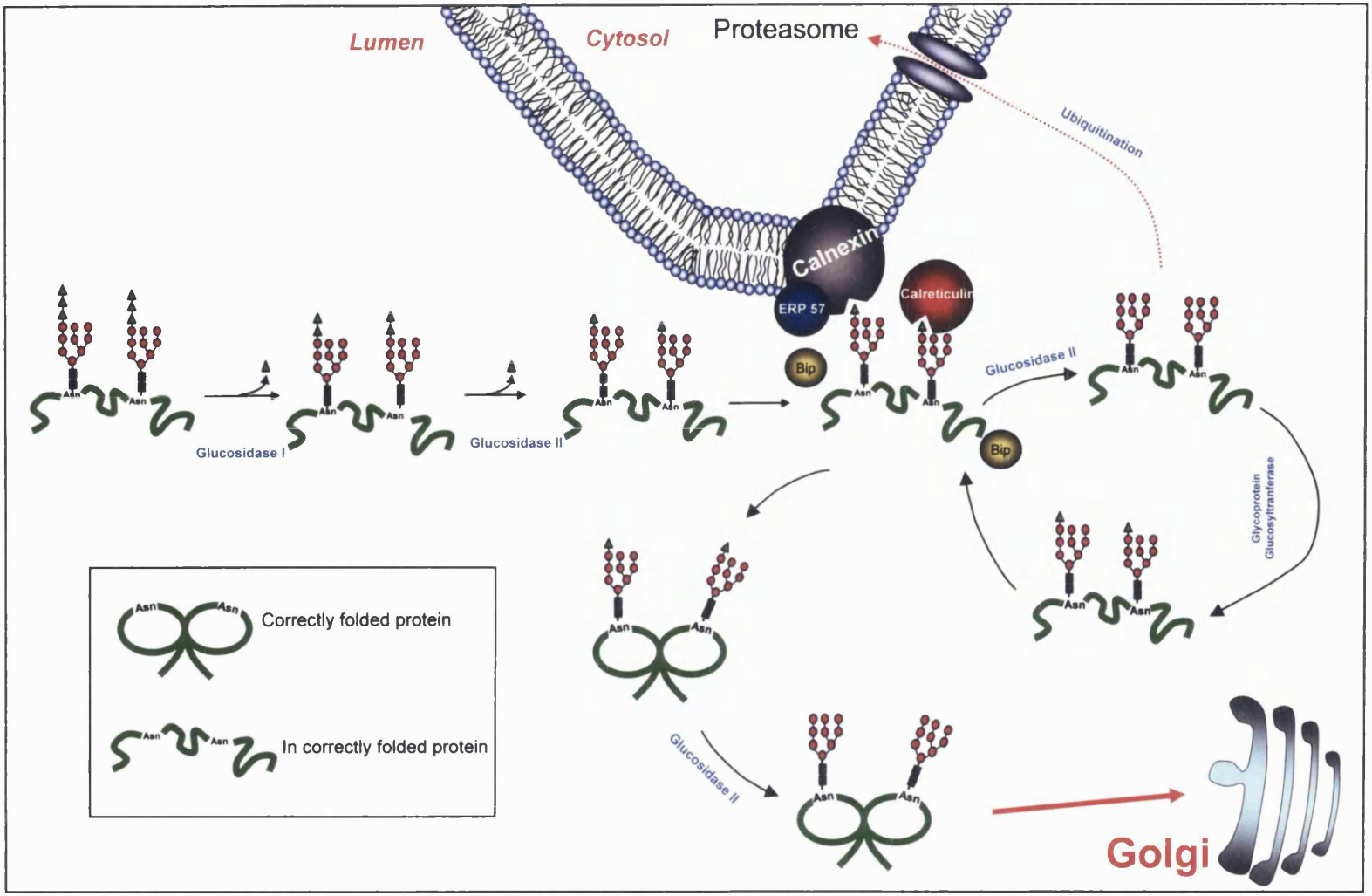
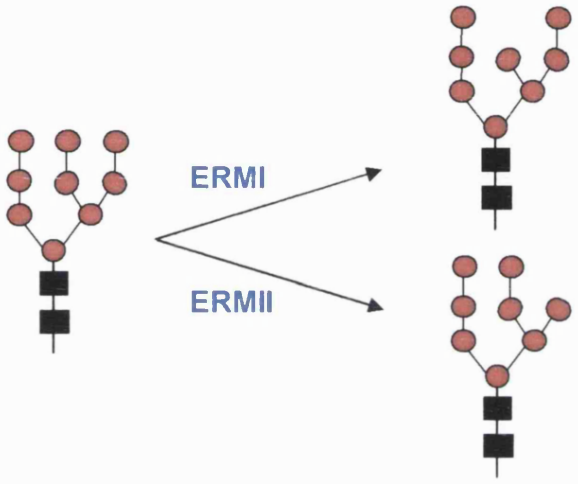


Figure 1.18. The interaction between glycoproteins and chaperones during folding in the ER

Once the glycoprotein has achieved its correct conformation, it is transported from the ER to the Golgi in transport vehicles (Lord *et al.*, 2000). However, the glycoprotein can undergo further processing in the ER before its transport to the Golgi complex, by the α-mannosidases ERMI or ERMII, to produce a $Man_8GlcNAc_2$ glycan, as depicted in Figure 1.19 (Moremen *et al.*, 1994).



ERM1- endoplasmic reticulum mannosidase I

ERMII- endoplasmic reticulum mannosidase II

Figure 1.19. Processing of N-linked glycans in the ER by α-mannosidases.

1.5.2.5 Proteasomal Degradation of incorrectly folded proteins.

If the glycoprotein molecule does not achieve its correct conformation after repeated cycles of deglucosylation and reglycosylation and binding and release from the chaperones, calnexin and calreticulin, it is exported to the cytosol and degraded in the proteasome. The misfolded molecule is translocated back into the cytoplasm through the translocon channel (sec 61) (Vashist *et al.*, 2001). The exact mechanism and the ER-resident components involved in this reverse translocation of misfolded proteins into the cytoplasm are not understood fully. In the cytosol a 76-amino acid polypeptide called ubiquitin is added to the misfolded protein. The ubiquitin is first activated in an ATP-driven reaction by the enzyme ubiquitin activating enzyme (E1). The activated ubiquitin is then attached to the protein targeted for degradation by the enzyme, ubiquitin-conjugating enzyme (E2). The ubiquitin is attached to the target protein by an iso-peptide bond between the COOH-terminal glycine residue of ubiquitin and the ε-amino group of a lysine residue on the protein. The ubiquitinated target protein can become poly-ubiquitinated by the addition of other ubiquitin molecules to lysine 48 of each ubiquitin attached to the target molecule catalysed by ubiquitin ligase (E3). In this way long chains of ubiquitin can be attached to the protein targeted for degradation. The transfer of the misfolded protein into the cytosol is necessary because the E1, E2 and E3 enzymes are positioned adjacent to the translocon channel and situated on the cytosolic surface of the ER. The ubiquitinated protein molecule is then transported to and degraded in the cytosolic proteasomes (de Virgilio *et al.*, 1998). This is the main pathway for the degradation of misfolded proteins in the ER but an alternative pathway has been described by Vashit and colleagues (2001).

1.5.2.6 Transfer from the endoplasmic reticulum to the Golgi apparatus

After transfer to the Golgi apparatus the glycoprotein undergoes further processing of the N-linked glycans before export to other cellular compartments or out of the cell. The transport of glycoproteins from the ER to the Golgi apparatus involves several distinct vesicular steps (Gorelick and Shugrue, 2001). Firstly, buds containing the glycoproteins and coated with Cop II proteins form on the ER. After budding off they travel to the vesiculo-tubular cluster (VTC). The COP II coat proteins mediate budding from the ER while the COP I proteins are required for forward movement from the VTC to the Golgi apparatus. The glycoproteins then pass from the VTC to the Golgi in COP I-coated vesicles or through the formation of tubular connections. After the transport of the glycoproteins from the ER to the Golgi is complete, resident ER proteins are returned to the ER via a recycling mechanism in COP II-coated vesicles (See Figure 1.20; Gorelick and Shugrue, 2001).

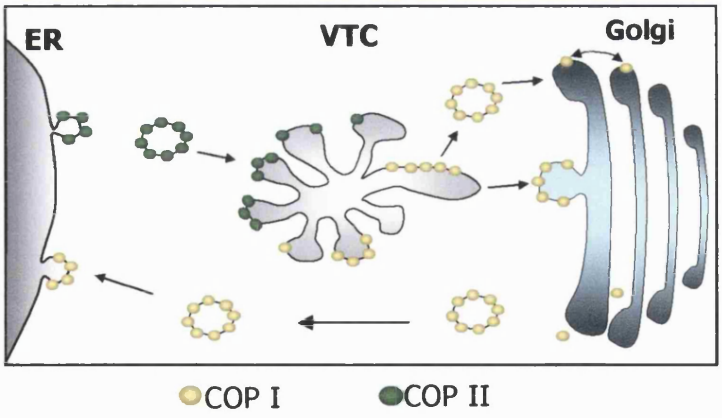


Figure 1.20 Transport of glycoproteins from the ER to the Golgi complex.

1.5.2.7 Processing of the protein N-linked glycans in the Golgi apparatus

The biosynthesis of N-linked glycans in the ER is highly conserved throughout eukaryotes from yeast to mammals (Helenius and Aebi, 2001). However, the N-glycan processing pathways in the TGN show significant divergence among different taxa in eukaryotes (Varki *et al.*, 1999a; Varki *et al.*, 1999b; Moremen 2002, Herscovics, 1999; Munro, 2001). Once in the Golgi complex, the glycoproteins undergo extensive processing of the N-linked glycans involving many specific enzymes. The Golgi apparatus is divided into three compartments, *cis-*, *medial* and *trans* Golgi, with each respective compartment containing a series of processing glycosidases and glycosyltransferases that are used to produce the final glycan structures present on the molecule. All mammalian N-glycans share a common core structure Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc-Asn (Figure 1.21). There is, however, an enormous variety and complexity in the oligosaccharides attached to this core structure. The major classes of N-linked glycan structures in the finally processed glycoproteins are shown in Figure 1.21

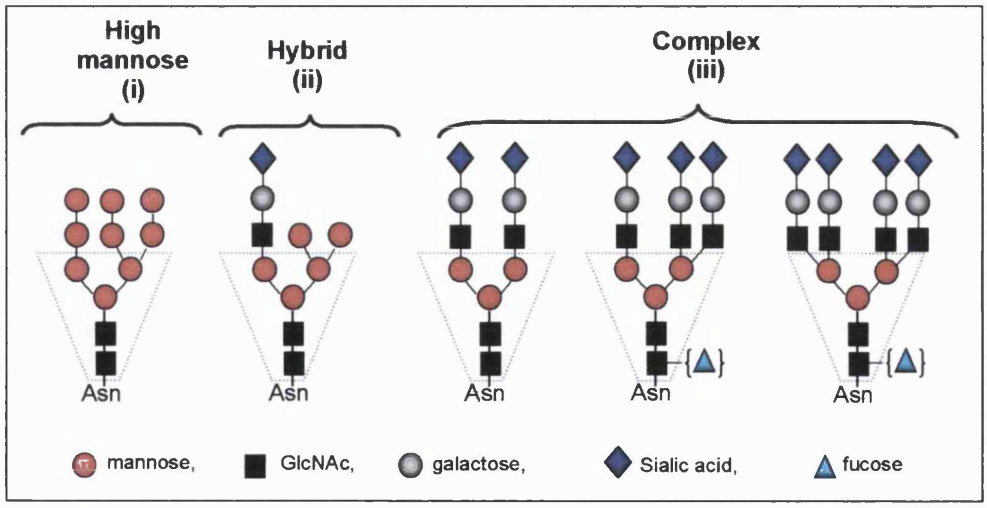


Figure 1.21. The three classes of glycan structures found on N-linked glycoproteins (the common pentasaccharide core is high-lighted in blue).

i) High mannose N-glycans contain only D-mannose (Man) residues attached to the pentasaccharide core (Schachter, 2001a).

ii) Hybrid N-glycans possess only Man residues on the Manα1-6 arm of the core and one or two antennae containing GlcNAc, Gal and sialic acid residues on the Manα1-3 arm (Schachter, 2001a).

iii) Complex N-glycans can have single or multi antennae attached to both Manα1-3 and Manα1-6 arms. These are brought about by the action of the GlcNAc-transferases (Schachter, 2000) and are further extended by the addition of D-galactose, N-acetyl-D-glucosamine, L-fucose, sialic acid and sulphate. The number of antennae results from the competition or cooperation between different GlcNAc-transferases (Figure 1.22; Adapted from Varki *et al.*, 1999a). All N-glycans except the high mannose type may be 'bisected' by a GlcNAc residue attached in β1-4 linkage to the β-linked Man of the core due to the action of GlcNAc-transferase III (Figure 1.22; Adapted from Varki *et al.*, 1999a). The Asn-linked GlcNAc of the core of all N-glycans except high mannose type may have a α1-6 linked fucose (Schachter , 2001a).

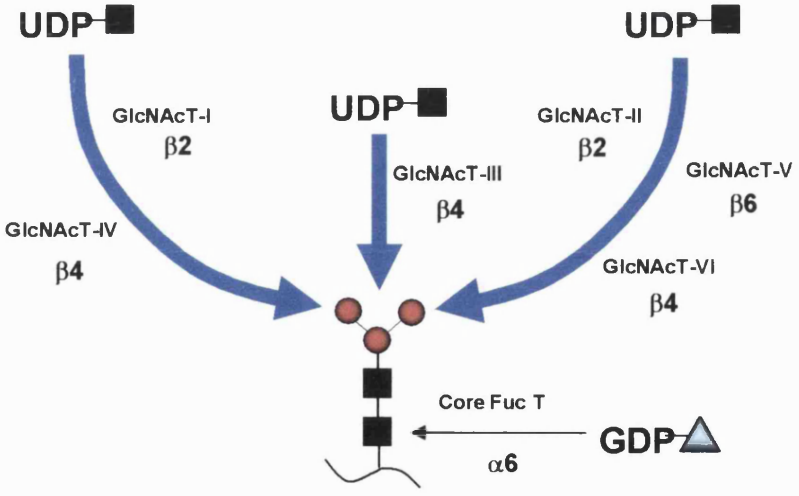


Figure 1.22 Diagram showing the possible actions of the GlcNAc-transferases on the common pentasaccharide core in the Golgi complex.

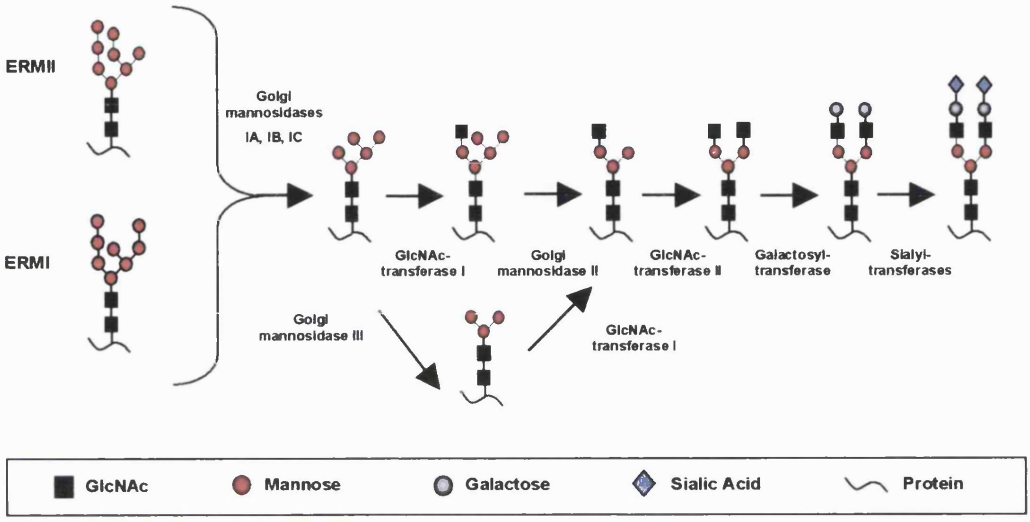


Figure 1.23 Processing of N-linked glycans in Golgi apparatus.

All the isoforms undergo further demannosylation reactions by Golgi mannosidase 1A, 1B and 1C to convert them into $Man_5GlcNAc_2$ structures (Tremblay and Herscovics, 2000) (Figure 1.23). The key enzyme for the conversion of high-mannose to complex and hybrid N-glycans is GlcNAc transferase I (T-I) which adds a GlcNAc in a β1-2 linkage to the Manα1-3Manβ1-4GlcNAcβ arm of the core (Figure 1.23). The presence of a β2-linked GlcNAc residue at the non-reducing terminus of this arm is essential for the subsequent actions of several enzymes in the processing pathway, i.e. α3/6-mannosidase II, GlcNAc T-II, core α6-fucosyltransferase and GlcNAc T-III and IV. GlcNAc-TI is therefore a 'go' signal for all these enzymes. Similarly, GlcNAc T-V needs the prior action of GlcNAc T-II (Schachter, 1991). However, an alternative pathway does exist where the $Man_5GlcNAc_2$ core structures can first undergo demannosylation steps by Golgi

mannosidase III generating $Man_3GlcNAc_2$ core structures which then become substrates for GlcNAc T-I to produce $GlcNAcMan_3GlcNAc_2$ structures (Chui *et al.*, 1997).

There are many instances in the N-glycan processing pathways at which more than one enzyme competes for a common substrate. The route taken by the synthetic pathways at a competition point is dictated primarily by the relative activities of the competing transferases. Some glycosyl residues serve as a 'stop' signal in the synthetic pathway, e.g., insertion of a bisecting GlcNAc prevents the actions of α3/6-mannosidase II, core α6-fucosyltransferase, and of GlcNAc T-II, IV and V (Schachter *et al.*, 1986; Brockhausen *et al.*, 1988). Although this reaction halts branching in the *medial* Golgi cisternnae, it does not prevent passage to the *trans* Golgi and subsequent addition of Gal or GalNAc, sialic acid or sulphate or L-fucose to the antennae.

In forming multi-antennary N-glycans, GlcNAc residues are added to the trimannosyl core by six different GlcNAc-transferases (I-VI) (Schachter, 2000) (Figure 1.22). GlcNAc T-I and -II must act to produce a complex N-glycan. All these enzymes have been cloned (Schachter, 2000). Core fucosylation is a common structural feature among mammalian N-glycans. Fucosyl transferases add fucose in an α1-6 linkage to the GlcNAc residue that is linked to asparagine. This modification is found on hybrid and complex N-glycans following GlcNAc T-I action and is inhibited by the bisecting GlcNAc added by GlcNAc T-III (Staudacher *et al.*, 1999).

The addition of β1-4 Gal onto GlcNAc residues in the hybrid and complex N-glycan is carried out by the β1-4 galactosyltransferases. A large family of human enzymes designated as β4Gal T1-T7 have been cloned, where individual enzymes display tissue specific and expression differences (Hennet, 2002; Amado *et al.*, 1999).

The sialic acids are typically found on the outermost ends of N-glycans. They are subject to a wide variety of modification. Further diversity in sialic acid presentation is generated by different α-linkages from the 2-carbon to underlying sugar chains. Of these the most common are to the 3- or 6-position of Gal residues or to the 6-position of GlcNAc residues. The structural diversity of sialic acids can determine and / or modify the recognition by antibodies. The various linkages and modifications of sialic acids show tissue-specific and developmentally regulated expression. The transfer of sialic acids from CMP-sialic acid to newly synthesised glycoconjugates in the Golgi is catalysed by a family of linkage specific sialyltransferases, many of which have been cloned and characterised (Harduin-Lepers *et al.*, 2001).

It is generally assumed that newly made glycoproteins move through the compartments of the Golgi from *cis* to *trans*, and hence the order in which they are exposed to the Golgi enzymes will be dictated by the distribution of the enzymes between the cisternae.

1.6 N-linked glycoproteins analysed in this study.

1.6.1 Choice of N-linked glycoproteins used in this study.

Two commercially purified human serum N-glycoproteins namely transferrin and α_1-antitrypsin as well as their native counterparts in the serum and plasma samples of normal controls and patients were studied in this project. Both proteins are clinically important and have been extensively studied and their normal structures are well defined. The third N-glycoprotein used in the present study was industrially produced human placental β-glucosylceramidase (Ceredase) and its recombinant form known as Cerezyme, which are used in the enzyme replacement therapy of Gaucher disease. The following section describes in detail, the structure, function and physiology of these glycoproteins.

1.6.2 Human transferrin.

1.6.2.1 Structure, function and physiology.

Human serum transferrin belongs to the transferrin superfamily which also contains ovotransferrin which is found in egg white and lactoferrin, which is found in milk and white blood cells. The transferrins specifically bind free Fe^{3+} ions at physiological pH in serum and transport Fe^{3+} ions into cells via transferrin-receptor mediated uptake (Baker and Lindley, 1992).

Human serum transferrin is a 79 kDa glycoprotein which is synthesised primarily in the liver and contains two N-glycosylation sites (Figure 1.24). The mature protein has a polypeptide chain of 679 amino acids but a pre-protein containing a 19 amino acid hydrophobic signal sequence is found in the liver. The transferrin molecule contains 19 disulphide bonds and has two distinct lobes which are termed the N- and C-lobes, inter-connected by a small sequence of amino acids. Each lobe has been further subdivided into two domains termed N1 and N2, and C1 and C2 respectively (Baker and Lindley, 1992). Since, both lobes have very similar sequence homology

and conformations, it is probable that transferrin originated by genetic sequence repetition (Williams, 1982). Similarly to α_1-antitrypsin, transferrin is also an extremely polymorphic protein with an excess of 30 distinct genetic variants having been described (Welch and Longmead, 1990).

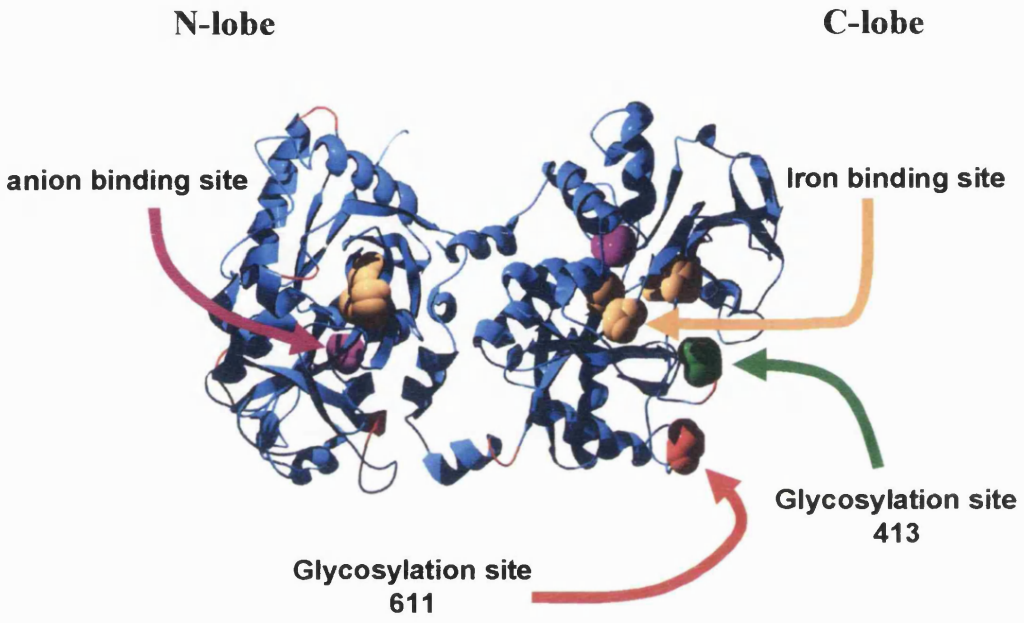


Figure 1.24 Ribbon view of the human serum transferrin showing Fe^{3+} and CO_3^{2-} (anion) and N-linked glycosylation sites.

Human transferrin contains two N-linked glycans at asparagines 413 and 611 (Figure 1.24), although in other mammals the number of glycans can vary from one to four, as in rabbit and bovine transferrins, respectively (Baker and Lindley, 1992). In all species the glycans are located on the C-lobe of the molecule. The glycans found on human serum transferrin are predominantly complex biantennary ones (82±3%), although smaller amounts contain one bi- and one triantennary (17±2%) and two triantennary complex glycans (>1%). Although transferrin is not an acute phase protein, the microheterogeneity of the glycan structures has been shown to increase during liver disease and pregnancy (Montreuil *et al.*, 1997).

Studies carried out on partially or completely deglycosylated transferrin, or on recombinant transferrin after site-specific mutagenesis of N-glycosylation sites have demonstrated that the glycosylation state of the molecule does not affect its Fe^{3+}-binding capacity or recognition by the plasma membrane-bound transferrin receptor (Montreuil *et al.*, 1997). Van Eijk and colleagues (1987) has also shown that microheterogeneity of the glycans, which varies from zero to five sialic acids residues per glycan, also does not affect its Fe^{3+}-delivery capacity to rat reticulocytes.

The total human body concentration of Fe^{3+} is approximately 3-5 g and is distributed amongst four major compartments; 1.5-3 g in the red blood cells, 0.1-0.3 g bound to proteins such as myoglobin, cytochrome c, sulphur proteins, peroxidases and catalases and approximately 1 - 1.5 g bound to ferritin mainly in the liver. The transferrin-bound Fe^{3+} in the serum and lymph tissues is only 3-4 mg, which in relation to the total Fe^{3+} in the human body, points to the efficiency of transferrin in maintaining the low levels of iron in the circulation (Montreuil *et al.*, 1997).

Iron-saturated transferrin is termed *holo*-transferrin and takes on a pink coloration while iron-free transferrin is termed *apo*-transferrin. Transferrin has the ability to bind two Fe^{3+}-ions together with two CO_3^{2-} anions, the binding of the CO_3^{2-} anions being a prerequisite for the binding of the Fe^{3+} ions. The binding site for each Fe^{3+} -ion is found deep within the inter-domain cleft. The amino acids involved in the binding of the Fe^{3+} ion include two tyrosine, one histidine and one aspartic acid residue; tyrosines 95, 188, histidine 249, aspartic acid 82 in the N lobe and tyrosines 426, 517, histidine 585, aspartic acid 392 in the C-lobes. The CO_3^{2-} -anion fits into the cleft containing the Fe^{3+} ion and is held in place by a series of hydrogen bonds between the anion, an arginine and an adjacent threonine residue (arginines 124 and 426, in the N- and C-lobes, respectively (Baker and Lindley, 1992)).

The binding and release of Fe^{3+} -ions involve large changes in the conformation of the transferrin molecule, with the apo-transferrin having an open lobed conformation while the Fe^{3+} -bound conformation has both lobes closed. At neutral pH, transferrin binds Fe^{3+} and CO_3^{2-} -ions very tightly and also binds to the transferrin receptor on the plasma membrane of cells (Montreuil *et al.*, 1997). The transferrin molecule delivers its cargo Fe^{3+} -ions to the cell by internalisation of the transferrin-receptor

complex into the cell. The molecule then releases the Fe^{3+} and CO_3^{2-} ions owing to a fall in pH to 5.6 in the acidic environment of the endosome (Hirose, 2000). The mechanism by which this process occurs has been termed the dilysine trigger mechanism and is thought to occur by protonation of lysine residues 206 and 296 in the N-lobe at low pH (Hirose, 2000). Thorstensen and Romslo (1990) have shown that cellular uptake of the Fe^{3+} ion from transferrin results in the reduction of the iron molecule to its Fe^{2+} state.

1.6.2.2 Transferrin and human disease.

Hereditary atransferrinaemia, a deficiency of plasma transferrin, is an extremely rare condition and has only been reported in 8 patients (Beutler *et al.*, 2000). Atransferrinaemia is characterised by paleness and extreme fatigue with microcytic anaemia, which results from diminished incorporation of iron into haemoglobin (7-55% of normal) (Goya *et al.*, 1972). Untreated patients have also been described with low iron in the marrow, hepatomegaly, recurrent infections, marked deposits of iron in the liver, pancreas, thyroid, myocardium and the kidneys.

Treatment of atransferrinaemia was initially carried out using plasma transfusions but this was eventually replaced by infusions of purified transferrin to reduce the risk of infectious diseases (Schwick, 1977). Treatment with augmentation therapy every two weeks is well tolerated and patients respond well and return to haemotologically normal levels. Iron stores in ferritin and the liver remain high but excess iron can be removed through phlebotomy (Beutler *et al.*, 2000).

The underglycosylation of transferrin and other serum glycoproteins in CDG patients is described in section 1.5. However, underglycosylation of serum glycoproteins is also evident in patients with long term or chronic alcohol dependence (Henry *et al.*, 1999). Henry and colleagues have shown that transferrin, α_1-antitrypsin, haptoglobin, clusterin and serum-amyloid protein are all underglycosylated in the plasma of alcoholic patients (1999). The underglycosylation of glycoproteins has been used in the detection of long term alcohol abuse or in the monitoring of its treatment. As a point of interest the discovery of CDG-I is based on a serendipitous observation made

by Jaeken and colleagues who noted similar plasma transferrin IEF patterns between alcoholic patients and those patients suffering with neurological and multi-system defects (Jaeken *et al.*, 1984). Transferrin has also been implicated in the reinforcement of the immune system and therefore constitutes an important defence mechanism in the body (Rossen, 1966).

1.6.3 Human α_1-antitrypsin.

1.6.3.1 Background.

Alpha$_1$-antitrypsin was first identified as α1-3,5-glycoprotein by Schultze and colleagues (1955) and later renamed α_1-antitrypsin because the inhibitory activity of serum towards trypsin was associated with this protein fraction. However, it was subsequently found that α_1-antitrypsin was a potent inhibitor of all proteases containing a serine at the active site and in particular neutrophil elastase. The association between the level of plasma α_1-antitrypsin and human disease was first described by Laurell and Eriksson (1963), who found a marked decrease in the plasma α_1-antitrypsin levels in early onset emphysema.

Although α_1-antitrypsin inhibits all serine proteases to varying degrees, its main inhibitory action is against neutrophil elastase in the lower respiratory tract where it prevents the destruction of the airway epithelial cells by neutrophil elastase. This is achieved by maintaining a correct balance between the concentration of inhibitor and protease in the circulation. Out of respect for these original findings and the contribution of Laurell and Eriksson, the name α_1-antitrypsin has been retained.

Alpha$_1$-antitrypsin is synthesized in the liver at a rate of approximately 34 mg per kg per day and has a steady state concentration in the plasma of approximately 1.3 mg/ml, although neutrophils also have the capability to synthesis α_1-antitrypsin in order to control the protease/inhibitor balance in microenvironments during bouts of inflammation and infection (Bironaite *et al.*, 2001). Alpha-$_1$-antitrypsin is an acute phase protein and its plasma level can increase as much as four fold during inflammation and infection. The half life of α_1-antitrypsin is between four and five days, and the removal of the sialic acid residues from the N-linked glycans results in a sharp fall in the half life to approximately 0.8 days (Brantly *et al.*, 1988).

In addition to its role as a protease inhibitor in the circulation, recent research has shown that α_1-antitrypsin may have other functions, including a role in cellular migration (Knoell *et al.*, 1998; Paakko *et al.*, 1996), as a chemoattractant and inhibitor of leukocyte migration (Bironaite *et al.*, 2001) and a binding protein for hydrophobic

metabolites in the liver. Some of these properties are discussed in more detail in Section 1.6.3.3.

1.6.3.2 Structure of α_1-antitrypsin.

Alpha$_1$-antitrypsin is a member of the serine protease inhibitor super-family of which there are over 40 members (Kim and Yu, 1996). All members of the serpin family are composed of three β-sheets and several α-helices and are capable of inhibiting, albeit to various degrees, those proteases which contain serine at their active site. The x-ray crystallographic structure of α_1-antitrypsin demonstrates that it is made up of 3 β-sheets and 8 α-helices (Cox, 2001). The three-dimensional structure of α_1-antitrypsin has been shown not to be the most thermodynamically stable folding intermediate possible. These modelling studies have also shown that α_1-antitrypsin contains a highly stressed amino acid 'loop' at the C-terminal end of the molecule, which contains the active site (Figure 1.25a). This loop has been demonstrated to be critical in maintaining the inhibitory function of the molecule.

The α_1-antitrypsin present in the circulation is made up of 394 amino acids although a larger pre-protein containing a 24 amino acid hydrophobic signal sequence can be detected in the liver and *in vitro*. The protein sequence contains only one cysteine (residue 232) and therefore it does not have disulphide bonds (Figure 1.24a). However, the cysteine residue of α_1-antitrypsin can form a disulphide bond with other proteins such as immunoglobulin α heavy chain, immunoglobulin κ light chain and glutathione (Cox, 2001). It is also known to interact with hydrophobic compounds such as cholesterol and bile acids (Janciauskiene and Eriksson, 1994). The cysteine residue of α_1-antitrypsin has recently been shown to be extremely important in regulating and introducing other functions to the molecule when it becomes nitrosylated (Miyamoto *et al.*, 2000) (see Section 1.6.3.3).

Mature serum α_1-antitrypsin is a globular glycoprotein of approximate mass of 52 kDa with as much as 12% of the molecular weight being contributed by carbohydrate (Cox, 2001). Alpha$_1$-antitrypsin has three potential N-glycosylation sites at

asparagine residues 46, 83 and 247, which are normally occupied by complex glycans giving α_1-antitrypsin a net negative charge (Figure 1.25a). Isoelectric focusing of plasma α_1-antitrypsin results in the detection of 8 bands, which are numbered M1 to M8 (anodal-low pH to cathodal-high pH) (Fagerhol and Laurell, 1967). The M4 and M6 bands are the most abundant and these isoforms make up 40 and 34 percent of the total plasma α_1-antitrypsin, respectively (Cox, 2001) (Figure 1.25b). The multiple forms of α_1-antitrypsin are due primarily to the presence of different numbers of sialic acid residues on the glycans. Isoforms containing more triantennary complex glycans, and thus more sialic acid residues, have lower pI values. Bands M4 and M6 have one tri-antennary plus two bi-antennary glycans and three bi-antennary glycans, respectively (Vaughan *et al.*, 1982). The two minor cathodal isoforms, bands M7 and M8, have the same carbohydrate structure as the major bands, M4 and M6, but lack the five N-terminal amino acids (Glu-Asp-Pro-Gln-Gly) (Jeppsson *et al.*, 1985). The loss of these five amino acids, by post-translational cleavage, causes an additional cathodal shift of the isoforms due to the loss of the negatively charged amino acids, glutamic acid (Glu 1) and aspartic acid (Asp 2). The proposed structure of the α_1-antitrypsin isoforms found in plasma is shown in Figure 1.25b. The genetic variation in the primary structure of α_1-antitrypsin and its role in human disease are discussed as a preface to Chapter 5.

Figure 1.25.

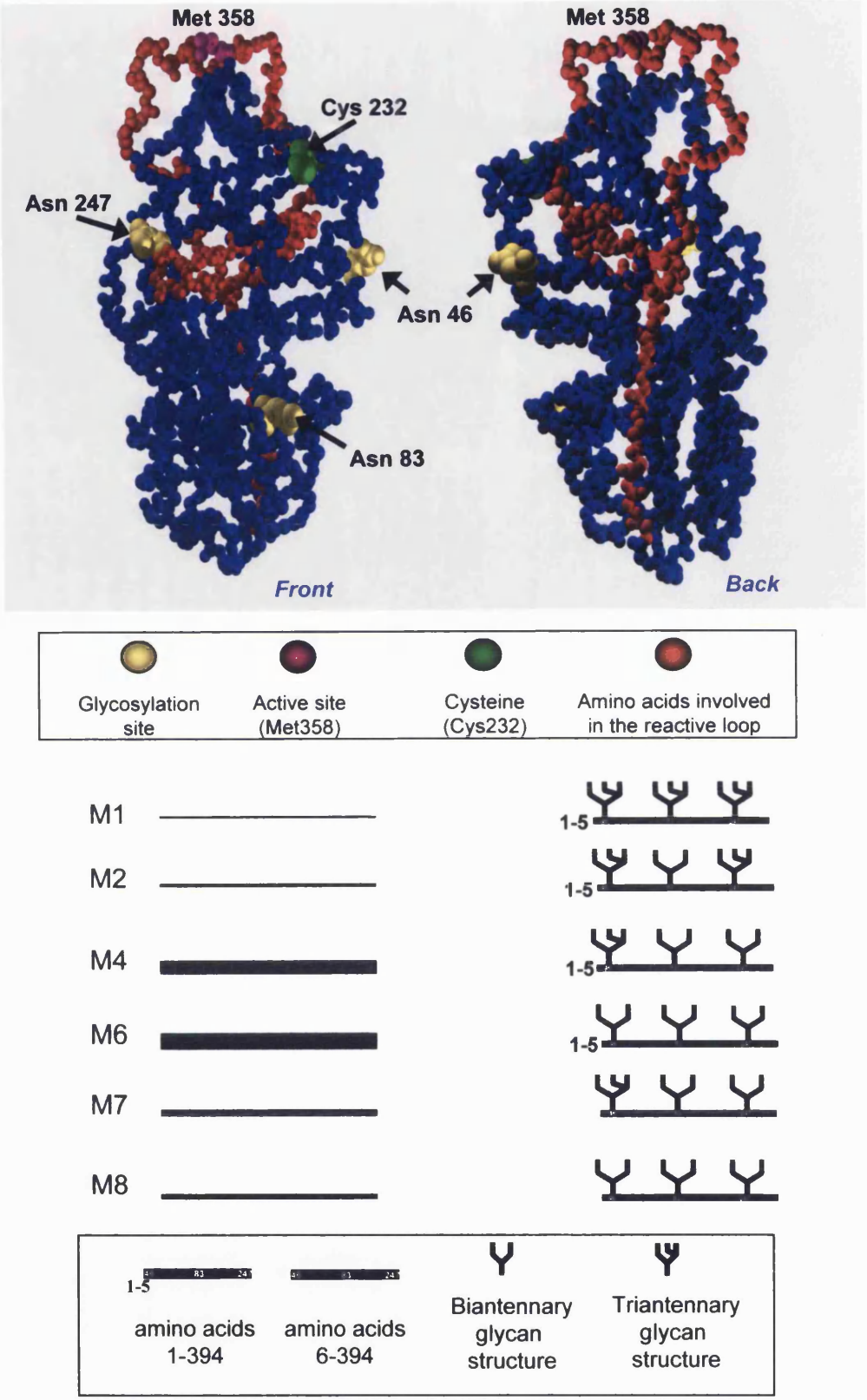


Figure 1.25 **(a) Three-dimensional structure of α_1-antitrypsin showing the potential glycosylation sites, reactive center loop and the free-cysteine residue which can be readily nitrosylated to modulate the function of α_1-antitrypsin.**
(b) Explanation of the heterogeneity of plasma α_1-antitrypsin observed during isoelectric focusing.

1.6.3.3 Function and mechanism of action of α_1-antitrypsin.

The main function of α_1-antitrypsin is generally considered to be its protective role in the lower respiratory tract of protecting lung endothelium from auto-proteolytic damage by neutrophil elastase. Any disturbance in this balance between protease and inhibitor due to a deficiency of α_1-antitrypsin can lead to destruction of the components of the lung extracellular matrix, especially elastin which provides elastic recoil in the connective tissue (Brantly *et al.*, 1988), and an eventual fall in respiratory capacity of the lung.

The active site of the α_1-antitrypsin molecule is centred around the amino acid, methionine 358, and is situated in a highly stressed loop, which extends from leucine 327 to the C-terminus (Brantly, 1988) (Figure 1.25a). Oxidation of methionine 358 destroys the inhibitory capacity towards neutrophil elastase. Methionine 358 is extremely susceptible to oxidation by environmental pollutants and patients with α_1-antitrypsin deficiencies are advised to avoid smoking in order to not accelerate the onset of emphysema (Brantly *et al.*, 1988). The serum levels of oxidised α_1-antitrypsin can be used as a marker of oxidative stress and have been shown to be elevated in conditions such as rheumatoid arthritis and in smokers (Ueda *et al.*, 2002).

The mechanism by which neutrophil elastase is inhibited has been investigated by studying the docking of the α_1-antitrypsin molecule into the active site of the protease (Carrell and Lomas, 2001). Neutrophil elastase is a protease secreted by neutrophils which cleaves peptide bonds at the carboxylic acid side of methionine residues. The protease cleaves the α_1-antitrypsin molecule between methionine 358 (active site) and serine 359 of the highly stressed loop. This results in a large conformational change in α_1-antitrypsin due to cleavage of the highly stressed loop which releases its energy like a coiled spring. Alpha$_1$-antitrypsin swings in a sling-shot motion to the side of the elastase and becomes tightly bound to the side of the protease, rendering it inactive (Carrell and Lomas, 2001). The inhibition reaction results in the suicide of both molecules and it has been calculated that it takes less than 0.2 seconds for one molecule of α_1-antitrypsin to inhibit and inactivate one molecule of neutrophil

elastase (Brantly *et al.*, 1988).

Recently it has been shown that the free cysteine residue at residue 232 in the amino acid sequence can form disulphide bonds with other proteins and molecules such as albumin and glutathione (Miyamoto *et al.*, 2000). It has also been demonstrated that cysteine 232 can react with a nitrosonium ion to form a nitrothiol. Nitrosylated α_1-antitrypsin ([NO]-α_1-antitrypsin) still retains its inhibitory activity against neutrophil elastase but also has other properties (Miyamoto *et al.*, 2000). [NO]-α_1-antitrypsin is a potent vasodilator of rabbit aortic rings *in vitro* in a dose-dependent manner and also has antimicrobial activity against both Gram-negative and Gram-positive bacteria. [NO]-α_1-antitrypsin has a protective affect on the liver after hepatic reperfusion injury. Reperfusion of rat livers with [NO]-α_1-antitrypsin *in vivo* after occlusion of the portal vein, was shown to have many beneficial features including decreased concentrations of liver enzymes in the plasma, increased hepatic blood flow, inhibition of neutrophil infiltration and decreased rates of apoptosis (Ikebe *et al.*, 2000). Reperfusion with non-nitrosylated α_1-antitrypsin did not produce these effects. Although the biochemical basis for these phenomena is unknown, it has been postulated that [NO]-α_1-antitrypsin can act as a reservoir for nitric oxide, allowing delivery of NO to sites of action where it is required (Miyamoto *et al.*, 2000).

1.6.4 Recombinant human β-glucosylceramidase (Cerezyme).

1.6.4.1 Gaucher disease and β-glucosylceramidase.

Gaucher disease is the lysosomal storage disease resulting from a deficiency of β-glucosylceramidase, which catalyses the hydrolysis of glucosylceramide to ceramide and glucose (Figure 1.26). It is characterised by the accumulation of glucosylceramide, a normal intermediate in the catabolism of globoside and gangliosides. It is an autosomal recessive inherited disorder and over 100 different mutations have been found in the gene. *In vivo* a sphingolipid activator protein, saposin A, is required for the hydrolysis of the lipid substrate (Beutler and Grabowski, 2001).

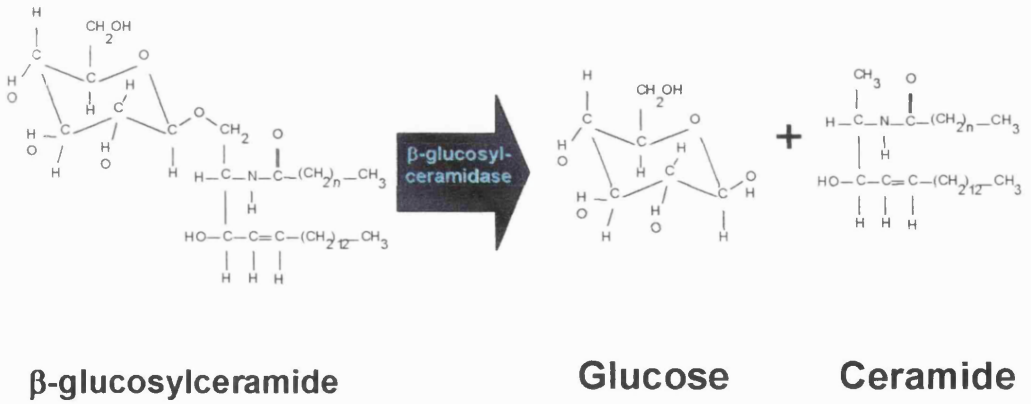


Figure 1.26 Reaction catalysed by β-glucosylceramidase.

Three clinical phenotypes are recognised, type I, the most common, being distinguished from type II and type III diseases by lack of neurological involvement. Type II, the acute neurological disorder, has an early onset culminating in death within the first two years of life. Type III is a sub-acute neuropathic disease, with a

later onset and a more chronic course than Type II. Hepatosplenomegaly, bone lesions and sometimes involvement of the lung and other organs may occur in all forms of Gaucher disease (Beutler and Grabowski, 2001).

1.6.4.2 Human placental and recombinant β-glucosylceramidase.

The amino acid sequence of human placental β-glucosylceramidase was first determined by amino acid sequencing of the purified enzyme (Osiecki-Newman *et al.*, 1986) and the complete sequence was deduced from sequencing of the gene (Tsuji *et al.*, 1986). The protein has about 11% leucine residues and almost 45% non-polar amino acids but transmembrane domains are not present in the mature polypeptide. The precursor form of the enzyme contains 536 amino acids, of which the first 39 amino acids constitute the signal sequence. The mature form of the enzyme (497 amino acids) has a calculated molecular weight of 55,575 Da and the glycosylated enzyme from placenta has a molecular weight of approximately 65,000 (Beutler and Grabowski, 2001). There is some controversy as to whether the enzyme is monomeric or dimeric *in vivo* (Choy *et al.*, 1986; Dawson and Ellory, 1985; Maret *et al.*, 1981; Maret *et al.*, 1983). The enzyme has three disulphide bonds and a free cysteine (126) (Beutler and Grabowski, 2001).

The positions of five N-glycosylation sites (asparagines 19, 59, 146, 270 and 462) are strictly conserved and the sequences are identical in the canine and human enzymes (O'Neill *et al.*, 1989). Typically bi- and triantennary complex N-linked oligosaccharides are present in the human placental form of the enzyme. Endoglycosidase F analysis and site-directed mutagenesis studies have showed that the first four glycosylation sites (asparagines 19, 59, 146 and 270) are utilised, and that the glycosylation of the first site is essential for the formation of an active conformer (Grace and Grabowski, 1990; Grace and Grabowski, 1993). Unlike the majority of lysosomal hydrolases, β-glucosylceramidase is not transported to lysosomes by the mannose-6-phosphate pathway.

1.6.4.3 Enzyme replacement therapy for Gaucher disease.

Early attempts at enzyme augmentation therapy proved disappointing owing to inability of the purified enzyme to be delivered to the sites of action. Subsequent studies showed that most of the enzyme administered was taken up by the asialoglycoprotein receptor in the liver. The work of Brady and colleagues (1994) showed that targeting of proteins to the macrophages could be achieved by utilising the mannose receptors on the surface of the cell. Therefore attempts were made to modify the complex N-linked glycans to expose the mannosyl residues to the receptors. The industrial scale production of enzymatically modified human placental β-glucosylceramidase made sufficient material available for prolonged clinical administration of large amounts of the enzyme to patients. The trials demonstrated a clear-cut response in the Gaucher patients (Barton *et al.*, 1991; Barton *et al.*, 1990). Clinical trials with recombinant β-glucosylceramidase (Cerezyme) indicated efficacy comparable or superior to the placenta-derived enzyme (Grabowski *et al.*, 1993).

Studies on the binding of recombinant β-glucosylceramidase enzyme to murine macrophages and human monocyte–derived macrophages showed that only a very small amount binds to the classical mannose receptor. Most of the mannose-terminated enzyme, is bound by a low-affinity, high copy number, mannose-dependent receptor that is present in many cells including endothelial cells (Sato and Beutler, 1993). It is possible that the small amount of the enzyme taken up by macrophages is sufficient to decrease the glucosylceramide burden or that the enzyme removes the glycolipid from tissues other than macrophages.

The recombinant β-glucosylceramidase or Cerezyme used for Gaucher therapy is produced in a rodent cell line (CHO) and chemically modified to expose mannosyl residues in the N-linked oligosaccharides. It also contains another modification in which the native arginine 495 residue is substituted by a histidine residue (R495H) (Grabowski *et al.*, 1995).

1.7 Aims of this thesis.

The general aim of this thesis was to develop methods for the investigation of post-translational and amino acid modifications in proteins using proteomic technology. Specific aspects of the research were:

(i) The establishment of techniques for identifying proteins separated by 2D-PAGE using mass spectrometry.

(ii) Development of a biochemical strategy for investigating mutations, polymorphisms and post-translational modifications using pure protein standards.

(iii) The application of this strategy to identify unequivocally:
   (a) The individual glycoforms of α_1-antitrypsin observed in control plasma.
   (b) Changes in the amino acid sequence of α_1-antitrypsin in patients with known neutral and charged amino acid substitutions.

(iv) The application of this strategy to investigate whether site-specific glycosylation of two marker proteins, transferrin and α_1-antitrypsin, was an ordered or random process.

Chapter 2

Materials and methods

Chapter 2 – Materials and methods

Contents

2.1 Materials and methods.

2.1.1 Chemical reagents.

All standard reagents were of analar grade or equivalent and obtained from the Sigma-Aldrich Chemical Company (Poole, Dorset, UK.). Carrier ampholytes (pH 4.5-5.5, pH 3-10, and pH 4-7) and 'Immobiline dry strip' isoelectric focusing strips (18cm, pH 4.5-5.5, pH 3-10, and pH 4-7) were obtained from Amersham Pharmacia Biotech (Little Chalfont, Bucks, UK.). Ultra-pure electrophoretic grade acrylamide (30% w/v) was obtained from National Diagnostics Ltd. (Hull, Humberside, UK). Piperazine diacrylamide (PDA) was obtained from BioRad (BioRad Ltd., Hemel Hempstead, UK.). 2,7-Naphthalenedisulfonic acid was obtained from Pfaltz and Bauer Inc. (Frankfurt, Germany). C-18 stationary phase (particle size 60 micron) and microcolumns were obtained from Jones chromatography (Bargoed, Mid-Glamorgan, Wales). 'SCX' ion-exchange cartridges and ZipTips were obtained from Millipore Ltd. (Millipore, Watford, UK.). All proteases used were of sequencing grade and were obtained from Promega Ltd. (Southampton, Hants, UK.). PNGase F, endoglycosidase F3 and endoglycosidase H were obtained from Glyko, Oxford, UK. All solvents used were HPLC grade and were obtained from BDH Ltd. (Poole, Dorset, UK.). α-Cyano-4-hydroxycinnamic acid (α-C4HA), 2,4,6 trihydroxyacetophenone (THAP) and 2,5-dihydroxybenzoic acid (DHB) were obtained from Fluka Fine Chemicals Ltd (Poole, Dorset, UK.).

2.1.2 Patient samples.

Control plasma samples: Age-matched control samples were obtained from the Enzyme Laboratory, Institute of Child Health and Great Ormond Street Hospital NHS Trust, 30 Guilford Street, London. WC1N 1EH. The samples were left over plasma from blood used for diagnostic assays. They were assayed anonymously. All samples used were from patients with unrelated diseases, which did not present with any clinical or biochemical markers of liver dysfunction.

α_1-Antitrypsin variant plasma samples: All samples were obtained with written consent from the MRC Human Biochemical Genetics Unit, Galton Laboratory, University College London. All amino acid substitutions had been confirmed previously using biochemical and genetic analyses.

CDG Type I plasma samples: All samples were obtained form the Enzyme Laboratory, Institute of Child Health and Great Ormond Street Hospital NHS Trust, 30 Guilford Street, London. WC1N 1EH. The initial diagnosis of each patient as CDG type I was made on the basis of a transferrin analysis performed by Dr Geoff Keir at the Institute of Neurology, Queens Square, London. Subsequent confirmation of each enzymatic deficiency was performed using biochemical and genetic analyses.

2.2 Purification of specific proteins from plasma and serum.

2.2.1 Immunopurification of plasma $\alpha_1$antitrypsin

Anti-α_1-antitrypsin affinity chromatography columns were constructed according to the manufacturer's instructions ('HiTrap' N-hydroxy-succinimide (NHS)-activated affinity columns, Amersham Pharmacia, Bucks, UK.). The top cap of the column was removed and 20 µl of ice cold 1 mM HCl was added. Care was taken to avoid any air bubbles at this stage. An adapter for a syringe (supplied with the kit) was attached to the top of the affinity chromatography column and the plastic seal at the other end of the column was removed. The isopropanol contained within the column was removed by washing the column with 6 x 1 ml washes of ice-cold 1 M-HCl. Immediately after the isopropanol was washed from the column, 1 ml of the solution containing the α_1-antitrypsin antibody (Sigma, Poole, Dorset, UK.) was added. The α_1-antitrypsin antibody was reconstituted in an appropriate volume of 0.2 M $NaHCO_3$ buffer, pH 8.3, containing 0.5 M NaCl, to make a final concentration of 7.5 mg/ml. If the resulting solution was cloudy the particulate matter was removed by centrifugation at 10,000 g for 1 min. The column was sealed and allowed to stand for 30 min at 25 ^{0}C.

Any active sites that had not coupled the antibody were deactivated by washing the column with 6 ml of 0.5 M ethanolamine, pH 8.3, containing 0.5 M NaCl, followed by 6 ml of 0.1 M sodium acetate, pH 4.0, containing 0.5 M NaCl. The column was finally washed with 6 ml of 0.5 M ethanolamine, pH 8.3, containing 0.5 M NaCl, and allowed to stand for 30 min. To ensure no remaining unbound $\alpha_1$antitrypsin antibodies were present, the procedure of washing with high and low pH buffers was repeated as described above. The affinity chromatography column was finally equilibrated with 6 ml of 0.05 M sodium phosphate buffer, pH 7.0 and stored at 4 ^{0}C until required.

The column was primed by washing with 3 ml of 0.2 M $NaHCO_3$ buffer, pH 8.3, containing 0.5 M NaCl, followed by 3 ml of 2 M Tris-HCl, pH 2.0 and 3 ml of

0.2 M $NaHCO_3$ pH 8.3, containing 0.5 M NaCl. The column was sealed and allowed to equilibrate for 30 min at 25 ^{0}C before addition of the sample.

Plasma samples were prepared for purification by diluting 50 µl of plasma with 950 µl of 0.05 M sodium phosphate buffer, pH 7.0. The diluted plasma was centrifuged at 10,000 g for 1 min to remove any particulate material and added to the affinity chromatography column. The column was washed with 3 ml of 0.2 M $NaHCO_3$ buffer, pH 8.3, containing 0.5 M NaCl, to remove any unbound material. The α_1-antitrypsin was released from the column by eluting with 2 ml of 2 M Tris-HCl, pH 2.0. The column was immediately re-equilibrated with 6 ml of 0.05 M sodium phosphate buffer, pH 7.0, sealed and stored at 4 ^{0}C. The column was not allowed to dry out at any time and flow rate did not exceed 1 ml per min.

2.2.2 Purification of transferrin using rivanol

Transferrin was selectively precipitated from human serum and plasma using rivanol as described by Mills *et al.*, (2001). Serum or plasma (500 µl) was diluted with 200 µl of H_2O and 300 µl of a rivanol solution (3%, w/v) was added drop-wise with mixing. The solution was allowed to stand at room temperature for 10 min before centrifugation at 10,000 x g for 10 min. The yellow supernatant was removed and added to 430 µl of a NaCl solution (25%, w/v) and mixed thoroughly for 1 min. The resulting suspension was centrifuged at 10,000 x g for 10 min to remove the precipitate formed and 800 µl of saturated ammonium sulphate was added per 1 ml of the clear supernatant. The solution was vortexed for 30 s and allowed to stand on ice for 10 min before centrifugation at 10,000 x g for 10 min. The resulting supernatant containing the transferrin was removed and desalted using an ultrafiltration cartridge as described in section 2.2.3. The yield of transferrin was measured using the modified Biuret protein assay method (Dunn, 1995b) and the purity of the sample ascertained by 1D-PAGE (see 2.3.1). Attempts to purify transferrin from less than 250 µl of serum or plasma resulted in the co-precipitation of albumin

with transferrin in approximately equal amounts. Therefore, samples of less than 500 µl were diluted to 500 µl with 0.05 mM phosphate buffered saline (PBS).

2.2.3 Proteins – Ultrafiltration.

'Amicon' YM-30, ultracentrifugation columns fitted with 30 kDa cut-off membranes were used to desalt transferrin and $\alpha_1$antitrypsin, according to the manufacturer's instructions. Protein solutions (500 µl) were added to the ultrafiltration column and centrifuged in a bench-top Eppendorf centrifuge at 7,000 x g for 6 min. The volume of liquid in the centrifugation cartridge was then checked to ensure that at least 80% of the solution had passed through the filter. If the majority of the solution had not passed through the filter, the microcolumn was re-centrifuged at 7,000 x g for 2 min intervals until approximately 50 µl of the protein solution remained in the upper reservoir. This phenomenon was often observed in solutions containing high salt concentrations. The proteins were de-salted by adding enough H_2O to bring the level of solution in the column to 500 µl. The column was re-centrifuged at 7,000 x g for 6 min and re-centrifuged at 7,000 x g for 2 min intervals until approximately 50 µl of the protein solution remained in the upper reservoir. The protein was recovered from the column by inverting the cartridge and 'pulsing' the centrifuge at 10,000 x g for 20 s. Any residual protein on the membrane was removed by washing the column with 50 µl of H_2O and 'pulsing' the centrifuge for at 10,000 x g for 20 s. The two solutions were combined, dried by centrifugal evaporation or freeze drying and stored at –20 ^{0}C until analysed.

2.3 Polyacrylamide gel electrophoresis (PAGE)

2.3.1 One-dimensional PAGE (1D-PAGE).

2.3.1.1 Resolving gel.

Proteins were separated according to molecular weight devised by Laemmli

(1970) with minor modifications in a vertical slab gel apparatus (Hoefer Scientific Instruments, San Francisco, CA, Table 2.1). Prior to casting the resolving gel, all solutions were de-gassed by sonication for 60 s. N,N,N',N'–tetramethylethylenediamine (TEMED) and ammonium persulphate (APS) were added to the polyacrylamide solution with stirring immediately prior to casting of the gels (see Table 2.1). Two gels (1.5 mm thickness) were cast at a time between 2 silanised glass plates (16 cm by 18 cm) by adding the gel solution to within 3.5 cm of the top of the plate. The gels were immediately over-layered with butan-2-ol, covered with aluminium foil and allowed to polymerise for 2 h.

2.3.1.2 Stacking gels.

A stacking gel (4%, w/v) was used to concentrate the proteins into tight bands prior to entering the resolving gel. Piperazine diacrylamide (PDA) was used as the cross-linker because the gels produced were more tensile and the wells less prone to collapsing on removal of the comb. Prior to casting the stacking gel (see Table 1), the butan-2-ol overlay was removed by inversion of the plates and any residual butan-2-ol removed by washing the surface of the gel with 5 ml of the stacking gel solution. After pouring the stacking gel on to the top of the resolving gel, a 15-well comb (1.5 mm thick) was placed in the stacking gel polyacrylamide solution. Any bubbles under the comb were removed by gently tapping the side of the plates with a metal spatula. Gels were allowed to polymerise at room temperature for 2 h prior to removal of the combs.

Table 2.1 Composition of gels and buffers used in 1D-PAGE.

Solution		Resolving gel (for 100 ml)		Stacking gel (for 50 ml)	Loading buffer (20 ml)
		7.5% *(proteins Mr>60kDa)*	10% *(proteins Mr<60kDa)*	4%	
Acrylamide (30%, w/v)		**25 ml**	**33.3 ml**	**6.67 ml**	-
5 mM sodium thiosulphate		**490 µl**	**490 µl**	-	-
1.34 M Tris-HCl pH 8.8		**25 ml**	**25 ml**	-	-
Crosslinker	PDA Analytical gel	97.5 mg	130 mg	**53 mg (PDA)**	-
	Bis-acrylamide Preparative gel	75.7 mg	100 mg		
SDS (10%, w/v) *Preparative gels only*		**1 ml**	**1 ml**	**0.5 ml**	4 ml
0.5 M Tris-HCl pH 6.8		-	-	**12.5 ml**	2.5 ml
Dithioerythritol (DTE)		-	-	-	0.31 g
Glycerol		-	-	-	2 ml
APS (10%, w/v)		**1 ml**	**1 ml**	**0.5 ml**	-
TEMED		**50 µl**	**50 µl**	**25 µl**	-
		Make up to 100 ml with H_2O	**Make up to 100 ml with H_2O**	**Make up to 50 ml with H_2O**	**Make up to 20 ml with H_2O**

2.3.1.3 Electrophoresis conditions.

Desalted protein samples (see section 2.8.1) were lyophilised in 1.5 ml Eppendorf tubes and reconstituted in 50 µl of the loading buffer (Table 2.1). The wells of the stacking gel were filled with the running buffer. The protein samples were heated at 90^0C for 1 min and allowed to cool to room temperature before addition to the stacking gel well. The gel was then allowed to stand for 20 min prior to electrophoresis, as this minimises the 'smiling' effect often observed in 1D-PAGE.

Electrophoresis was performed at 15 ^{0}C with 30 mAmp per gel, 60 V for 1 h or until the bromophenol blue line entered the resolving gel. After 1 h the voltage was increased to 250 V and electrophoresis was stopped when the bromophenol blue front was observed to leave the resolving gel (approximately 5 h). Gels were either stained immediately or placed in fixative.

2.3.2 Two-dimensional PAGE (2D-PAGE).

2D-PAGE separations were carried out using isoelectric focusing on immobiline strips in the first dimension followed by SDS-PAGE at right angles in the second dimension.

2.3.2.1 First dimension – isoelectric focusing

Isoelectric focusing pH strips (18 cm) were rehydrated overnight in a rehydration cassette in 350 µl of 6 M-urea, 2 M thiourea, CHAPS (2%, w/v), DTE (10 mM), bromophenol blue (0.01%, w/v) and carrier ampholytes (2%, v/v) (BioRad, Hemel Hempstead, UK) containing the sample to be analysed. For analytical gels 2.5 µl of plasma was used and 10 µl for preparative mass mapping analyses. Glycan analyses were performed using 30 µl of plasma. The rehydration solution was spread evenly over the surface of the cassette well and the isoelectric focusing strip was placed directly onto the rehydration solution with the gel side down. The strips were covered in enough mineral oil to completely immerse each strip and allowed to rehydrate at room temperature for 12 h (Sanchez, 1997).

Isoelectric focusing was performed at 20 °C in a LKB-Multiphor II focusing unit (Amersham Pharmacia Biotech, Bucks, UK.) for 75 kV h. The conditions were 0-300 V for 5 min, 300-3500 V over 4 h and 3500 V until 75 kV focusing h were achieved. The focused strips were snap-frozen in liquid nitrogen and stored at – 80 ^{0}C until analysed.

2.3.2.2 Resolubilisation of proteins after isoelectric focusing.

The focused proteins were resolubilised and the cysteine residues blocked by carbamidomethylation in two steps according to Diettrich *et al.*, (1998) prior to SDS-PAGE. The first step involved heating the IEF strips at 90 °C for 1 min in a solution of 50 mM Tris-HCl, pH 6.8 containing glycerol (30%, v/v), SDS (2%, w/v) and DTE (2%, w/v). The IEF strips were allowed to cool to room temperature and equilibrated for 15 min in a solution of 50 mM Tris-HCl pH 6.8, containing 6 M urea, glycerol (30%, v/v), SDS (2%, w/v), iodoacetamide (2.5%, w/v) and a trace of bromophenol blue.

2.2.2.3 Second dimension – resolving gel.

Polyacrylamide gels (2 mm thick) were cast between silanised 22 cm x 20 cm glass plates using the Protean casting chamber (BioRad, Hemel Hempstead, Bucks, UK.). Gels were cast in batches of 8 with each set of plates separated in the gel casting chamber by acetate sheet to stop the plates sticking together on polymerisation of the acrylamide. Approximately 2 litres of de-gassed polyacrylamide solution were added to the casting chamber or until the level of the polyacrylamide reached to within 2 cm of the top of the plates (Table 2.2). The gels were immediately layered with butan-2-ol, covered with aluminium foil and allowed to polymerise for 3 h. Gels were used immediately after polymerisation although they can be stored, wrapped in cling film, at 4 ^{0}C for up to a week after casting.

Table 2.2 Composition of gels for SDS-PAGE in the second dimension.

Solution		Resolving gel *(for 2L)*	
		7.5% *(proteins Mr>60kDa)*	10% *(proteins Mr<60kDa)*
Acrylamide (30% w/v)		**500 ml**	**33.3 ml**
5mM sodium thiosulphate		**9.8 ml**	**490 µl**
1.34M Tris-HCl pH 8.8		**500 ml**	**25 ml**
Cross-linker	PDA (1.6 %) Analytical gel	**1.95 g**	**3.2 g**
	Bis-acrylamide Preparative gel	**1.5 g**	**2.0 g**
SDS (10% w/v) *Preparative gels only*		**20ml**	**20 ml**
APS (10% w/v)		**10 ml**	**10 ml**
TEMED		**1 ml**	**1 ml**
		Make up to 2 L with H_2O	**Make up to 2 L with H_2O**

2.3.2.4 Electrophoresis conditions.

Prior to IEF, strips were cut to size and positioned on the resolving gel by overlaying the surface of the gel with the running buffer and allowing the IEF strip to sink onto the surface of the resolving gel. Any air bubbles observed between the gel surface and the IEF strip were removed by gently tapping the gel plate with a metal spatula. SDS-PAGE was carried out immediately to minimise buffer exchange between the IEF strip and the running buffer.

Electrophoresis was carried out using 40 mA per gel at 60 V for 1 h followed by 40 mA per gel at 250 V until the bromophenol blue was observed to leave the resolving gel according to Laemmli *et al* (1970).

2.3.2.5 Analytical and preparative PAGE.

Although analytical and preparative gels were essentially the same, some minor modifications to the protocol were made to optimise the results. PDA was used as the cross-linker in the analytical gels because it produced more tensile gels, gave higher resolution of the proteins and significantly reduced background staining. It was replaced in preparative gels with bis-acrylamide to improve subsequent in-gel protease / glycanase digestions. The percentage of cross-linker was also decreased in the preparative gels to facilitate in-gel digestions (Table 2.3).

Table 2.3 The % cross-linker required in PAGE for optimal enzymatic digestions.

Protease / Glycanase	% cross-linker
Trypsin	1% bis-acrylamide
Chymotrypsin	0.2% bis-acrylamide
PNGase F	0.2% bis-acrylamide

SDS was omitted from analytical gels to minimise the formation of micelles in the polyacrylamide matrix that can be observed as local distortions during silver staining. SDS was included in the preparative gels as it was observed to facilitate subsequent in-gel protease digestions.

2.4 Detection of proteins after PAGE.

2.4.1 Silver staining.

2.4.1.1 The method of Shevchenko and colleagues (Shevchenko *et al.*, 1996)

PAGE gels were stained manually according to the method of Shevchenko *et al.*, (1996) with minor modifications. Immediately after electrophoresis gels were fixed in a solution of methanol: acetic acid: H_2O (50:10:40 v/v) for 20 min. The gel was washed with 50% methanol (v/v) for 10 min prior to washing with H_2O for a further 10 min to remove any remaining acetic acid. The gel was sensitised by incubating with sodium thiosulphate (0.02%, w/v) for 1 min and washed twice with ice cold H_2O for 1 min to remove the excess sodium thiosulphate. The gel was then incubated in an ice-cold silver nitrate solution (0.1%, w/v) for 30 min. The silver nitrate solution was discarded and the gel was washed twice with ice cold H_2O for 1 min. The gel was then treated with a formaldehyde solution (0.04%, v/v) containing sodium carbonate (2%, w/v) for 1 min. Fresh formaldehyde was added and the gel was agitated vigorously until the proteins started to develop. It was sometimes necessary to include another change of the formaldehyde solution if a persistent brown precipitate was observed. The developing process was terminated when the gel background started to darken, by removing the developing solution and vigorously agitating the gel in acetic acid (5%, v/v). The gels were then decreased in size and volume by the addition of a solution of methanol: acetic acid: H_2O [50:1:49, v/v] for 1 h, before storage in sealed plastic envelopes.

2.4.1.2 The method of Hochstrasser and Merril (1988)

Immediately after electrophoresis, gels were fixed in a solution of methanol: acetic acid: H_2O [50:10:40, v/v]. Gels can be stored in the fixative solution for several months without any effect on the staining or mass spectrometry analyses.

After washing in a solution of methanol: acetic acid: H_2O [5:5:90, v/v] for 30 min followed by 3 washes with deionised H_2O for 10 min, the gels were incubated with a solution of 0.5 M sodium acetate for 30 min. Glutaraldehyde was omitted from the original protocol because it covalently modifies the proteins and affects subsequent mass spectral analysis. Excess sodium acetate was removed by washing the gel 3 times in H_2O. To obtain a homogenous dark brown staining of the proteins, the gels were soaked twice in a solution of 2,7-naphthalene disulphonic acid (0.05%, w/v) for 30 min. All residual 2,7 naphthalene disulphonic acid was removed by rinsing the gel 4 times in H_2O for 10 min. It is important to remove the excess 2,7 naphthalene disulphonic acid to prevent the gel from turning irreversibly opaque on the addition of the silver solution. Gels were stained in a freshly made ice-cold, ammoniacal silver nitrate solution for 30 min. The silver solution is made by the drop-wise addition of 30 ml of a silver nitrate solution [6 g in 30 ml] to 160 ml of an ice-cold, stirring solution of H_2O containing 10 ml of concentrated ammonium hydroxide (35%, v/v) and 1.5 ml of 10 M sodium hydroxide. A transient dark brown precipitate was often observed but it re-dissolved rapidly. The silver nitrate solution was diluted with chilled H_2O to a final volume of 750 ml, covered in aluminium foil and kept on ice until used. The silver solution was chilled to prevent the ammonia evaporating, since significant losses of ammonia led to a general loss in sensitivity of the stain.

Excess silver nitrate solution that has not reacted with the proteins was removed by washing in H_2O for 2 min, 4 times. Failure to wash the gels sufficiently during this procedure could lead to high background staining. Proteins were visualised by the addition of the developing solution containing citric acid (0.01%, w/v) and formaldehyde (0.1%, v/v). Addition of the developing solution was accompanied by the vigorous agitation of the developing tray. Developing times varied from 2-10 min. When a slight background stain was observed, the developing process was stopped by the addition of acetic acid (2%, v/v). The gels were kept in acetic acid (2%, v/v) for 15 min to ensure the developing process was completed. The gels were then shrunk by the addition of methanol: acetic acid: H_2O [50:1:49, v/v] for 1 h, before storage in sealed plastic envelopes.

2.4.2 Reverse staining using copper II chloride (Lee *et al.*, 1987)

Gels were briefly washed in H_2O for 10 s prior to immersion in a solution of 0.3 M $CuCl_2$. The gels were gently rocked in the $CuCl_2$ solution for 10 min and then washed in H_2O for 5 min to remove any excess reagent. Proteins were visualised in the gel as transparent areas whereas areas of the gel that contained no protein were coloured green-blue. Gels were stored in sealed plastic envelopes until analysed.

2.4.3 Reverse staining using zinc sulphate (Ortiz *et al.*, 1992)

Gels were gently washed in a solution of sodium carbonate (1%, w/v) for 5 min. The gels were then incubated in a solution of 0.2 M imidazole containing SDS (0.1%, w/v) for 10 min. The gels were briefly washed with H_2O for 10 s to remove the excess SDS and developed by a rapid incubation with 0.2 M $ZnSO_4$ for 40 s. Development was terminated by four washes with H_2O for 10 s, 20 s, 4 min and 5 min, respectively. The proteins were visualised in the gel as transparent areas whilst areas of the gel that contained no protein were coloured white. Gels were stored in sealed plastic envelopes until analysis.

2.4.4 Coomassie Blue (Meyer and Lamberts, 1965)

Immediately after electrophoresis gels were incubated in methanol: acetic acid: H_2O [50:10:40, v/v] containing coomassie blue stain R250 (0.1%, w/v) for 12 h or overnight. The proteins were visualised by de-staining the gel in a solution of methanol: acetic acid: H_2O (50:10:40, v/v) until the protein bands or spots appeared blue. The de-staining process can involve several changes of the methanol: acetic acid: H_2O (50:10:40, v/v) solution. However, to reduce the number of solution changes, pieces of absorbent paper were placed in the corner of the de-staining tray to absorb the coomassie blue as it left the gel matrix.

2.4.5 Automated procedure.

As the silver staining of gels involves many time-consuming steps it was carried out automatically using 'Pharmacia Biotech gel stainers' according to the method of Hochstrasser and Merril (1988). The automation of the staining process allowed the staining of two gels simultaneously, using the same reagents thereby removing some inter-assay variations that could occur through human error. The delivery volume of each reagent wash was set at 350 ml of each solution and 30 s was deducted from the timing of each step because the pumping out of reagent waste takes approximately 30 s.

2.5 Analysis of proteins in polyacrylamide gels

2.5.1 In-gel derivatisation of cysteine residues.

The gel spots or bands were transferred to 2 ml Eppendorf tubes and washed 3 times with 200 µl of 100 mM ammonium bicarbonate buffer, pH 7.8. Gel slices were dehydrated partially by the addition of 500 µl of acetonitrile and mixed on a rotary mixer for 20 min. The acetonitrile was discarded and the gel pieces were dehydrated completely by centrifugal evaporation for 1 h at 37 °C. Disulphide bridges in the protein were reduced by the addition of 200 µl of 10 mM dithioerythritol in 100 mM ammonium bicarbonate buffer, pH 7.4 and incubated for 1 h at 56 °C. The gel pieces were allowed to cool to room temp and the cysteine residues were carboamidomethylated by the addition of 300 µl of 55 mM iodoacetamide in 100 mM ammonium bicarbonate and incubated at room temp for 45 min in the dark (Schevenko *et al.*, 1996).

This procedure was not necessary for the analysis of proteins from 2D-PA gels because the proteins have already been carboamidomethylated prior to SDS-PAGE.

2.5.2 In-gel enzymic digestion

2.5.2.1 Trypsin

In-gel tryptic digestions were performed as described by Schevenko and colleagues (Schevenko *et al.*, 1996) with minor modifications. Proteins were reduced and carboamidomethylated prior to proteolysis as described in section 2.5.1. Protein spots or bands were cut out from gels using clean 1 ml pipette tips with the orifice cut to the size of the diameter spot. Bands from 1D-PAGE were cut out using a sterilised scalpel. The spots or bands were transferred to 2 ml Eppendorf tubes and washed 3 times with 200 µl of 100 mM ammonium bicarbonate buffer, pH 7.8. Gels were dehydrated partially by the addition of 500 µl of acetonitrile and mixed on a rotary mixer for 20 min. The acetonitrile was discarded and the gel pieces were dehydrated completely by centrifugal evaporation for 1 h at 37 °C. Proteolysis was performed by adding 60 µl of a solution of trypsin (12.5 ng/µl in 50 mM ammonium bicarbonate buffer, pH 7.8) to a dehydrated gel piece, which was incubated for 12 h or overnight in a water bath at 37 ^{0}C.

2.5.2.2 Chymotrypsin.

In-gel chymotryptic digestions were performed in a similar manner to the trypsin in-gel digestion protocol but with minor modifications. The gel slices were incubated overnight with an excess of chymotrypsin (60 µl of 25 µg/µl) and the ammonium bicarbonate buffer was replaced by 100 mM Tris-HCl, pH 7.8.

2.5.2.3 PNGase F.

Glycans were released from glycoproteins by in-gel digestion according to Küster and colleagues (Küster *et al.*, 1997) with minor modifications. Protein spots were cut out from gels using clean 1ml pipette tips with the orifice cut to the size of the diameter spot. Bands from 1D-PAGE gels were cut out and the proteins were reduced and carboamidomethylated as described earlier (see

section **2.5.1**). The dried gel pieces were incubated with 60 µl of PNGase F (100 U/ml) in 20 mM $NaHCO_3$ buffer, pH 7.0 for 16 h to release the glycans.

2.6 Extraction from polyacrylamide gels

2.6.1 Peptides and glycopeptides

Peptides were extracted from the gel-matrix by shaking the gel slices with 300 µl of 50% acetonitrile (50%, v/v) containing TFA (0.1%, v/v) on a rotary mixer for 60 min. The acetonitrile was removed from the gel slice and dried down by centrifugal evaporation. Any peptides remaining in the gel slice were extracted by shaking the gel slice with 300 µl of 4 M urea on a rotary mixer for 30 min. After the second extraction, the urea solution was removed from the gel slice and added to the dried peptides extracted with acetonitrile. This solution was vortexed thoroughly for 60 s to re-solubilise all peptides prior to de-salting (see 2.8). The de-salted peptide solution was dried by centrifugal evaporation and re-constituted in 10 µl of 0.1 % TFA for mass spectrometry.

2.6.2 Glycans

The released glycans were extracted from the gel matrix by three changes of 200 µl of H_2O with sonication for 20 min (Küster *et al.*, 1997). The extracted glycans were de-salted using graphite columns, freeze dried and reconstituted in 10 µl of H_2O prior to analysis by mass spectrometry.

2.7 Digestion of glycoproteins in-solution.

2.7.1 Derivatisation of cysteine residues

All proteins were reduced and carboamidomethylated before proteolysis. 5-100 µg of protein were 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 ^{0}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 ^{0}C for 30 min in the dark to carboaminomethylate the cysteine residues. Excess reagent and low molecular weight products were removed by ultrafiltration (see section 2.2.3) and the protein solution was concentrated by centrifugal evaporation.

2.7.2 Proteolytic digestions

Derivatised proteins were reconstituted in H_2O to a final concentration of 1 μg/μl and 5 μl was added to 190 μl of the appropriate reaction buffer: for trypsin, 50 mM ammonium bicarbonate buffer, pH 7.8; 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) were added to the reaction mixture. After incubation at 37 ^{0}C for 16 h the reactions were stopped by the addition of a cocktail of serine protease inhibitors (Sigma) or by heating at 90 ^{0}C for 30 s. The digestion mixtures were desalted on silanised microcolumns (see 2.8.2.1). 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 PNGase F.

2.7.3 Release of glycans

2.7.3.1 PNGase F

For digestion with PNGase 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 24 h at 37 °C.

2.7.3.2 Endoglycosidase F3

All detergents were omitted from the endoglycosidase reactions because of the adverse affects observed on the subsequent enzymatic 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 24 h at 37 °C.

2.7.3.3 Endoglycosidase H

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 250 mM sodium phosphate buffer, 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 24 h at 37 °C.

2.8 Desalting and removal of contaminants from proteins and digestion mixtures

2.8.1 Peptides

2.8.2.1 Using C-18 stationary phase columns

Digestion mixtures were desalted using silanised microcolumns, containing 5 mg of C-18 stationary phase (Jones Chromatography, U.K.). The columns were silanised by the addition of 100 μl of hexane containing Repelcote V8™ (BDH Ltd., Poole, Dorset, UK.) silanising agent (1%, v/v). The column was primed by the stepwise addition of 1 ml of 90% acetonitrile (90%, v/v) containing 0.1% TFA (v/v), 1 ml of 10% acetonitrile (10%, v/v) 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 2 volumes of 50 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 PNGase F.

2.8.2.2 Using 'Zip-Tip' C-18 microcolumns

Digestion mixtures were desalted on 'Zip-Tip' microcolumns according to manufacturer's instructions (Millipore). The columns were primed by the stepwise aspiration of 5 x 20 µl of 50% acetonitrile containing 0.1% TFA and 5 x 20 µl of 0.1% aqueous TFA. The digestion mixture was applied to the column by repeated aspiration of the solution and washed by aspiration with 5 x 20 µl of 0.1% aqueous TFA. The desalted peptides / glycopeptides were eluted from the column with 2 washes of 10 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/ PNGase F.

2.8.2.3 Ion-exchange using 'SCX' columns

Digestion mixtures were desalted using micro-centrifuge columns fitted with ion-exchange resins and used according to the manufacturer's instructions. Columns were positioned within a 2 ml Eppendorf vial in a bench-top micro-centrifuge with the 'marked' area of the tube facing the inside of the centrifuge. The column was primed by adding 500 µl of methanol : ammonium hydroxide (1:1, v/v) to the column and centrifuging at 7000 x g for 45 s. The priming procedure was completed by adding 500 µl of 0.05% formic acid and centrifuged at 7000 x g for a further 45 s. The sample was added in 500 µl of 0.05% formic acid and centrifuged at 1200 x g for 60 s. The peptides were desalted by

washing the column with 500 µl of 0.05% formic acid and centrifuged at 1200 x g for 45 s. After all washing and priming steps the eluants were discarded.
The microcolumn was transferred to a new vial and the desalted peptides eluted by centrifuging twice with 75 µl of methanol : ammonium hydroxide (1:1, v/v) at 7000 x g for 30 s. The eluate was dried under nitrogen and the desalted peptides stored at –20 ^{0}C until analysed.

2.8.2.3 Glycans

The released glycans were desalted using graphite Glyco H columns (Glyko, Oxford, UK.). Columns were primed with 3ml of 50% acetonitrile containing 0.1% TFA and 3 ml of 5% acetonitrile containing 0.1% TFA. Samples were applied to the column and washed with 3 ml of H_2O and 3 ml of 5% acetonitrile containing 0.1% TFA. Glycans were released from the column with 50% acetonitrile containing 0.1% TFA. Samples were freeze-dried prior to analysis.

2.9 Preparation of matrices for matrix-assisted laser desorption time of flight mass spectrometry (MALDI TOF MS).

2.9.1 α cyano-4-hydroxy acid (αC4HA) matrix.

Markedly increased sensitivity and lower background signals were achieved using recrystallised α cyano-4-hydroxy acid (αC4HA) as the matrix. Commercial αC4HA was purified by recrystallisation from an ammoniacal solution by the stepwise addition of concentrated hydrochloric acid. 100 mg of αC4HA was added to 100 ml of H_2O on a stirrer and ammonium hydroxide (35%, v/v) was added until the αC4HA dissolved completely. Concentrated HCl was added drop wise to the stirring ammonical solution of αC4HA until a brilliant yellow precipitate was observed. The solution was allowed to cool and

the precipitate purified by filtering the solution through a volumetric flask fitted with a glass filter. The brilliant yellow precipitate of αC4HA was washed twice with 100 ml of 0.1 M-HCl and transferred to a round-bottomed flask. The αC4HA precipitate was freeze dried for 16 h using a 'Modulyo' freeze drier (Edwards Ltd, Surrey, UK.).

The matrix preparation was made by dissolving 10 mg of αC4HA in 1ml of acetonitrile / ethanol (50:50, v/v). 1.5 µl of a desalted peptide digestion mixture which had been 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. The target was allowed to air-dry at room temperature for 30 min before being placed into the mass spectrometer.

2.9.2 αC4HA / fucose matrix.

10 mg of αC4HA was dissolved in 1 ml of acetonitrile: ethanol (50:50, v/v) and mixed with an equal volume of aqueous 50 mM fucose. 1.5 µl of a desalted digestion mixture which has been 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 air-dry at room temperature for 30 min before being placed into the mass spectrometer.

2.9.3 Dihydroxybenzoic acid (DHB) 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 homogenous distribution of crystals over the target surface (Küster *et al*, 1998).

2.9.4 Analysis of peptides and glycans using trihydroxyacetophenone.

Trihydroxyacetophenone (2 mg) was dissolved in 1 ml of acetonitrile and mixed with an equal volume of 20 mM ammonium citrate. For peptide analyses, 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. 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.

For glycan analyses 0.75 µl of a desalted glycan mixture, reconstituted in H_2O, was added to the target and 0.75 µl of matrix solution added directly on to the glycan mixture. The droplet formed was collapsed by gently tapping the target which was then dried immediately under vacuum to form a homogenous crystalline film (Papac *et al.*, 1996). The target was allowed to stand at room temperature for 30 min before being placed into the mass spectrometer.

2.10 Matrix-assisted laser desorption ionisation time of flight mass spectrometry (MALDI TOF MS).

2.10.1 Peptide analysis

Mass spectrometry was carried out on a MALDI TOF MS instrument, fitted with a reflectron and a 337 nm u.v. laser (TOF Spec E, MicroMass, Manchester, U.K.). Peptide analyses were performed in positive ion mode with the following voltages, source 20 kV, extraction 19.95 kV, focus 16.5 kV, reflectron 25 kV and a pulse voltage of 2900 V. Spectra were acquired by averaging over a period of 5 scans of highest signal. Data were acquired in reflectron mode, operating over a mass range of 6000 m/z with matrix suppression set at 650 Da.

2.10.2 Glycan analysis

Glycan analyses were performed in both positive and negative ion mode with voltages of 20 kV for the source, 19.95 kV for the extraction, 16.5 kV for the focus and a pulse voltage of 900 V. Spectra were acquired by averaging over a period of 10 scans of highest signal. Data were acquired in linear mode, operating over a mass range of 5000 m/z with matrix suppression set at 800 Da. Glycan structures were observed as the pseudomolecular ions $[M-H]^-$ in negative ion mode and $[M+Na]^+$ in positive ion mode.

2.11 Mass spectrometry data analyses.

Data analysis was carried out using MassLynx and Biolynx data analysis software. Database searching and mass mapping studies were performed using Protein Prospector database software ('MS-FIT') at University of San Francisco (http://falcon.ludwig.ucl.ac.uk/mshome3.2.htm). Peptide and glycopeptide mass calculations were performed using PAWS proteomic analysis software (http://prowl.rockefeller.edu/).

Chapter 3

Optimisation of the 2D-PAGE and mass spectrometry protocols

Chapter 3 -Optimisation of the parameters involved in the analysis of proteins using PAGE and MALDI TOF MS.

Contents

3.0 Optimisation of the 2D-PAGE / mass spectrometry protocols.

The aim of this project was to develop sensitive techniques to separate and characterise aberrant proteins from biological samples such as plasma, using proteomic technology. These techniques could then be used to study changes in protein expression, amino acid substitutions and post-translational modifications in patients with genetic metabolic diseases such as the α_1-antitrypsin deficiencies and the congenital disorders of glycosylation (CDG).

The combination of the separation of proteins by PAGE followed by mass spectral analysis has become an extremely powerful tool for the rapid identification of proteins. Proteins are usually identified after PAGE by the mass spectral analysis of the peptides generated from in-gel proteolytic digestion of separated proteins. The masses of these proteolytic peptides are used to identify the protein, by comparing the experimental masses with theoretical digestions of protein sequences that are available in protein databases such as Expasy (www.expasy.ch). This process is termed as mass mapping identification and can be utilised for high throughput identification of proteins after 2D-PAGE when used in conjunction with MALDI TOF MS. Additionally, the identity of a protein can be determined using partial sequence information obtained by mass spectral peptide fragmentation studies of peptides. The peptide sequence information obtained can be used to perform expressed sequence tag (EST) searches of the same protein databases to obtain a percentage match or score of the identity of the protein. Although MALDI TOF MS instruments do have sequence capabilities, such sequencing studies can be readily and efficiently performed using ion-trap or QTOF mass spectrometry techniques. However, MALDI TOF mass spectrometers are usually used for mass mapping studies, where it is the greater number of peptides detected and hence the higher sequence coverage, that increase the chance of a successful identification of the protein during database searches. In contrast to mass mapping studies, sequencing studies using QTOF or ion trap mass spectrometers require significantly more material than MALDI TOF MS analyses. In both cases the

probability of gaining a positive identification of a protein from a gel depends on both the quality and quantity of peptides generated from the in-gel proteolytic digestion.

The common protocol for the analysis of proteins using PAGE and mass spectrometry can be broken down into 7 critical steps as listed below and are shown schematically in Figure 3.01

.

1 Selection of the optimal pH-range for isoelectric focusing.
2 The detection of proteins after electrophoresis.
3 The in-gel proteolytic digestion of proteins in PAGE gels.
4 The extraction of the peptides from the gel matrix after digestion.
5 The removal of contaminants prior to MALDI TOF MS.
6 The handling of small amounts of peptides to minimize sample losses.
7 Optimum parameters for the mass spectral analysis of peptides u sing MALDI TOF MS.

All these steps have been evaluated and optimised for the development of an effective protocol for the analysis of proteins using PAGE and MALDI TOF MS.

Figure 3.01

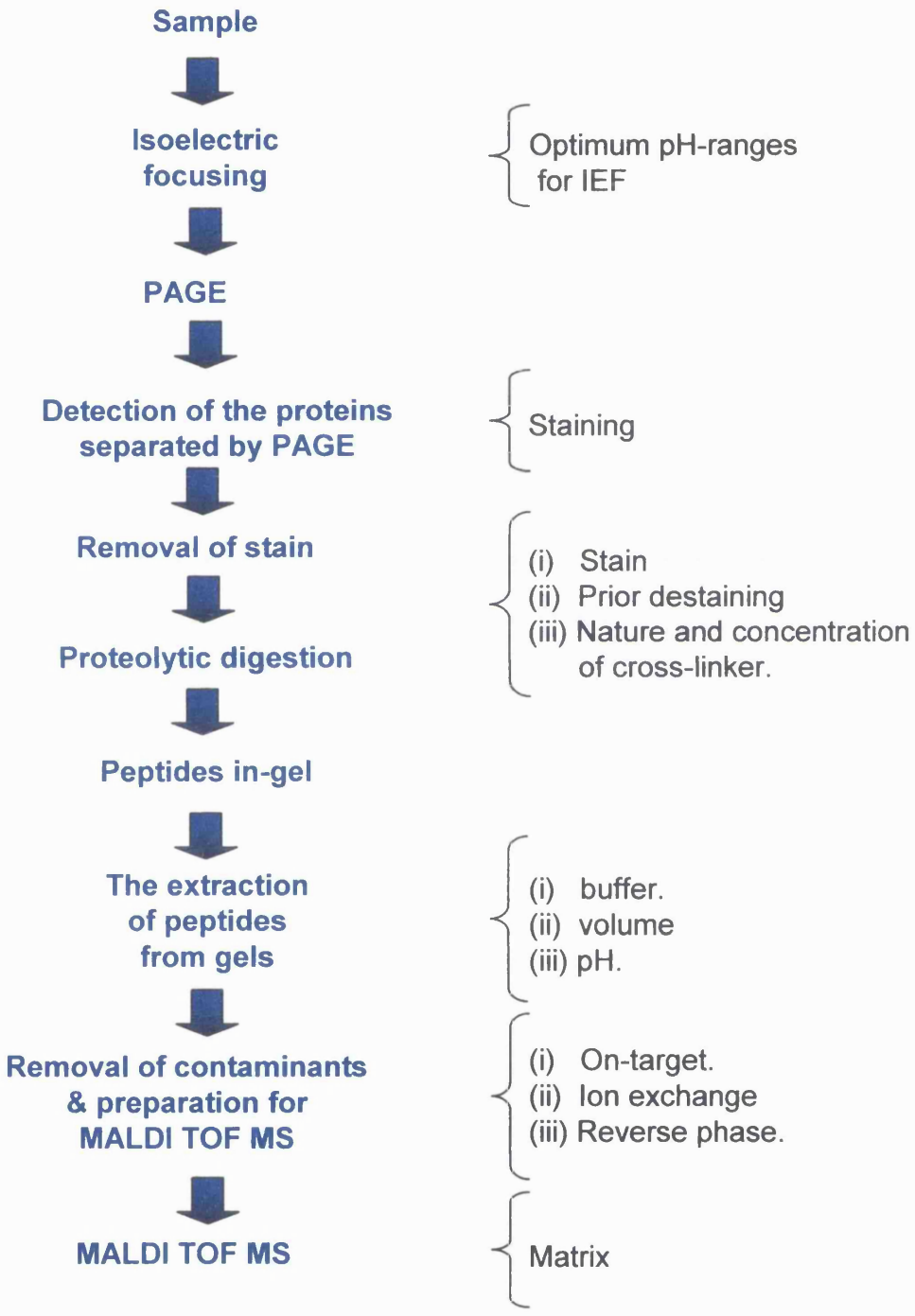


Figure 3.01 Schematic representation of the critical steps evaluated in this section.

3.1 Selection of the optimal pH-range for isoelectric focusing.

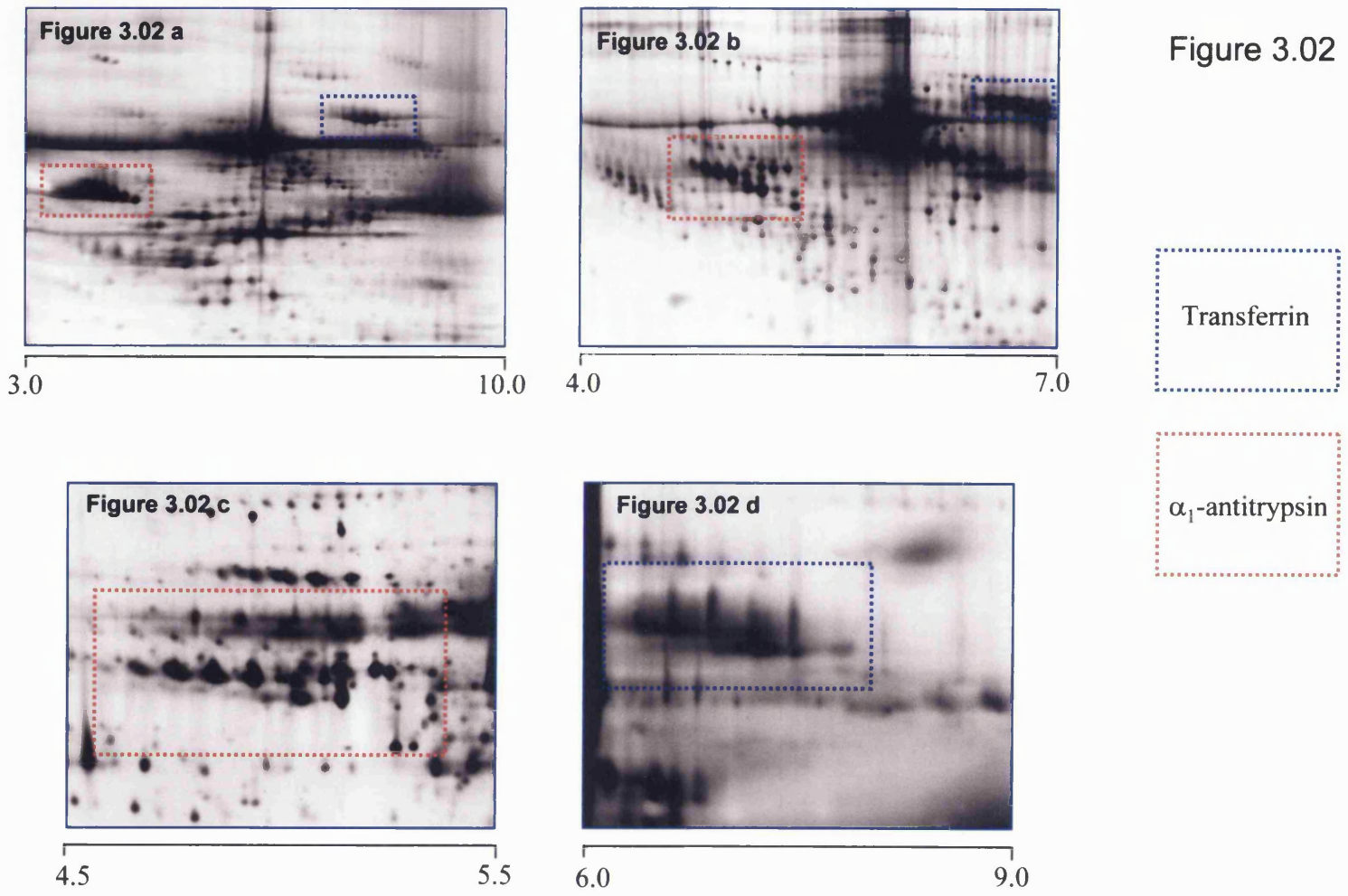
Since its development by O'Farrell in the mid-1970s, 2D-PAGE has undergone many refinements and improvements (O'Farrell, 1975) to overcome the problem of reproducibility that was associated with the original procedure. The recent commercial availability of immobilised pH-gradients and narrow pH-range focusing strips has improved the resolution and reproducibility of 2D-PAGE.

The first critical step in the analysis of plasma proteins using 2D-PAGE is to determine the optimal pH-range for IEF in the first dimension. Westbrook *et al.*, (2001) have shown that narrow range pH gradients are superior to larger range pH gradients for the separation of different protein isoforms over during 2D-PAGE analysis.

All the diseases studied in this work are rare and as the samples are often obtained from neonates the amount of plasma available is limited. It was therefore important to use the plasma samples economically.

3.1.1 Optimisation of IEF ranges.

Throughout the duration of this work, the only commercially available IEF strips were 18cm long and covered the pH-ranges of 3-10, 4-7, 4.5-5.5 and 6-9. Strips of these pH-ranges were evaluated for their suitability for the analysis of plasma proteins by 2D-PAGE. The 2D-PAGE analysis of CDG plasma proteins focused over the pH-range 3-10 (Figure 3.02a) shows that the pI of the majority of the plasma proteins lies within the pH-range 4-7. It should also be noted that the resolution of the individual proteins over this pH-range is poor overall as only the major plasma proteins such as albumin, α_1-antitrypsin, transferrin and the immunoglobulins (light and heavy chains) are recognisable. Figure 3.02b shows the 2D-PAGE of the same CDG plasma focused over the pH-range 4-7. It can be observed clearly that there is a marked improvement in the resolution of individual proteins as well as in the number of proteins being separated (Figure 3.02a). This increase in resolution due to the use of the narrower pH-range, is particularly noticeable for α_1-antitrypsin (highlighted by a red box) and transferrin (blue box).



Figures 3.02 **Comparison of 2D-PAGE of human plasma using different IEF strips**
(a) IEF over pH range 3-10, (b) IEF over pH range 4-7, (c) IEF over pH range 4.5-5.5, (d) IEF over pH range 6-9.

The α_1-antitrypsin isoforms in the pH 3-10 gel are relatively poorly resolved with each isoform overlapping with the next (see Chapter V for an explanation of α_1-antitrypsin isoforms). However, by decreasing the pH-range from 3-10 to 4-7, a significant increase in the resolution and almost complete separation of each α_1-antitrypsin and transferrin isoform can be achieved. The use of high resolution IEF pH-strips with narrow pH-ranges, 4.5-5.5 and 6-9, accomplished the separation of all individual isoforms of α_1-antitrypsin and transferrin (Figure 3.02c and d, respectively).

The results shown in Figures 3.02c and 3.02d demonstrate that use of pH-strips over a narrower pH-range, allows the detection of subtle variations in the pI and M_r of each isoform, which could not be observed using wide range pH-strips. The use of narrow pH-range IEF also eliminates albumin, the most abundant plasma protein which makes up approximately 80% of the total protein content. The presence of albumin at higher protein loads during plasma 2D-PAGE analyses (> 5 µl of plasma) often results in presence of 'streaking' in the second dimension. This can consequently result in contamination of other smaller proteins with albumin and make mass mapping identification of lesser abundant proteins more difficult due to contamination. By choosing a pH-range outside the pI of albumin (5.5.-6.5) it is possible to exclude the albumin and to avoid its effects on the analysis of other proteins.

The greater resolution of the narrow range pH strips also allows the distinct separation of proteins with similar pI's that could not be achieved using the larger pH-ranges. This becomes extremely important in the subsequent mass spectrometry analyses. When proteins observed in the 2D-PAGE shown in Figure 3.02a were re-analysed using the high resolution IEF ranges (Figures 3.02c and d) they were often found to comprise of multiple protein molecules. Since identification of proteins by mass mapping and database searching depends upon the purity of the protein analysed, the better the resolution of the protein the greater the chance of successful identification.

Therefore a decision had to be made on the use and choice of pH-range for the 2D-PAGE, over which the initial analyses would be performed, with each pH-range evaluated having its advantages and disadvantages. The wide range IEF strips (pH 3-10) allowed the analysis of the majority of the proteins found in plasma, but gave poor resolution of individual proteins. The use of pH-range 4-7, significantly improved the

resolution of the individual proteins as observed in the pH 3-10 analyses, whilst resulting in the loss of only one major class of proteins, the immunoglobulins (light and heavy chains). However, the use of narrow range pH strips (4.5-5.5 and 6-9), improved the individual resolution of proteins even further and also resulted in the loss of the abundant contaminating protein, albumin, from the analyses.

It was decided therefore, that all initial analyses for the detection of markers of disease would be performed on a 2D-PAGE focused over the pH-range 4-7. There were two main reasons for this;

(i) The use of two narrow range IEF strips (pH 4.5-5.5 and 6-9) would require two separate analyses and involve the use of twice as much sample.

(ii) Although significant increases in resolution were observed using the narrow pH-ranges, which also result in the removal of albumin from the analyses, the use of a combination of two pH-ranges also results in the loss of all the proteins that fall within the pH-range 5.5 – 6.0. Thus this area of the 2D-PAGE gel which contains many proteins, would be totally excluded from the screening process.

3.2 The optimal staining technique for the detection and analysis of proteins after PAGE.

The detection of proteins within a polyacrylamide gel usually involves chemical staining techniques that are based on the aggregation of dyes or metal ions around the protein spots or bands (Patton, 2001). It is therefore important that the staining technique should be sensitive enough to detect the minor protein components of plasma and it should not covalently modify the protein molecule or interfere with any subsequent analyses such as proteolytic digestion. Although many staining techniques are available for the identification of proteins in PAGE, four were evaluated because of their amenability to subsequent analysis by mass spectrometry. These were silver nitrate, coomassie blue, copper chloride and zinc sulphate staining; the latter two staining processes are reverse stains whereby the gel takes up the stain and the protein bands remain clear. Variable amounts of transferrin (16 ng to 4.2 μg) were run in duplicate, on four identical 1D-PA gels (10% acrylamide with 1.3% PDA cross-linker). Each gel was developed using a different protein staining technique (Figure 3.03). These experiments were used to determine the levels of sensitivity of each staining technique what effect, if any, each stain had on any in-gel proteolytic digestions and the overall sensitivity possible for the analysis of proteins in gels.

3.2.1 Silver staining.

Two methods for silver nitrate staining were evaluated for their sensitivity and the amenability of stained proteins to mass spectrometry. The lower limit of detection of transferrin was 79 ng for both methods (Figure 3.03a, Table 3.1). However, the technique of Schevenko and colleagues (1996) produced a significant background staining relative to the method of Hochstrasser and Merril (1988). Further the method of Hochstrasser also gave much darker stained bands and this combined with the lower background staining produced much higher resolution of the individual protein bands.

3.2.2 Coomassie blue staining

The identical gels with the same protein loading were stained with coomassie blue (Dunn, 1995a) (Figure 3.03b). The lower limit for detection of transferrin was 1.2 µg of protein (Table 3.1)

3.2.3 Copper Chloride reverse staining

The 1D-PA gel was stained with copper chloride reagent (Lee *et al.*, 1987) to deduce the relative sensitivity of the staining technique (Figure 3.03c, Table 3.1). The lower limit of detection of transferrin was again 1.2 µg, although some protein was visible at 0.8-0.3 µg of protein (Table 3.1).

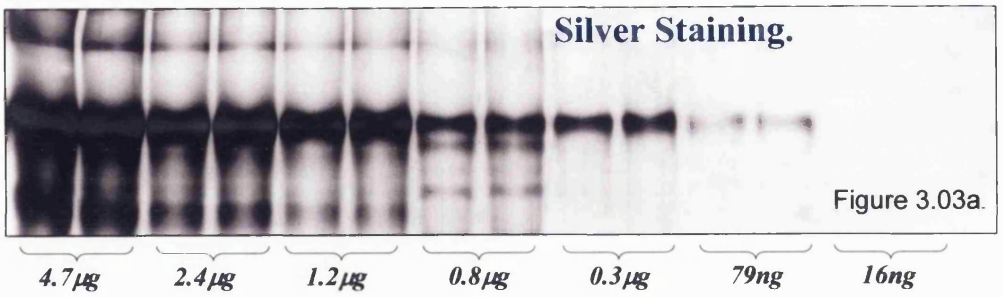
3.2.4 Zinc reverse staining

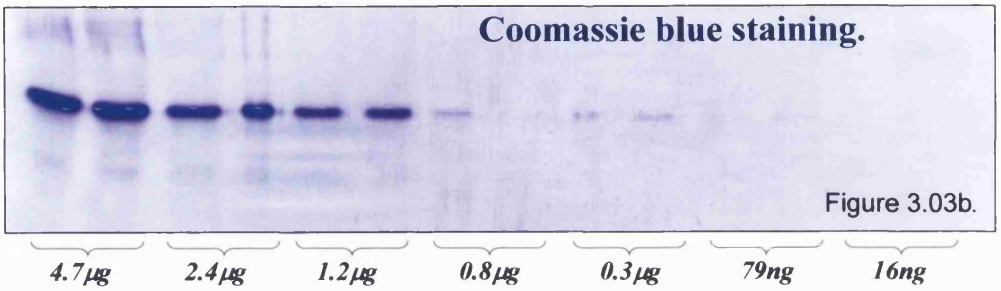
The 1D-PAGE was stained with zinc sulphate according to Ortiz *et al.*, (1992). Figure 3.03d shows the lower limit of detection for transferrin with this method was 1.2 µg, although some protein was visible at the 0.8 µg and 0.3 µg level (Table 3.1)

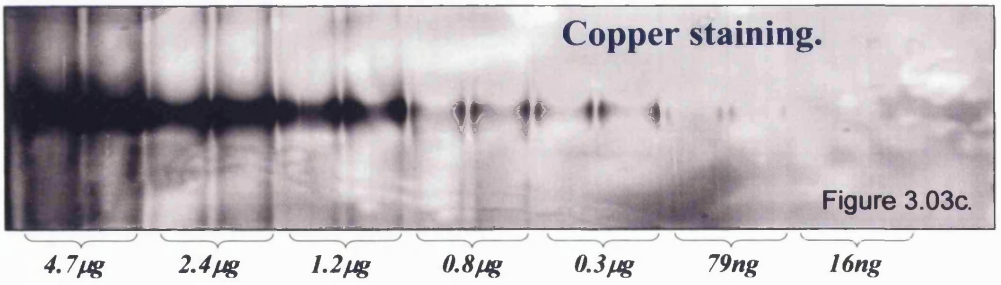
Table 3.1 Sensitivity of silver, coomassie blue, copper and zinc staining techniques

PAGE conditions 10% acrylamide: 1.3% PDA	(60 pmol) 4.7 µg	(30 pmol) 2.4 µg	(15 pmol) 1.2 µg	(10 pmol) 0.8 µg	(4 pmol) 0.3 µg	(1 pmol) 79 ng	(0.2 pmol) 16 ng
Silver (Hochstrasser *et al*)	✓	✓	✓	✓	✓	✓	X
Coomassie blue	✓	✓	✓	X	X	X	X
Copper	✓	✓	✓	X	X	X	X
Zinc	✓	✓	✓	X	X	X	X
✓- protein detected, X- protein not detected							

Figure 3.03







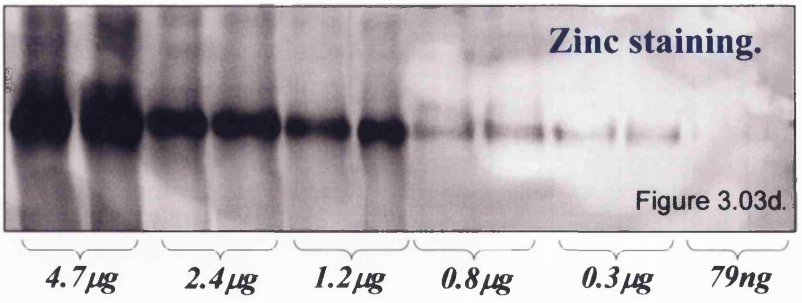


Figure 3.03 Investigation of the sensitivity of different staining procedures. Figures 3.03c and d have been produced as negative images for greater clarity.

3.2.5 Conclusions

Silver staining was the most sensitive technique with a limit of detection of 79 ng of protein (Figure 3.03 and Table 3.1). The other stains evaluated had similar levels of sensitivity for the detection of proteins in PA gels of between 0.8 and 1.2 μg of protein. This approximates to an order of magnitude less sensitive than silver staining.

Although no significant difference in the sensitivity of the two silver staining methods was observed, the Hochstrasser staining technique (Hochstrasser and Merril, 1988) takes approximately 4 hours per gel while the Schevenko silver staining method takes approximately 1 hour for 2 gels (Schevenko *et al.*, 1996). It was therefore decided that all analytical gels would be stained using the longer method of Hochstrasser because of the higher resolution, diminished background and higher resolution of individual spots. However, the preparative gels with higher loading of samples, could be stained using the silver stain method of Schevenko, coomassie blue or copper chloride staining. Zinc staining was omitted as an option for staining of preparative gels because it was considered the most technically difficult of the stains evaluated.

3.3 The optimisation of the in-gel proteolytic digestion of proteins in PA gels.

3.3.1 Introduction

Although silver staining had been shown to be the most sensitive detection procedure it does not necessarily mean it is the most appropriate visualisation method for the subsequent analysis of detected proteins. Therefore, the effects of each stain on subsequent in-gel digestion of the proteins and the analysis of the resultant peptide mixtures by MALDI TOF MS were investigated.

3.3.2 The effect of the staining technique on the in-gel digestion of proteins.

To evaluate the effects of each of the staining processes on in-gel proteolytic digestion, each of the transferrin bands from the 1D-PAGE staining experiments (see section 3.2), were excised from each gel and digested with trypsin in-gel as described by Schevenko and colleagues (1996). Where the staining techniques were not sufficiently sensitive to detect the lower concentrations of protein, the approximate area of the gel that the protein would have migrated to, was cut out and removed for analysis. The limit of sensitivity of each method was determined by the presence or absence in the mass spectrum of any tryptic peptides originating from transferrin. The results of the tryptic digestions of each differentially stained protein band from PA gels are shown in Table 3.2. The greatest sensitivity was achieved using the copper chloride staining technique (79 ng), even though the stain itself was insufficiently sensitive to visualise the protein band (Figure 3.03a). No tryptic peptides originating from transferrin were detected from the bands excised from the silver, coomassie blue and zinc sulphate-stained gels containing less than 1.2 μg of protein. The only peptides detected originated from keratin or were auto-catalytic products of trypsin, Figure 3.03b). More notably, the level of sensitivity for in-gel tryptic analyses attained from the silver-

level of sensitivity for in-gel tryptic analyses attained from the silver-stained gel was significantly below the levels described in the literature (Rosenfeld *et al.*, 1992).

The main difference between the staining techniques is that the copper chloride staining method is a reversible stain that does not result in the permanent fixing of the protein within the gel matrix. After the staining and visualisation of the proteins using copper chloride, the protein can be eluted from the gel matrix by chelation of the copper ions using Na_2^+EDTA (Lee *et al.*, 1987). This allows the tryptic digestion of the transferrin molecule to be performed in conditions more similar to those observed in solution than within the polyacrylamide matrix of the gel, as in the other staining techniques. These observations, therefore suggest that the composition of the polyacrylamide gels could be interfering with the subsequent in-gel proteolytic digestions.

Table 3.2 Limit of detection of proteins by tryptic digestion in PA gel stained by different procedures.

PAGE conditions 10% acrylamide : 1.3% PDA	(60 pmol) 4.7 µg	(30 pmol) 2.4 µg	(15 pmol) 1.2 µg	(10 pmol) 0.8 µg	(4 pmol) 0 .3 µg	(1 pmol) 79 ng	(0.2 pmol) 16 ng
Silver	✓	✓	✓	X	X	X	X
Coomassie blue	✓	✓	✓	X	X	X	X
Copper	✓	✓	✓	✓	✓	✓	X
Zinc	✓	✓	✓	X	X	X	X
✓- **transferrin peptide detected,** X- **transferrin peptide not detected**							

3.3.3 The effect of PAGE cross-linkers on the in-gel digestion of proteins.

The previous studies (section 3.3.1) suggested that the composition of the polyacrylamide gel may have played a part in the failure of in-gel tryptic digestion of proteins at the lower levels described in the literature (Rosenfeld *et al.*, 1992). The only obvious deviation from the original electrophoresis protocol of Laemmli (1970), during the initial experiments, was the substitution of the *bis*-acrylamide cross-linker by PDA. Therefore, *bis*-acrylamide was substituted for PDA to see if it had any effect on the PAGE and subsequent in-gel proteolytic digestions.

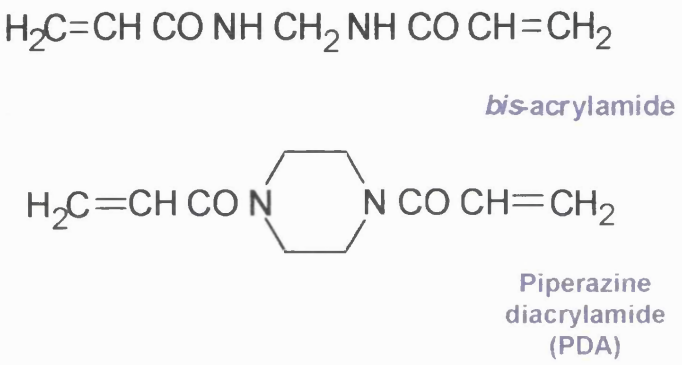


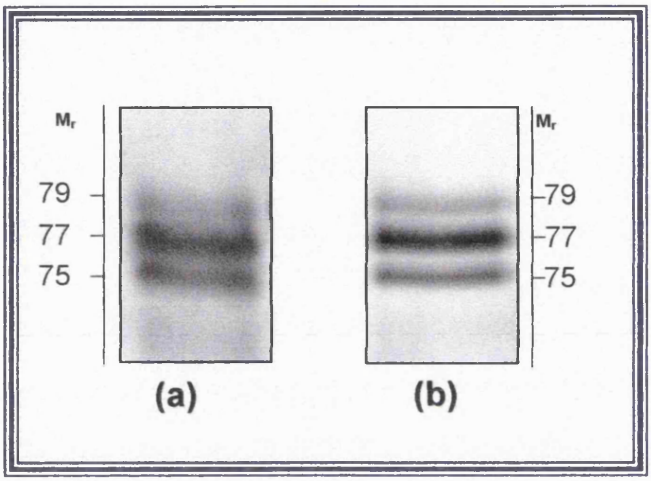
Figure 3.04 The chemical structures of methylene bis-acrylamide and piperazine diacrylamide

Comparable amounts of transferrin were analysed by 1D-PAGE as described earlier (section 3.2) using gels cross-linked with *bis*-acrylamide or PDA, respectively (10% resolving gel, 1.3% cross-linker). Each gel was then stained using silver according to the method of Schevenko and colleagues (1996).

Comparison of the two gels showed marked differences in both the resolution of the individual proteins and the background staining (Figure 3.05). 1D-PAGE with the resolving gel containing *bis*-acrylamide resulted in lower resolution of the individual protein bands than the 1D-PAGE with PDA cross-linked gel. However, the most

marked difference between the types of gels was the resultant background staining. The PDA-gel showed significantly less background than its *bis*-acrylamide cross-linked counterpart.

Figure 3.05 Comparison of silver staining of transferrin after 1D-PAGE using different cross-linkers (a) *bis*-acrylamide (b) PDA.



Although PA gels cross-linked with PDA gave better resolution and lower background staining than their *bis*-acrylamide equivalents, it was important to evaluate the possible effects of cross-linkers on subsequent in-gel tryptic digestions of proteins. Therefore, a control transferrin band was excised from both types of gel and digested with trypsin.

The limits of sensitivity of detection of tryptic peptides were lower in the *bis*-acrylamide gels (Table 3.2) than those cross-linked with PDA (Table 3.3), with all the staining procedures except copper chloride. Transferrin peptides were detected from 79 ng of transferrin in the *bis*-acrylamide cross-linked gels, which is the same sensitivity as was observed with the copper chloride stained analyses in PDA cross-linked gels (section 3.3.2). These levels of sensitivity are similar to those reported previously (Rosenfeld *et al.*, 1992) and resulted in an increase in the sensitivity of approximately 15-fold over the amount detected in PDA cross-linked gels (79 ng→1.2 μg).

PAGE conditions 10% acrylamide:1.3% *Bis*	(60 pmol) 4.7 µg	(30 pmol) 2.4 µg	(15 pmol) 1.2 µg	(10 pmol) 0.8 µg	(4 pmol) 0.3 µg	(1 pmol) 79 ng	(0.2 pmol) 16 ng
Silver	✓	✓	✓	✓	✓	✓	X
Coomassie blue	✓	✓	✓	✓	✓	✓	X
Copper	✓	✓	✓	✓	✓	✓	X
Zinc	✓	✓	✓	✓	✓	✓	X
✓- **transferrin peptide detected,** X- **transferrin peptide not detected**							

Table 3.3 Limits of tryptic digestion detected after 1D-PAGE in gels cross-linked with *bis*-acrylamide.

Therefore, the effect of changing the cross-linker from PDA to *bis*-acrylamide, increased the limit of sensitivity of the in-gel tryptic digestion analyses that could be achieved. Hence tryptic peptides could be generated from medium to weak silver-stained proteins and potentially identified using MALDI TOF MS and mass mapping.

3.3.4 The effect of the concentration of the *bis*-acrylamide cross-linker on in-gel proteolytic digestion.

As the type of cross-linker can affect the in-gel proteolytic digestion of proteins, the effect of the concentration of cross-linker was also investigated. The results observed in section 3.3.3 had shown that at a concentration of 1.3% *bis*-acrylamide the sensitivity of peptide mass mapping of separated proteins was much better in-gels cross-linked with *bis*-acrylamide than PDA. It had been shown previously (Kuster *et al.*, 1997) that the use of a lower percentage of *bis*-acrylamide cross-linker (0.1%), increased the efficiency of the deglycosylation of glycoproteins by the endoglycosidase PNGase F. Therefore, 1D-PAGE of the same amounts of unglycosylated transferrin was carried out in 1% and 0.1% *bis*-acrylamide cross-linked gels. After electrophoresis, both gels were silver-stained and the protein bands cut out and analysed by in-gel tryptic digestion. Although no difference in the sensitivity of detection of peptides was

detection of peptides was observed between the two gel compositions, differences in the spectrum of peptides and coverage of transferrin were observed.

The differences in peptide profiles were even more exaggerated when another unglycosylated plasma protein, α_1-antitrypsin, was digested in low- and high percentage cross-linked gels and in-solution (Figures 3.06 and 3.07). The total amino acid sequence coverage of α_1-antitrypsin obtained with 0.1% *bis*-acrylamide gels was approximately 67%, whereas the coverage increased to 73% in gels containing 1.0% *bis*-acrylamide. Conversely, the coverage was decreased to 49% when the same tryptic digestion was performed in solution (Figure 3.06). It can be seen from the mass spectra that the different digestion conditions result in markedly different peptide responses. The differences in the mass spectra obtained in each reaction medium, were illustrated clearly by plotting the response for each peptide ion observed on the MALDI TOF MS against the amino acid sequence (Figures 3.07a, b and c). The results demonstrate that alterations in the proteolytic digestion conditions can result in changes in the digestion profile of the protein. This phenomenon can be exploited during protein analysis for example, when monitoring site-specific glycosylation sites (Mills *et al.*, 2000). Figure 3.07a shows the tryptic digestion of deglycosylated α_1-antitrypsin in a low percentage cross-linker PA gel. Peptides containing two out of the three glycosylation sites (sites 83 and 247) were detected, whilst a peptide covering the first glycosylation site (46) was not evident. However, the same analysis performed in a higher cross-linked gel (1.0% *bis*-acrylamide) resulted in peptides containing all three glycosylation sites being detected (Figure 3.07b) and a more extensive sequence coverage of the molecule than was attained during the same analysis in the lower percentage cross-linked gels (0.1% *bis*-acrylamide). The tryptic digestion of α_1-antitrypsin performed in solution also resulted in a change in the digestion profile. However, unexpectedly a fall in the sequence coverage was evident with an amino acid sequence coverage of 49% (Figure 3.07c) and the detection of only one of the 3 glycosylation sites (Asn 83). The reasons for these changes in sequence coverage and mass spectra are unclear but they possibly result from the different digestion matrices affecting the protein conformation or the way in which it is presented to the protease during the digestion.

Figure 3.06

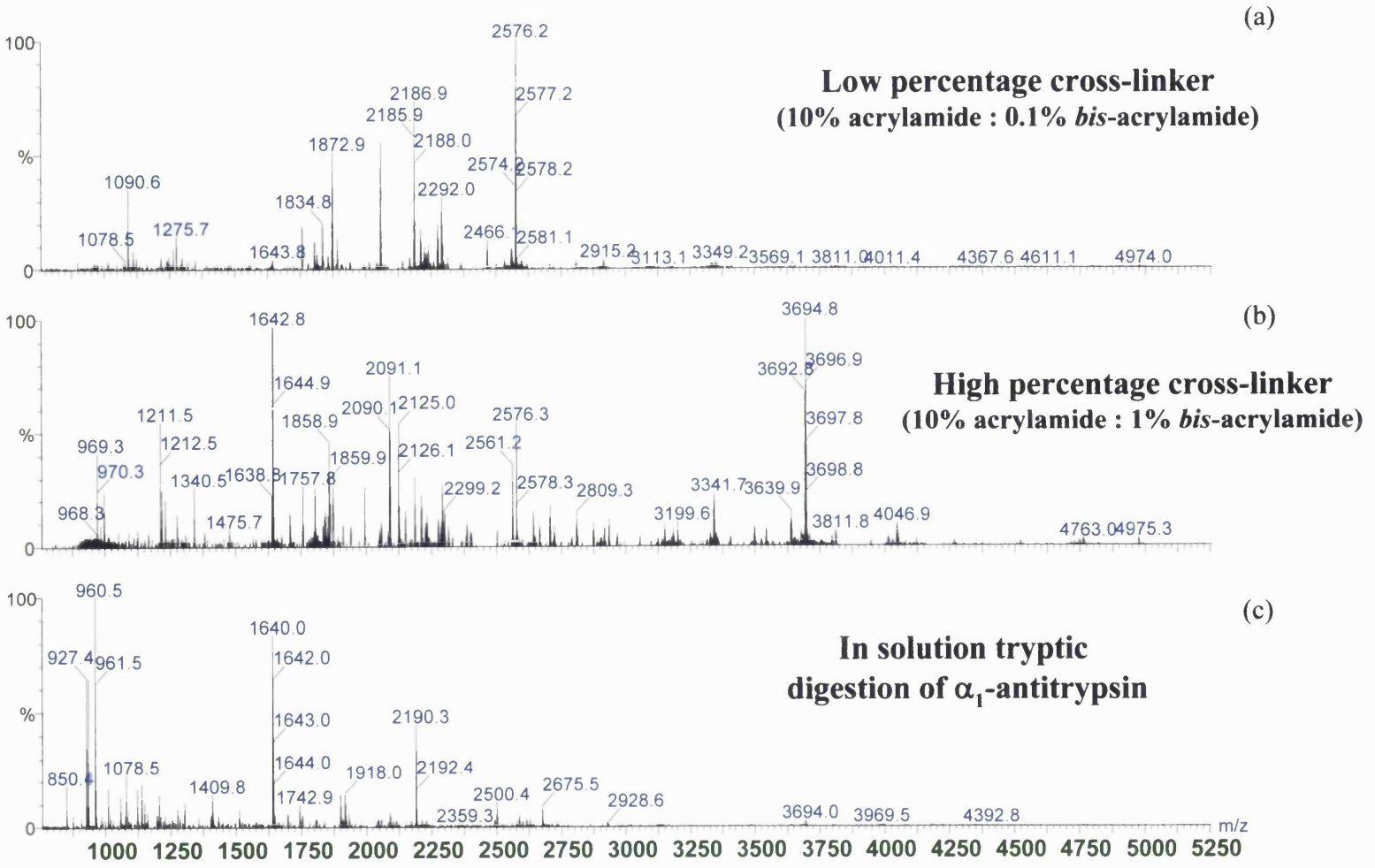


Figure 3.06 Tryptic digestion of α_1-antitrypsin in (a) Low % gels (b) High % gels (c) In solution.

Figure 3.07

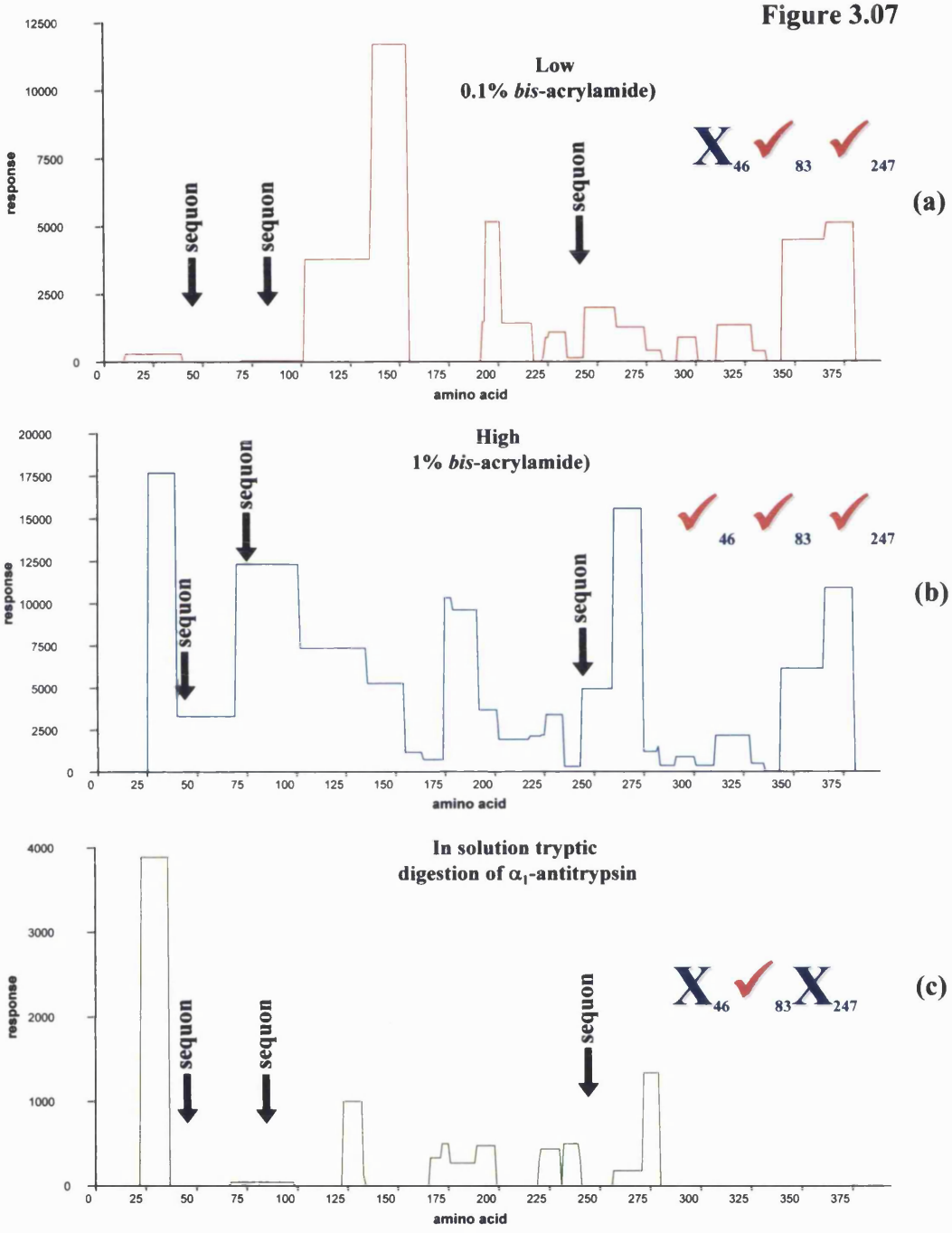


Figure 3.07 Tryptic digestion of unglycosylated α_1-antitrypsin. Intensity of masses and coverage of amino acid sequence.

X- peptide containing a glycosylation site not covered.
✓- peptide containing a glycosylation site detected

3.3.5 Conclusions

The effect of changing the cross-linker from PDA to *bis*-acrylamide resulted in a significant effect on each staining technique and any subsequent in-gel digestions. Although the sensitivity of the silver stain was not affected by the two different cross-linkers, marked increases in the background staining were observed during silver staining of the *bis*-acrylamide gels compared to the PDA cross-linked gels. This effect became more pronounced during the staining of low concentrates of protein when longer development times are necessary for the proteins to be visualised. This phenomenon is due to the amide hydrogen in the *bis*-acrylamide molecule which binds silver ions during staining (Hochstrasser *et al.*, 1988a). However, in PDA the amide hydrogens are replaced by a piperazine ring (Figure 3.04) which does not cause the binding of silver ions and lower background staining is observed. Another advantage of using PDA gels was the high tensile strength of the gels produced, an important factor when handling the large and fragile 2D-PAGE gels, which are extremely susceptible to tearing during the staining process.

Nevertheless, the use of PDA as a cross-linker had detrimental effects on proteolytic in-gel digestions, severely affecting the level at which in-gel tryptic digestion of proteins was possible. The reason for this phenomenon is unclear but it may be due to the increased rigidity of the gels cross-linked with PDA, relative to their *bis*-acrylamide counterparts, creating a steric hindrance that affects the action of the protease. The replacement of PDA with *bis*-acrylamide resulted in significant increases in the sensitivity of in-gel digestions. *Bis*-acrylamide gels resulted in greater sensitivity of in-gel tryptic digestions than their PDA equivalents but the lower limit for in-gel digestion was the same as that for digestion in free solution (~1 pmol). The reason for this is unclear but below 1 pmol of transferrin, both in solution and in-gel, the only peptides observed were tryptic auto-digestion products. Subsequent experiments varying the ratio of protease to protein or including detergents with the digest did not result in any changes in this limit of digestion (data not shown).

The lowering of the amount of *bis*-acrylamide cross-linker did not result in any advantage in terms of sensitivity of tryptic digestions. However, it was found

subsequently that the efficiency of chymotryptic in-gel digestions was increased in gels containing 0.1% cross linker. In concentrations of *bis*-acrylamide greater than 0.1%, chymotrypsin produces no peptide products and only auto-catalytic proteolytic products are observed in the MALDI TOF MS (data not shown). This is similar to the findings of Kuster and colleagues (1997), who have shown that the concentration of cross-linker is critical for the optimal efficiency of N-glycanase.

Therefore it was decided, that if sufficient sample was available both analytical and preparative gels would be run. High resolution analytical gels with PDA as the cross-linker would be used to determine any subtle changes in pI or Mr due to changes in post-translational modification or amino acid substitutions. The use of PDA gives greater resolution of individual proteins and lower background staining. Once any changes in the pI or Mr have been detected, the same samples would be re-analysed on PA gels with *bis*-acrylamide cross-linker with the option of varying the concentration of cross-linker to suit the subsequent proteolytic digestion, deglycosylation with endoglycosidases or level of sequence coverage required.

3.4 The optimisation of the extraction of peptides after in-gel proteolytic digestion.

3.4.1 Introduction

Existing protocols for the extraction of peptides from gels usually include washes of the gel slice or spot with aqueous buffered solutions, followed by several washes with another solvent-based system (Shevchenko *et al.*, 1996). Therefore it was decided to assess several extraction solutions used traditionally for the solubilisation of peptides or small proteins, to see whether any improvements could be made to the original protocols for the extraction of peptides from PAGE matrices and their subsequent analysis by mass spectrometry. The solutions evaluated included aqueous, solvent and urea based solutions, and combinations of these, to deduce the optimum elution solution for the extraction of peptides from PAGE gels after tryptic digestion.

3.4.2 Evaluation of the optimum elution solutions of peptides from PA gels

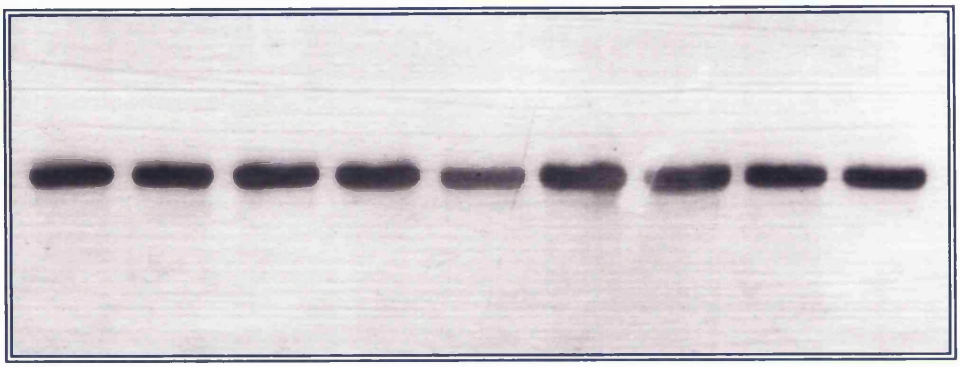
To assess the extraction efficiency of each solution, multiple samples of transferrin (1.2 µg) were analysed by 1D-PAGE containing 10% acrylamide and 1.3% *bis*-acrylamide (Figure 3.08a). The gels were stained with silver (Shevchenko *et al.*, 1996) and two protein bands per analysis, were excised, carboamidomethylated and digested in-gel with trypsin as described previously (Shevchenko *et al.*, 1996). The peptides were extracted from each protein band by shaking vigorously for 30 min with 300 µl of each elution solution evaluated (Table 3.4). When several elution solutions were used to extract the peptides, the solutions were combined and reduced to dryness and reconstituted in 100 µl of 0.1% TFA (containing 1 pmol of substance P internal standard) prior to desalting using C-18-microcolumns. In the two stage solvent- and urea-based systems (combination 6 and 7, Table 3.4), the first solvent-based extraction was reduced to dryness and the second urea wash (containing 1 pmol of substance P) was then added to the vessel containing the dried first extract. The urea solution was vortexed to reconstitute fully the dried peptide extracts and desalted using C-18 microcolumns. For consistency, all the extracted peptides for each elution solution were desalted using C-18 microcolumns prior to mass spectrometry. Peptide extracts were reconstituted in 10 µl of 0.1% TFA and 1.5 µl of this extract was used for mass spectral analysis.

Table 3.4 Elution solutions evaluated for the optimal extraction of peptides from a gel slice after in-gel tryptic digestion.

	Solution 1	Solution 2	Solution 3
1	300µl of 0.1% [TFA] *(30 min)*	*n/a*	*n/a*
2	300µl of 50%ACN :0.1% [TFA] *(30 min)*	*n/a*	*n/a*
3	300µl of 60%ACN :0.1% [TFA] *(30 min)*	*n/a*	*n/a*
4	300µl of 0.1% [TFA] *(30 min)*	300µl of 50%ACN :0.1% [TFA] *(30 min)*	300µl of 50%ACN :0.1% [TFA] *(30 min)*
5	300µl of 50%ACN :0.1% [TFA] *(30 min)*	300µl of 60%ACN :0.1% [TFA] *(30 min)*	*n/a*
6	300µl of 50%ACN :0.1% [TFA] *(30 min)*	300µl of [4M] urea (20min)	*n/a*
7	300µl of 60%ACN :0.1% [TFA] *(30 min)*	300µl of [4M] urea (20min)	*n/a*
8	300µl of [4M] urea (30min)	*n/a*	*n/a*
9	300µl of [8M] urea (30min)	*n/a*	*n/a*

Figure 3.08

(a)



(b)

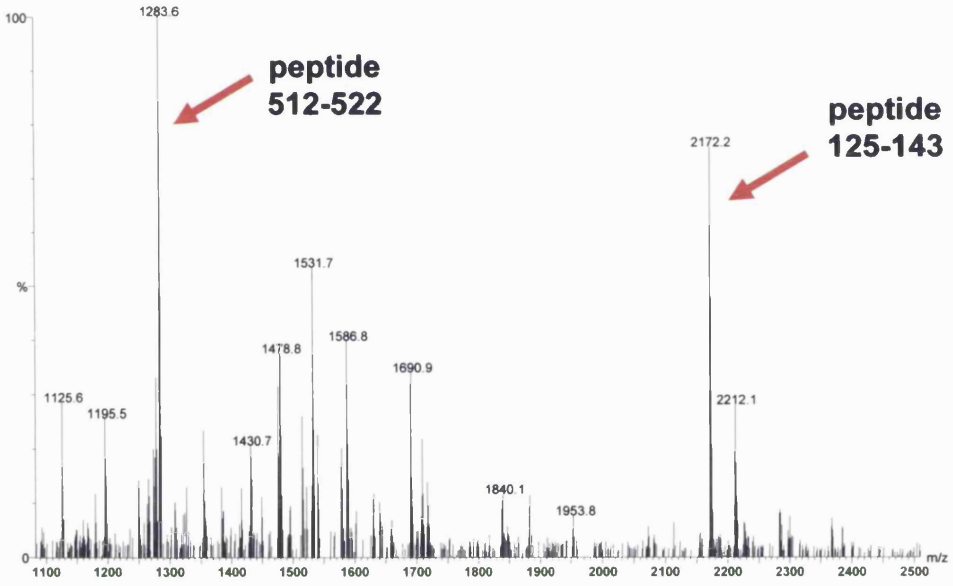


Figure 3.08

(a) 1D-PAGE of transferrin (1.2 μg) used to determine the optimum extraction solution of peptides from gels.

(b) MALDI TOF mass spectrum of the tryptic in-gel digestion of transferrin from Figure 3.08a

Figure 3.09

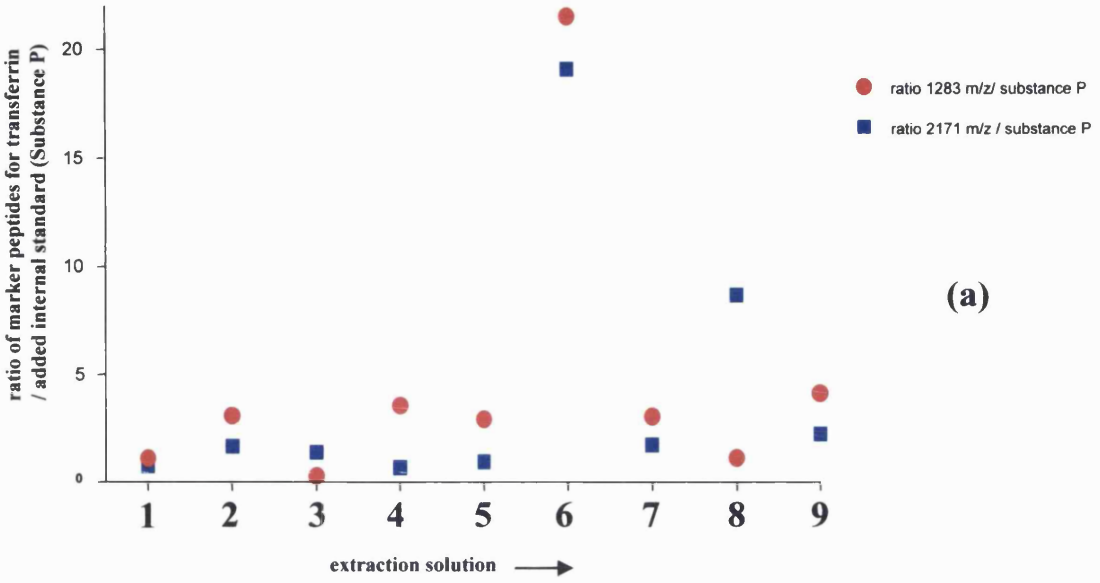


Figure 3.09a Determination of the optimal extraction procedure for peptides after an in-gel proteolytic digestion

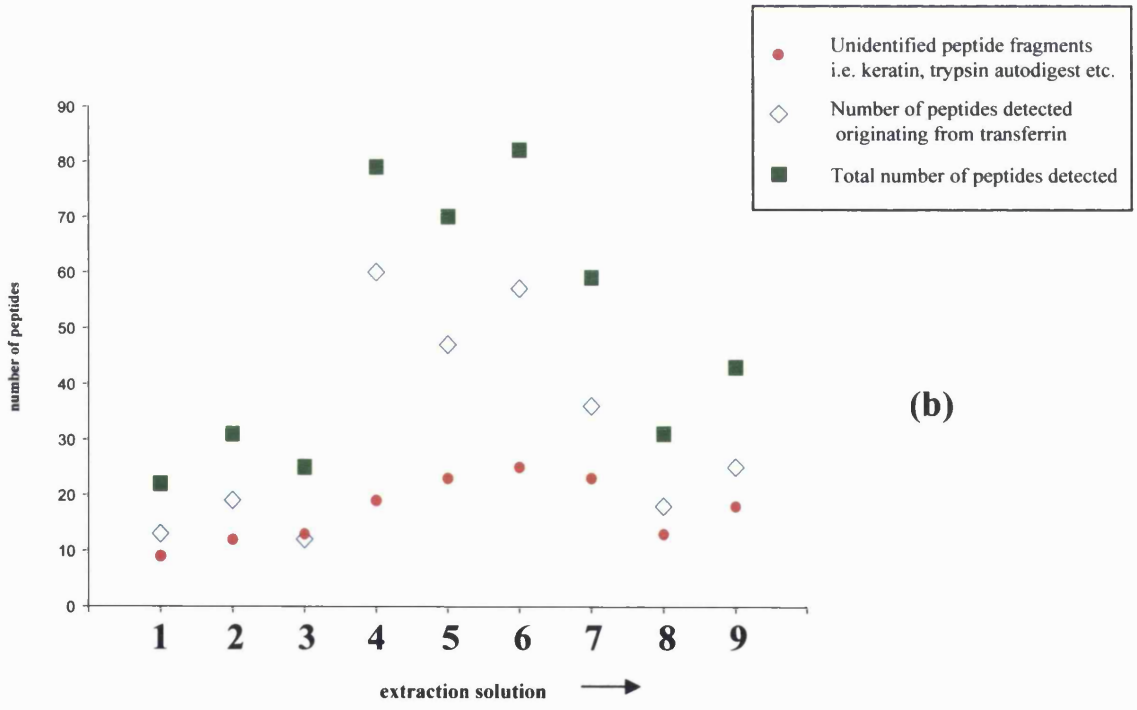


Figure 3.09b Total number of peptides observed by MALDI TOF MS after an in-gel digestion of transferrin and extraction using different procedures.

The recovery of the peptides from each gel slice was determined by comparing the peak height of the two major ions 1283.6 m/z and 2172.2 m/z, amino acids 512-522 and 125-143, respectively (Figure 3.08b), to that of the internal standard substance P (Figure 3.09a). The best extraction solution was combination 6 which consisted of 300µl of 50% acetonitrile containing 0.1% TFA wash for 30 minutes, followed by a further extraction with 300 µl of 4 M urea for 30 minutes (Table 3.4). The yield of the peptides observed from the tryptic digestion of transferrin using this elution protocol was considerably higher than for all the other extraction solutions.

The total number of peptides observed during the extraction process is also dependent on the elution protocol used (Figure 3.09b). Although three elution protocols resulted in more than 70 peptides being observed in the mass spectra, the most efficient combinations were the extraction combinations 4, 5 and 6. However, the yield of the peptides observed with combination 6 was greater and therefore it was considered the optimum elution procedure.

With all the elution solutions, approximately 25% of the peptides detected could not be attributed to the tryptic digestion of transferrin. Although most of these peptides can be attributed to keratin or to trypsin autodigest products, in some cases as much as 10% of the total number peptides observed could not be assigned to one of these proteins. The origin of the peptides is hard to deduce without determining their sequence but they probably originate from non-specific cleavages of the transferrin molecule by trypsin.

3.4.3 Conclusions

The ideal method for assessing the recovery of peptides after proteolytic digestion of proteins in gels is to use radioactively labelled proteins. However, Nicola and colleagues (1995) have shown that internal standards and MALDI TOF MS can be used to quantitate the yield of peptides as long as the component to be measured and the internal standard are within 3 orders of magnitude of one another. Preliminary studies showed that the amount of substance P needed to spike the eluted peptides was 1 pmol for the in-gel tryptic digestion of 1.2 µg of transferrin. At no time during any

quantitative analyses was the signal observed or allowed to saturate the detector.

The best elution protocol was found to be combination 6 (Table 3.4), 50% ACN : [0.1% TFA] followed by 4 M urea (combination 8). The recovery of peptides using this combination was approximately 5 times greater than with 4M urea. The recovery was also an order of magnitude greater than that for the original protocol (Shevchenko *et al.*, 1996) which consisted of washing with 100 mM ammonium bicarbonate buffer, followed by 3 washes with 50% ACN: 0.1% TFA. Unexpectedly, the yields from the extraction of the peptides from the gel slice using either 50% acetonitrile [0.1% TFA] or 4M urea alone were not additive. An explanation for this observation could be that the second extraction solution not only increased the yield of the peptides extracted from the gel slice, but more probably contributed to the re-solubilising those peptides, that may have been absorbed onto the surface of the vial during the solvent reduction procedure. However, if urea was used in any extractions of peptides from polyacrylamide gels, it would have to be removed prior to any MALDI TOF MS analyses as it suppresses the ionisation process of the peptides. Surprisingly, 4M urea showed a preference for extraction of the peptide 2171.2 m/z over the peptide 1283.6 m/z, which was not observed in any of the other extraction protocols. The reason for this phenomenon is unclear but it probably reflects the different solubility of the respective peptides in the different solutions.

3.5 The optimisation of the removal of contaminants prior to MALDI TOF MS.

3.5.1 Introduction.

MALDI TOF MS is often described as the most salt-tolerant mass spectrometry technique. This coupled to the high throughput of samples, makes it the most popular choice for the analysis of peptides after in-gel proteolytic digestion. However, various groups have shown that removal of contaminants prior to, or following in-gel digestion, significantly improves the peptide signal observed by MALDI TOF MS. The destaining of silver-stained proteins, prior to in-gel proteolytic digestion has been shown to improve the signal to noise ratio and increase sensitivity during subsequent

mass spectrometry analyses (Gharahdaghi *et al.*, 1999). Removal of contaminants using a C-18-solid phase extraction (C-18-SPE) after in-gel digestion, also significantly improves the quality of MALDI TOF MS (Bagshaw *et al.*, 2000). To evaluate whether removal of buffers and contaminants, does improve the analysis of peptide mixtures by MALDI TOF MS samples, of transferrin, which had been analysed by 1D-PAGE, were digested with trypsin in the gel and subjected to different procedures to remove contaminants before mass spectrometry.

3.5.2 Destaining of gel slices, C-18-SPE, ion-exchange and on-target washing as methods of removing contaminants prior to MALDI TOF MS.

Transferrin samples were separated using 1D-PAGE and silver-stained as described earlier (see Figure 3.08). Each protein band was excised and analysed to assess the effects of prior destaining or post digestion purification on the MALDI TOF MS spectra of the peptide mixtures. Figure 3.10a shows the mass spectrum of an in-gel tryptic digestion mixture of transferrin, which has not been desalted prior to analysis. It can be seen that the spectrum is relatively weak, with a low signal to noise ratio. However, the spectrum did provide sufficient ions for the identification of transferrin by mass mapping. The spectrum observed after the transferrin band had been destained (Gharahdaghi *et al.*, 1999) prior to in-gel tryptic digestion had a slightly higher signal to noise ratio and a greater yield of peptides was observed (Figure 3.10b). The spectra obtained permitted correct identification of the protein by mass mapping and with an increase in the percentage of transferrin peptides detected from 33% to 47% during the database search.

On-target desalting of peptides is a quick and simple method to remove contaminants and improve the spectra obtained using MALDI TOF MS (Beavis and Chait, 1990). A small amount of H_2O (~1 µl) is put onto the surface of the sample/matrix to allow the contaminants to diffuse into the H_2O overlay. After a few seconds, the H_2O containing the contaminants is removed carefully using a pipette and discarded, and the target is re-analysed.

Figure 3.10

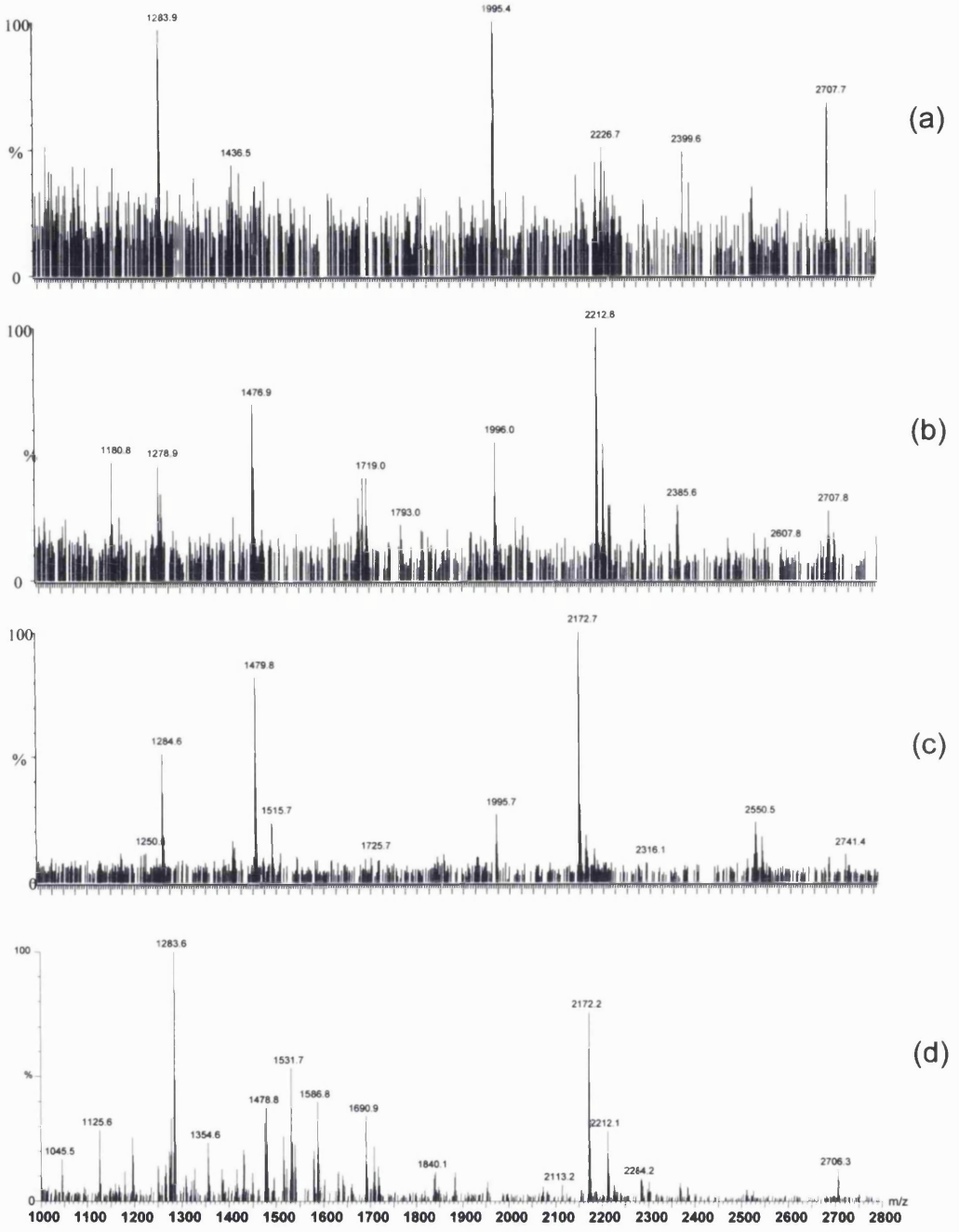


Figure 3.10 In-gel digestion of transferrin (1.2 μg)

(a) No desalting.
(b) Gel slice de-stained prior to tryptic digestion and no desalting.
(c) Peptides de-salted on target.
(d) Peptides de-salted using C18 solid phase extraction columns.

The spectrum obtained after the on-target desalting of the same peptide is shown in Figure 3.10a. It can be seen that there was an improvement in the signal to noise ratio compared with the original non-desalted sample (Figure 3.10a). Unexpectedly, this was also accompanied by a change in the spectrum observed relative to the original analysis. This variation in the spectra is probably due to changes in the way the matrix recrystallises after removal of the contaminants using H_2O. Nevertheless, the spectrum obtained was an improvement on the initial non-desalted sample in terms of signal to noise even though the number of peptides matching transferrin only increased marginally from 33% to 37%.

Figure 3.10d shows the analysis of peptides from an in-gel tryptic digestion, from which contaminants and buffers have been removed using a C-18-SPE microcolumn prior to MALDI TOF MS (see Materials and Methods). It can be seen clearly that the spectrum is better in terms of signal to noise, intensity and yield of peptides observed. The desalting step increased the total number of peptides observed from less than 10 in the previous analyses (Figures 3.10a, b and c) to greater than 60.

Preliminary experiments with peptide standards, demonstrated that ion-exchange columns had the same efficiency for desalting mixtures of standards in solution as the C-18-microcolumns. Unusually, the analysis of tryptic peptides of transferrin which had been desalted using ion-exchange cartridges (SCX, Millipore) produced no peptide ions, including those of keratin, during MALDI TOF MS. The reason for this remains unclear but may be due to unknown contaminants from the PAGE gel extract affecting the binding or retention of the peptides by the ion-exchange column during the desalting procedure.

3.5.3 Conclusions.

The increase in sensitivity observed by desalting proteolytic digests indicates that either buffer contaminants from the digest (ammonium bicarbonate) or residual contaminants from the PA gels, suppress the ionisation of the peptides. Silver-stained PA gels may contain many contaminants and it has been shown that chelation of silver ions prior to tryptic in-gel digestion significantly improves the spectra obtained during

MALDI TOF MS (Gharahdaghi *et al.*, 1999). Although, improvements in the spectra were observed by de-staining the gel slices prior to proteolytic digestion, they were not as marked as expected. In contrast, the removal of contaminants from the peptide mixtures by solid phase extraction resulted in marked improvements in the mass spectra. This indicates that the presence of precipitated silver in the PA gel from the staining process does not affect the proteolytic action of trypsin appreciably, but may result in the suppression of peptides during MALDI TOF MS. The removal of the silver ions before or after the tryptic in-gel digestion does have some beneficial effects on the spectra observed during MALDI TOF MS. However, the removal of the contaminants after in-gel digestion clearly removes additional contaminants which have a detrimental effect on the mass spectrometry but which cannot be removed completely by on-target desalting. Thus it is more efficient to remove contaminants by the single step use of C-18- SPE after in-gel digestion prior to mass spectrometry to obtain better spectra. The inclusion of a prior destaining step does not seem to provide any significant advantages (Gharahdaghi *et al.*, 1999).

3.6 Treatment of peptides after their extraction.

3.6.1 Introduction.

Various protocols have been used to extract peptides from gels (Section 3.4) but they all include several extractions using aqueous solutions or solvent-based solutions. The eluted fractions are then combined and the volume is decreased by centrifugal evaporation or freeze drying, prior to analysis by mass spectrometry. Typically, the amount of peptide present in a silver-stained protein from a 2D-PAGE is in the low fmol range but the volume of extraction solution can be relatively large (>1 ml Shevchenko *et al.*, 1996). Subsequently, the volume of the elution solution has to be decreased to concentrate the peptides for analysis by mass spectrometry. Experiments were carried out to see whether the volume of the extraction solution affected the yield of peptides extracted and whether the nature of the buffer and the pH affected the recovery of peptides during concentration.

3.6.2 The effect of extraction volume on the recovery of peptides during concentration

To analyse whether the extraction volume affects the recovery of peptides during their concentration, 1 pmol of two peptides (angiotensin I and ACTH (clip 18-39)) was added to various volumes (3 µl to 1.2 ml) of 20 mM ammonium bicarbonate buffer / 50% ACN [0.1% TFA] (1:3 v/v) and taken to dryness by centrifugal evaporation in silanised Eppendorf tubes (1.8ml). Each sample was then resolubilised in 10 µl of 0.1% TFA containing 1 pmol of substance P and 1.5 µl was taken for analysis by mass spectrometry. The effect of sample volume on recovery, was deduced by ratioing the peak heights of the peptide standards, angiotensin I and ACTH (clip 18-39) to the peak height of the internal standard Substance P, obtained on mass spectrometry (Figure 3.11a).

The zero volume indicates the response when the two peptides (200 fmol) were spotted directly onto the target without going through a drying procedure. It can be seen clearly that the centrifugal evaporation of large volumes of extraction solution decreased the ratios of the peptide standards / internal standard, indicating considerable losses for both peptides for the more dilute solutions. This effect was more pronounced in volumes greater than 10 µl, compared with volumes less than 10 µl. This loss of peptides at volumes > 10 µl probably results from the adsorption of the peptide standards onto the surface of the vessel as the volume is decreased in the centrifugal evaporator. Similar results were also observed when samples were taken to dryness by freeze-drying (data not shown). These data imply that to minimise losses of peptides during concentration, all volumes should be kept to a minimum.

Figure 3.11

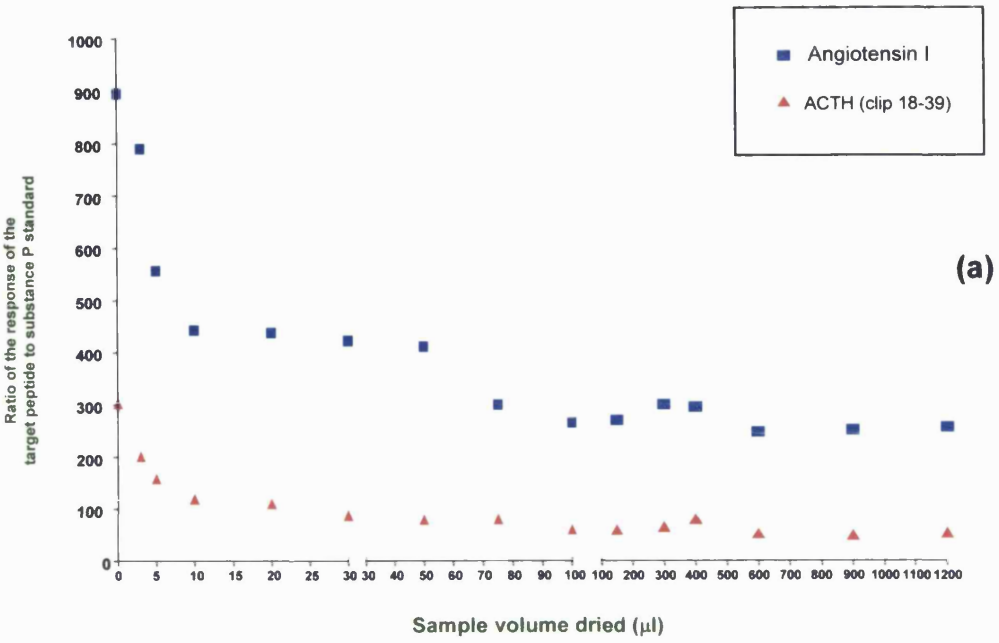


Figure 3.11a Analysis by MALDI TOF MS of the effect of volume on sample recovery.

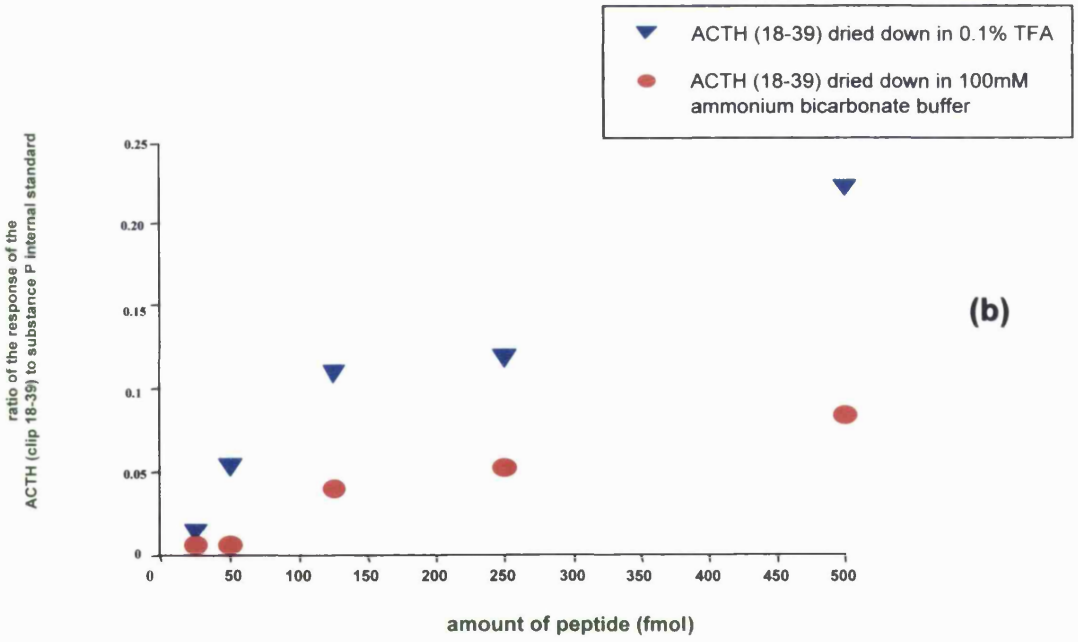


Figure 3.11b The effect of buffers on the recovery of sample during volume reduction.

3.6.3 The effect of pH and buffers on the recovery of peptides during concentration of extractions.

To analyse whether pH and the nature of the buffers have any effect on sample recovery during the drying process, different amounts of ACTH standard (12.5 fmol→500 fmol) were added to either 100 µl of 100 mM ammonium bicarbonate buffer or 100 µl of 0.1% TFA. Each sample was taken to dryness by centrifugal evaporation and resolubilised in 10 µl of 0.1% TFA, containing 0.5 pmol of substance P and 1.5 µl was taken subsequently for analysis by mass spectrometry. As in previous experiments, the sample recovery during the drying process was analysed by determining the ratio of the peak height of the ACTH to the peak height of the internal standard, Substance P (Figure 3.11b).

It can be seen clearly that greater losses of peptide occur when the peptides are concentrated in the ammonium bicarbonate buffer than in the 0.1% TFA solutions. These losses became worse at the lower concentrations of ACTH. The lower limit of detection of the ACTH standard was ~50 fmol of peptide in the ammonium bicarbonate solution, while 12.5 fmol was the lower limit of detection in the 0.1% TFA solution.

3.6.4 Conclusions.

The loss of sample during the drying process is shown in Figures 3.11a and b. These losses are far worse if the sample volumes are decreased in ammonium bicarbonate buffer. It is apparent that the presence of buffers and large sample volumes contribute to sample loss during the drying process, presumably due to absorption onto the sides of the vessel. These losses are more pronounced between the volumes of 50 µl and 1.2 ml, but the observed loss of peptides between these two volumes was not linear. Between the sample volumes, 50 µl → 1.2 ml, approximately 40% of peptides were recovered. This rose sharply between the sample volumes, 3 µl to 50 µl, resulting in approximately 80% recovery of peptides. The reason for this sharp rise in peptide recovery in the smaller volumes is presumably due to the volume of H_2O (10 µl) used

to resolubilise the peptides, being insufficient to completely resolubilise the peptides lost on the walls of the vessel when larger volumes of solution are used. The constant 40% recovery of peptides for the volumes 75 µl to 1.2 ml may also imply that a specific amount of peptide is lost selectively during the drying step. This may indicate that a finite number of binding sites are present on the surface of the vessel and that once they are occupied no further loss of peptides occurs.

The results shown in section 3.5 indicate that the inclusion of a C-18-SPE column to clean up peptides after in-gel digestion results in improved sensitivity. Therefore, if a C-18-SPE step is included in every analysis this would allow the use of urea during the extraction process, which had been shown previously to be beneficial (section 3.5). The incorporation of a urea wash after the first acetonitrile extraction presumably facilitates resolubilising the dried peptide extracts and eliminates the problem of larger volumes and sample loss. The inclusion of a C-18-SPE step prior to MALDI TOF MS removes the urea and other contaminants from the in-gel digestion, and concentrates the peptides. This allows the use of many other non-volatile, digestion buffers and therefore does not limit the type of enzymatic reactions to those which only work in volatile buffered solutions. Thus the combination of a new peptide extraction procedure coupled with a C-18-SPE desalting step, should result in more efficient extraction, sample handling and mass spectral analysis of peptides from a gel.

The reasons for the greater losses of peptides in the ammonium bicarbonate solution are unclear, since this buffer is volatile and should be removed during the drying process and by the addition of the 0.1% TFA solution containing the internal standard (substance P). A possible explanation for the losses of the ACTH (clip 18-39) standard could be the pH of the solutions to which the peptide was added. The pH of the ammonium bicarbonate buffer used during the experiment was approximately 8, while the pH of the 0.1% TFA was approximately 1.5. At pH 1.5, the majority of peptides will be protonated and less likely to be attracted or absorbed onto the surface of the vessel.

However, the mass spectral analysis of the samples containing the ammonium bicarbonate buffer also proved more difficult than the analysis of the samples dried in 0.1% TFA. The crystallisation of the α-cyano-4-hydroxy cinnamic acid matrix was

uneven in the buffer containing samples and hence the distribution of the sample on the target was varied. The reason for this effect on the crystallisation of the matrix is unclear, since all traces of buffer should theoretically have been removed prior to analysis.

3.6.5 Optimisation of the C-18-SPE desalting process.

C-18-SPE columns are commercially available in the form of disposable 50 µl pipette tips (Zip Tip, Millipore, UK). These pipettes can be used easily and efficiently to remove contaminants from peptide solutions. The peptides are adsorbed selectively onto the solid phase and the salts and other contaminants are removed by washing the solid phase several times with acidified H_2O. Finally the peptides are eluted using 50% ACN : [0.1% TFA]. However, if a final volume of 300 µl of 4M urea containing the extracted peptides is to be used, then the 50 µl volume of the pipette is too small. Therefore, microcolumns were constructed in house with a similar bed volume of C-18 solid phase (5-10 mg) to the Zip Tips but immobilised between two 5 mm (20 µm) frits, contained in a microcolumn with a 1 ml reservoir (Jones Chromatography, Bargoed, Wales). Unlike the commercial Zip Tips, in which the peptides are selectively removed and desalted by repeated passing of the peptide solution over the C-18 stationary phase, the microcolumns allowed the peptides to be extracted by one passage through the C-18 bed under gravity. As for the procedure for the Zip Tips, the peptides are washed with H_2O and subsequently eluted with acidified ACN solutions. Each step involved had to be validated to optimise the retention of the peptide samples, while maximising removal of contaminants. This section describes the optimisation of the priming, washing and elution parameters for the construction of C-18 microcolumns.

3.6.6 Optimal column priming, washing and elution parameters.

The commercially available Zip Tips, which use octadecasilane (C-18) as the stationary phase, are used according to the manufacturer's instructions by priming with

of the Zip Tips, the optimum conditions for their use i.e. priming, washing and elution volumes, had to be evaluated first. The following points were highlighted as potential problems in the cleaning up of peptide mixtures and are evaluated in this work.

(i) Failure of the peptides to bind to the column.

(ii) Release of bound peptides during the H_2O washing step.

(iii) Failure to elute all the peptides bound to the stationary phase.

To determine the optimal volume to condition the columns prior to the addition of the sample, four separate columns were primed with different volumes of priming solutions as shown in Table 3.5.

	Priming solutions **(i) 50% ACN: [0.1% TFA],** **(ii) 0.1% [TFA]**
Column 1	1 ml, 1 ml
Column 2	500 µl, 500 µl
Column 3	250 µl, 250 µl
Column 4	100 µl, 100 µl

Table 3.5 The volume of priming solution used to condition each column.

After conditioning of each column, 2 pmol of ACTH (clip 18-39) standard was added in 300 µl of 0.1% TFA. No urea or buffers were included because they would affect the subsequent mass spectral analyses. Therefore all experiments were conducted in volatile solutions which contained no contaminants. All eluate and wash fractions from each column were reduced to dryness in the centrifugal evaporator or freeze-dried. Each dried fraction was reconstituted in 100 µl of 0.1% TFA containing 1 pmol of substance P internal standard and analysed by MALDI TOF MS. The amount of ACTH (clip 18-39) present in each vial was determined by ratioing the peak height of

of substance P internal standard and analysed by MALDI TOF MS. The amount of ACTH (clip 18-39) present in each vial was determined by ratioing the peak height of ACTH (clip 18-39) against the peak height of the substance P standard, to obtain the approximate efficiencies of each priming volume.

The analysis of the amount of ACTH (clip 18-39) peptide present in the eluate that did not bind to the column, is shown in Figure 3.12a. It can be seen that as the priming volumes were decreased, the loss of sample increased. For the columns primed with 1000, 500, 250 and 100 µl volumes, the approximate amounts of peptides that failed to bind to the C-18 column were 50, 135, 470 and 490 fmol, respectively (Figure 3.12a). These data indicated that approximately 2.5%, 6.8%, 23.5% and 24.5% of the total sample added to the column was lost when the columns were primed with 1000, 500, 250 and 100 µl volumes, respectively.

Each column was then washed with 1000 µl of 0.1% TFA (used to remove the urea and salts in the analysis of proteins after their extraction from PA gels) and the washing eluant retained for analysis to determine the amount of sample lost during the washing step for each column. Figure 3.12b shows the amounts of ACTH (clip 18-39) detected in the 0.1% TFA washing eluant, after the columns were primed and washed with volumes of 1000 µl, 500 µl, 250 µl and 100 µl, respectively. It can be seen that further increases in losses of peptides occurred during the washing step when insufficient volumes of priming solutions were used. These losses were inversely proportional to the volumes of solutions used to condition the columns and similar to those losses seen in the sample break through fractions (peptides that failed to bind to the column). For the columns primed with 1000, 500, 250 and 100 µl volumes, the approximate amounts of peptides present in the wash fractions were 110, 245, 450 and 490 fmol, respectively.

It can be concluded that at priming volumes below 1000 µl, the column was insufficiently conditioned for optimal retention of the peptides. Increasing the column priming volumes from 1 to 3 ml did not result in an increase in the efficiency of the column (data not shown), which remained at approximately 90%.

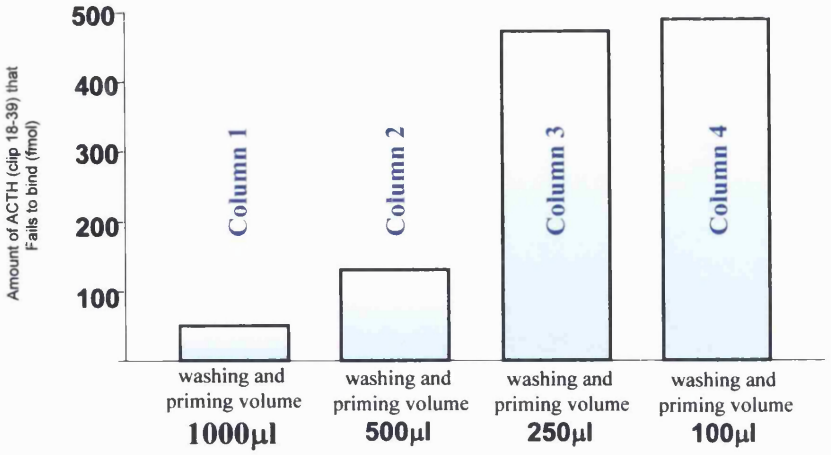
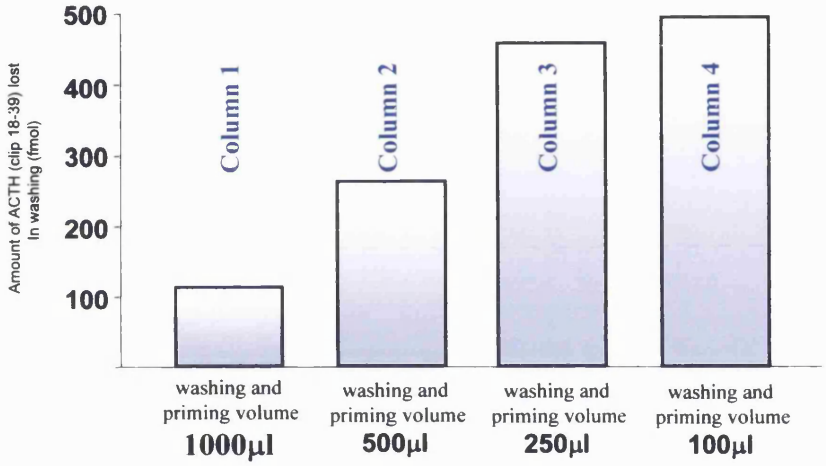
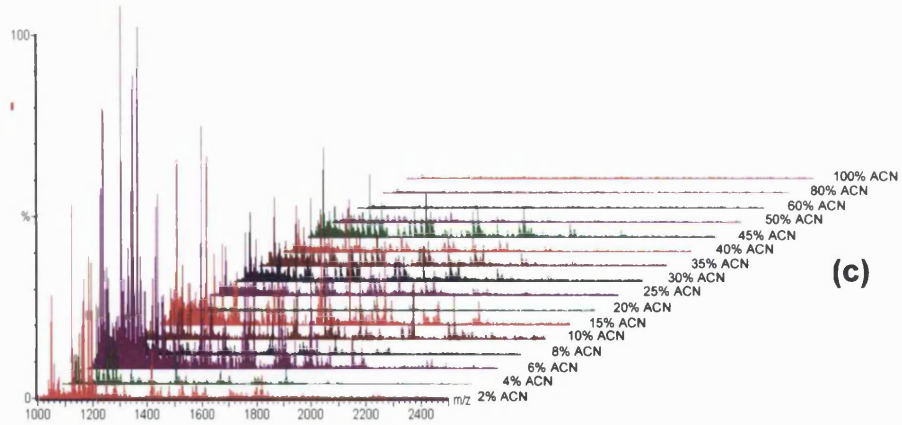


Figure 3.12

(a)



(b)



(c)

Figure 3.12

(a) Relationship between the amount of sample loss occurring due to peptides failing to bind to the column and diminishing priming volumes.

(b) Relationship between the amount of sample lost in the washing step (0.1% TFA) and volume used for prior priming of the column.

(c) Mass spectral profile of α_1-antitrypsin tryptic peptides eluted from a C18 column with increasing percentages of ACN.

The next parameters to be optimised were the composition and volume of the solution used to release the peptides from the C-18 stationary phase. Various groups have shown that increasing the percentage of organic solvent results in the elution of peptides from the C-18 stationary phase (Pipkorn *et al.*, 2002). To analyse the optimal concentration of acetonitrile required to elute peptides from the microcolumns, tryptic digestions of transferrin and α_1-antitrypsin (5 µg) in 300 µl of 0.1% TFA, were added to separate columns which had been optimally primed as described above (1 ml priming volumes). The break through and wash fractions were stored for analysis and the peptides eluted with 100 µl aliquots of solutions containing increasing percentages of acetonitrile (2%→100%). All eluants and washes were dried in a centrifugal evaporator prior to analysis by mass spectrometry. The mass spectral analysis of the eluate and aqueous wash fractions from both columns, showed that a small amount of peptide had not bound to the column (<10%). The elution profile of tryptic peptides of α_1-antitrypsin from a C-18 column are shown in Figure 3.12c. It can be seen that peptides are eluted from the column in concentrations of ACN as low as 2% and were observed to reach a maximum between 6% and 10% ACN, for both α_1-antitrypsin and transferrin. However, some peptides were found to continue to be eluted from the column with concentrations of ACN up to 50%, after which no further peptides were observed. This phenomenon was observed in both the transferrin and α_1-antitrypsin analyses and indicated that most peptides were eluted from the column with 50% ACN. These results are in accordance with the Zip Tip methodology, which utilises 50% ACN: [0.1% TFA] to elute all peptides from the C-18 stationary phase.

3.6.7 Conclusions.

A final protocol was established for the desalting and concentration of peptides for analysis by mass spectrometry. This method involved the construction of C-18 microcolumns that contained ~5 mg bed volume of C-18 stationary phase. These were primed with 1 ml of 50% ACN [0.1% TFA] and 0.1% TFA, prior to the addition of the sample. With priming volumes below 1 ml significant losses of peptides were observed. However, when the column was working optimally, the recovery of peptides

was >90%. The loss of approximately 10% of the sample during desalting was considered acceptable relative to the gains in both the sensitivity and quality of the mass spectra observed when the sample was not desalted prior to mass spectrometry (Figure 3.10a).

The results presented in Figure 3.11 show that all volumes should be kept to a minimum prior to any volume reduction steps using either centrifugal evaporation or freeze drying. Although 50% ACN: [0.1% TFA] was the optimal concentration of ACN for the elution of peptides, it was found that a minimum volume of 100 µl was required to remove all the peptides from the column (data not shown). Although use of the columns did not allow the elution of the peptides in small enough volumes to minimise sample loss during drying procedures, the benefits observed in terms of the greater yield of peptides eluted from gel slices using urea, far exceeded these losses. Therefore, until a column or a device is developed that can remove contaminants from a peptide mixture of approximately 300 µl with elution of the peptides in less than 10µl, potential improvements to the method are still possible.

3.7 The optimisation of the analysis of peptides using MALDI TOF MS.

3.7.1 Introduction.

Many matrix preparations have been described for the analysis of peptides and proteins by mass spectrometry (Kussmannet and Roepstorff., 2000). However, several groups have shown that addition of co-matrices to traditional matrix preparations can increase the performance of the MALDI process (Gusev *et al.*, 1995). In particular, Billeci and Stults (1993) have shown that the use of fucose as a co-matrix with 2,5-dihydroxybenzoic acid (DHB) resulted in increased sensitivity, suppression of matrix ions, increased signal to noise ratio, improved resolution and the analysis of glycopeptides. Gusev *et al.*, (Gusev *et al.*, 1995) have shown that DHB when used with fucose and/or methyl salicylic acid allowed the quantitative analysis of peptides using MALDI TOF MS. It was therefore decided to evaluate the use of DHB : fucose and other matrix preparations, to optimise matrix preparation for the analysis of small amounts of peptides.

3.7.2 Development of a new matrix preparation for the analysis of peptides.

3.7.2.1 Comparison of the sensitivity of various matrices for peptide analysis.

Initially five matrix preparations were evaluated for their sensitivity and applicability to the analysis of peptide mixtures using MALDI TOF MS. These included sinapinic acid (Brown and Gilfrich, 1991), α–cyano-4-hydroxycinnamic acid (thin film method, Vorm *et al.*, 1994), THAP/ammonium citrate (Papac *et al.*, 1996) and DHB/fucose co-matrices (Billeci and Stults, 1993). As discussed earlier the analysis of proteins after PAGE often requires the analysis of less then 100 fmol of peptides. The sensitivity of each matrix was compared by analysing serial dilutions of the peptide standards angiotensin I and ACTH (clip 18-39) by MALDI TOF MS (Table 3.6).

Table 3.6 Sensitivity of MALDI TOF MS during the analysis of standard peptides with different matrices.

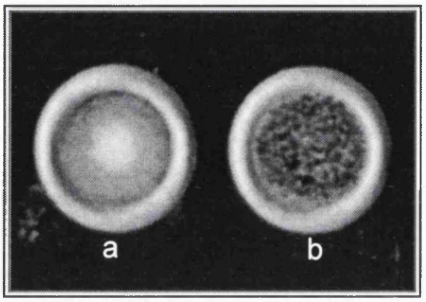
	500 fmol	250 fmol	125 fmol	62.5 fmol	30 fmol	15 fmol	7.5 fmol
(a) sinapinic acid	✓	✓	✓	✓	X	X	X
(b) α–cyano-4-hydroxy cinnamic acid (thin film method)	✓	✓	✓	✓	✓	✓	X
(c) THAP: ammonium citrate	✓	✓	✓	X	X	X	X
(d) DHB: fucose	✓	✓	✓	✓	X	X	X
(e) αC4HA : fucose	✓	✓	✓	✓	✓	✓	✓
✓ peptide detected, X peptide not detected							

It can be seen that out of the four matrix preparations assessed initially, the thin film method of Vorm *et al.*, (1994) was the most sensitive (Table 3.6). Therefore, it was decided to investigate whether incorporating fucose with α–cyano-4-hydroxycinnamic acid as a co-matrix could yield the same advantages as were observed with the DHB. Several concentrations of α–cyano-4-hydroxycinnamic acid and fucose, in different solvent systems were evaluated for the optimal configuration of the two matrices. It was found that a concentration of 10 mg/ml of α–cyano-4-hydroxycinnamic acid in ethanol: acetonitrile (1:1 v/v) when diluted 1:1 (v/v) with 50 mM fucose produced the optimum parameters for the analysis of peptides by MALDI TOF MS (data not shown). When this matrix preparation was mixed with 1:1 (v/v) with a peptide solution in 0.1% TFA it was discovered that the α–cyano-4-hydroxycinnamic acid precipitated selectively in the centre of the target whilst the fucose crystallised around the perimeter (Figure 3.13a).

Figure 3.13 Optimisation of matrix crystallisation.

(a) Crystallisation of α–cyano-4-hydroxycinnamic acid when fucose is incorporated into the matrix preparation.

(b) Crystallisation of α–cyano-4-hydroxycinnamic acid when no fucose is incorporated into the matrix preparation.



In order to evaluate whether the incorporation of fucose with α–cyano-4-hydroxycinnamic acid had any advantages to using α–cyano-4-hydroxycinnamic acid on its own, the same dilutions of peptide standards, were reanalysed to assess the sensitivity of the new matrix preparation. Table 3.6 shows that the use of fucose with α–cyano-4-hydroxycinnamic acid leads to a further increase in the sensitivity of the MALDI TOF MS analysis of peptides from 15 fmol with the thin film method, to below 7.5 fmol of peptide with the new preparation. The least sensitive of the matrix preparations was THAP: ammonium citrate (125 fmol) followed by sinapinic acid (62.5 fmol) and DHB (62.5 fmol). The use of α–cyano-4-hydroxycinnamic acid, either on its own or in combination with fucose, provided both significant improvements in terms of sensitivity and resolution over the other methods. This improvement in resolution of the peptides probably resulted from the other matrices requiring more energy (50% laser setting) to obtain peptide ions than is required with both α–cyano-4-hydroxycinnamic acid preparations (20% laser energy setting).

3.7.2.2 Comparison of various matrices for the analysis of peptide mixtures

It was also important to evaluate the sequence coverage and yield of peptides, during the analysis of protein digests, using each matrix. Therefore, 10 μg of two glycoproteins fetuin (bovine) and transferrin (human) were reduced, carboamidomethylated and digested with trypsin (see Materials and Methods, Chapter II). Each digestion mixture was desalted using microcolumns as described earlier (section 3.5), reduced to dryness in a centrifugal evaporator and reconstituted in 0.1% TFA to give a final concentration of 500 fmol / μl. 1.5 μl of each tryptic digestion of transferrin and fetuin were subsequently analysed using the same matrix preparations as in the sensitivity studies (Table 3.6).

Figures 3.14 and 3.15, show the results of the mass spectral analysis of the tryptic peptides of fetuin and transferrin, respectively, when various matrices were employed. It can be seen clearly that for both protein analyses, the DHB: fucose preparation gave the poorest mass spectra of all the matrices evaluated. The greatest numbers of peptides observed in both analyses were observed using the THAP: ammonium citrate matrix. However, >20% of these peptides could not be identified as peptides or glycopeptides derived from fetuin and transferrin. The amino acid sequence coverage attained with each matrix preparation is shown in Table 3.7.

	Sinapinic Acid	**αC4HA (thin film method)**	**THAP: ammonium citrate**	**DHB : fucose**	**αC4HA : fucose**
Fetuin *(sequence coverage)*	**37.0%**	**37.0%**	**35.0%**	**31.0%**	**41.6%**
Transferrin *(sequence coverage)*	**38.2%**	**37.1%**	**41.8%**	**31.5%**	**48.7%**
Limit of Sensitivity	62.5 fmol	10 fmol	100 fmol	62.5 fmol	1 fmol

Table 3.7 Total amino acid sequence coverage observed with each of the matrices evaluated and their lower limit of sensitivity.

Figure 3.14

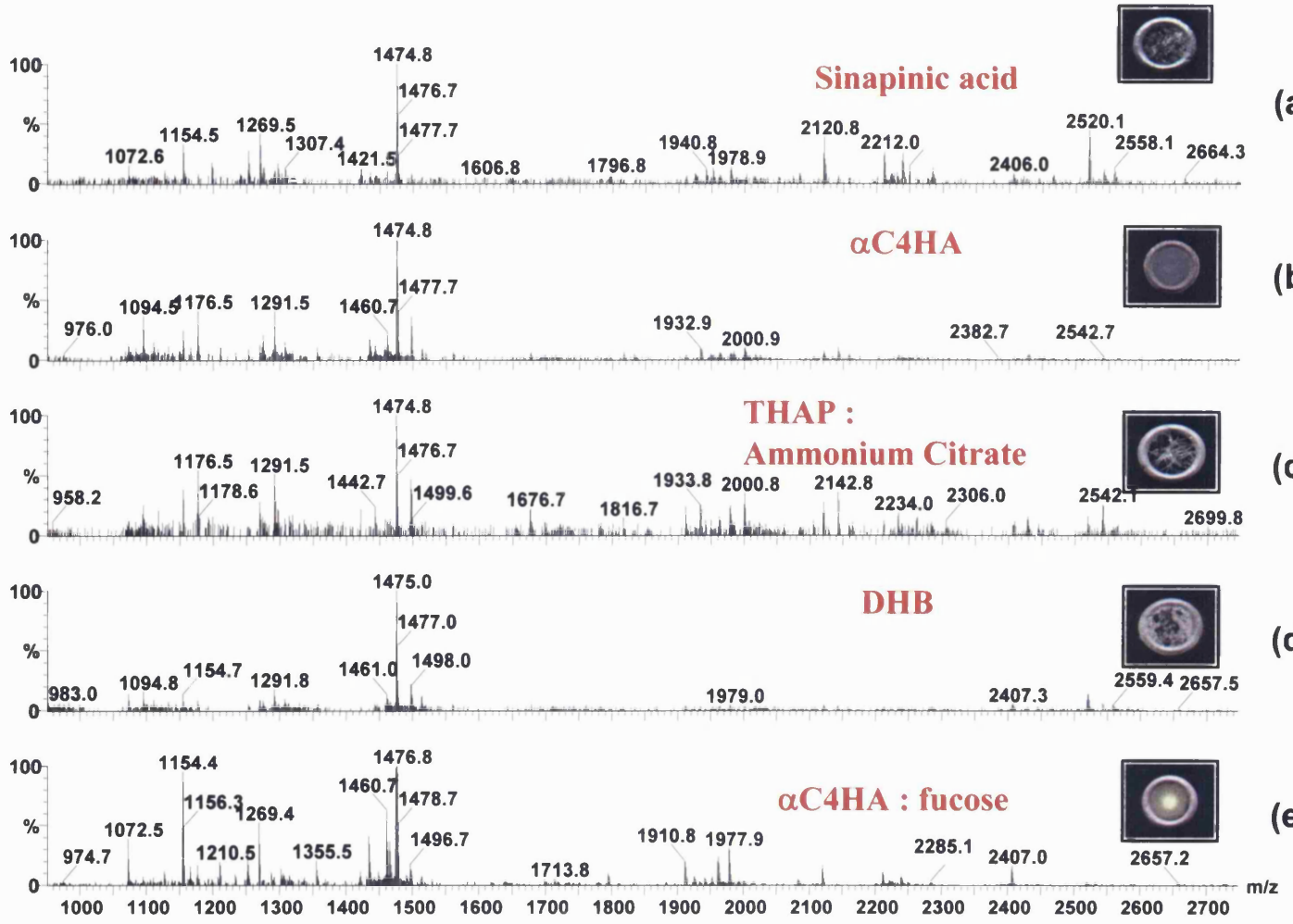


Figure 3.14 Analysis of fetuin tryptic peptides in various matrices by MALDI TOF MS.

Figure 3.15

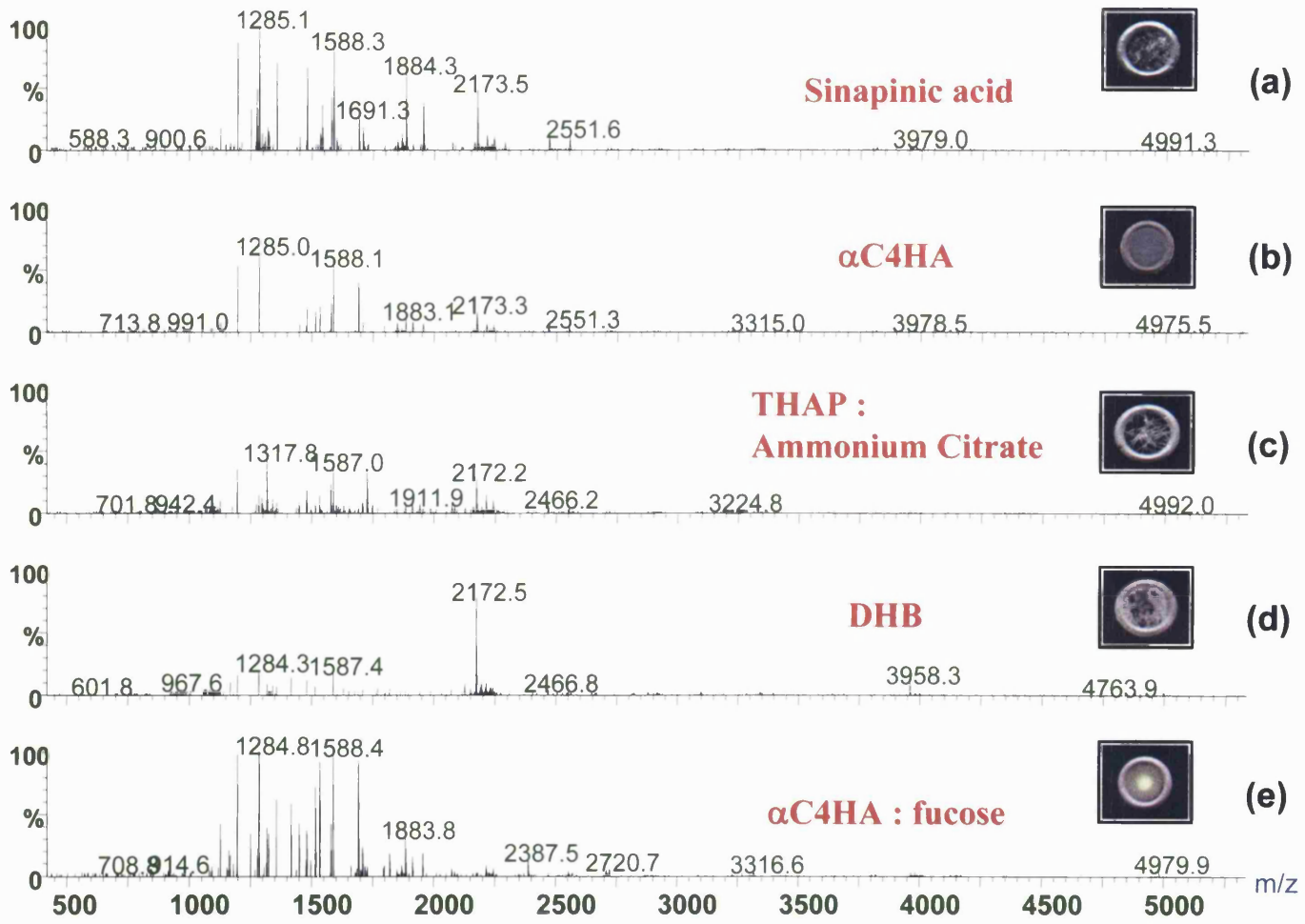


Figure 3.15 Analysis of transferrin tryptic peptides in various matrices by MALDI TOF MS.

The greatest amino acid sequence coverage was attained using the new matrix preparation of α–cyano-4-hydroxycinnamic acid: fucose, with 41.6% and 48.7% coverage of fetuin and transferrin, respectively. The DHB: fucose combination resulted in the lowest sequence coverage of both proteins (Table 3.7).

As for the earlier analyses of peptide standards to determine the limit of sensitivity of each matrix (Table 3.6), the tryptic peptides from the transferrin and fetuin analyses were serially diluted and analysed to determine the limit of sensitivity of each matrix (Table 3.7). In accordance with the previous results, both α–cyano-4-hydroxycinnamic acid preparations were the most sensitive matrices for the analysis of tryptic peptide mixtures. Similarly, the DHB: fucose combination was the least sensitive matrix, being an order of magnitude less sensitive than the sinapinic acid and THAP: ammonium citrate matrices. The α–cyano-4-hydroxycinnamic acid analyses without fucose demonstrated that peptides could be detected at the 10 fmol level. However an order of magnitude greater sensitivity was attained when fucose was used as a co-matrix, with a lower limit of detection of 1 fmol of peptide. Figure 3.16 shows the MALDI TOF MS analysis of transferrin and fetuin tryptic peptides in both α–cyano-4-hydroxycinnamic acid alone and α–cyano-4-hydroxycinnamic acid: fucose matrices, at the 50 fmol level. This amount was below the limit of sensitivity of the other matrices evaluated. It can be seen from the spectra that the yield of peptides observed was significantly higher in the α–cyano-4-hydroxycinnamic acid: fucose co-matrices than when α–cyano-4-hydroxycinnamic acid was used alone

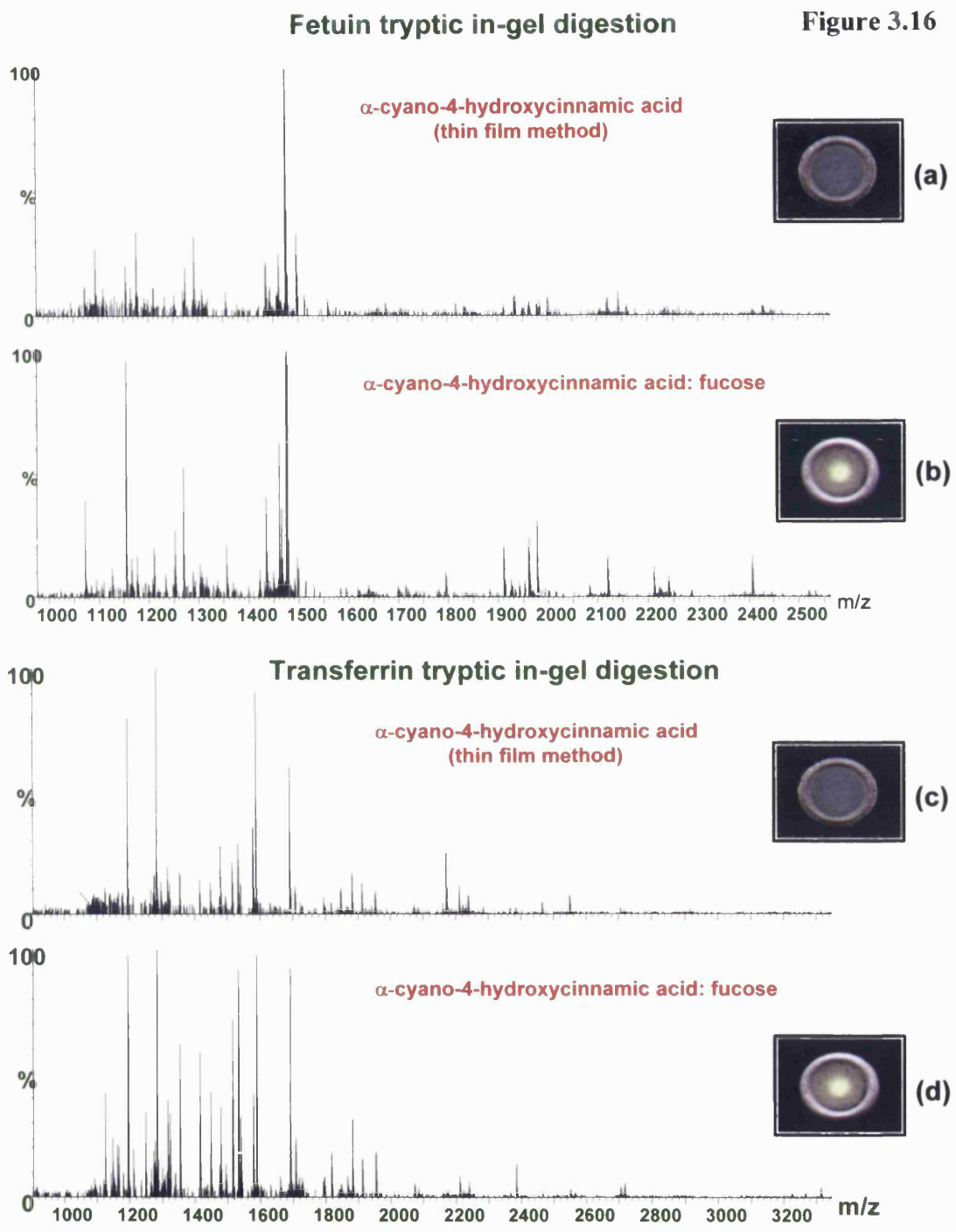


Figure 3.16 Analysis of tryptic peptides of fetuin and transferrin by MALDI TOF MS using;

(a) Thin film matrix method, (b) α-cyano-4-hydroxycinnamic acid: fucose
(c) Thin film matrix method, (d) α-cyano-4-hydroxycinnamic acid: fucose

3.7.2.3 The effect of α–cyano-4-hydroxycinnamic acid/fucose matrix on post-source decay (PSD) studies.

The use of fucose as a co-matrix with α–cyano-4-hydroxycinnamic acid had been shown to be beneficial in the analysis of peptide mixtures in terms of sensitivity and the total amino acid sequence coverage attained. However, if post-source decay (PSD) analyses were to be employed for sequencing of peptides, the possible effect of the incorporation of fucose into the matrix preparation needed to be evaluated. ACTH (clip 18-39, 20 pmol / μl) was analysed using the matrix combinations α–cyano-4-hydroxycinnamic acid and α–cyano-4-hydroxycinnamic acid with fucose, to assess whether any advantages were gained in adding fucose as a co-matrix during sequence studies. Figure 3.17a and b show the results of the PSD of ACTH (clip 18-39) in both matrices. No significant differences in the spectra were observed or any advantages in the level of sensitivity attained. However, significantly increased 'burn times' were observed in the analyses with the matrices containing fucose. The term 'burn time' describes the maximum time available for the analysis of the peptides before the laser 'burned' through a particular layer of crystals and led to a cessation of signal. Hence, the incorporation of fucose as a co-matrix made both PSD and mass mapping studies significantly simpler than analyses in the absence of the fucose. Analysis of peptide mixtures using the thin film technique (Vorm *et al.*, 1994) requires the constant repositioning of the laser position in order to observe peptide ions. This is necessary because of the rapid burn through of the matrix observed (~ every 1-2 sec). However, the incorporation of the fucose as a co-matrix resulted in a stable and constant signal of peptides, for periods greater than 10 sec. Once the peptide signal was observed to be deteriorating, the signal response could be revived by increasing the laser energy to burn through to the next layer of crystals until peptide ions were detected again, thereby avoiding the necessity to reposition the laser.

Figure 3.17

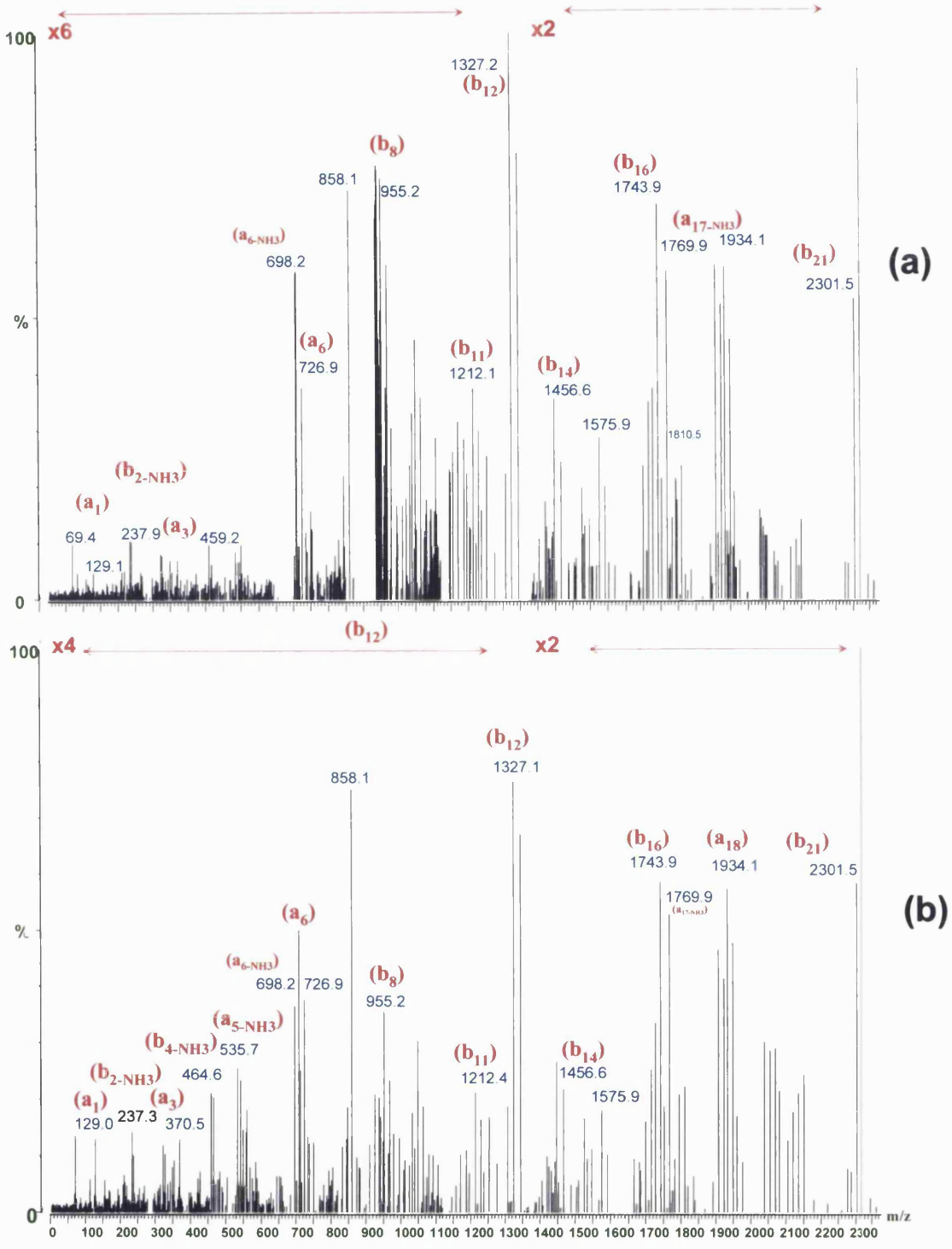
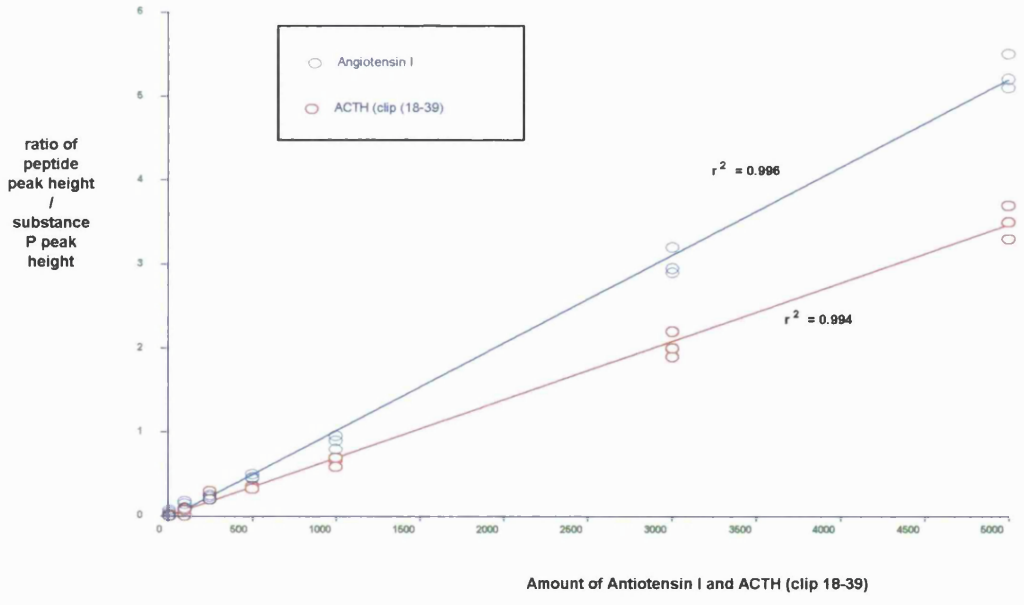


Figure 3.17 Post source decay MALDI TOF MS spectra of ACTH (clip 18-39).
(a) using α-cyano-4-hydroxycinnamic acid matrix.
(b) using α-cyano-4-hydroxycinnamic acid : fucose matrix.

3.7.2.4 The effect of α–cyano-4-hydroxycinnamic acid and fucose on the quantitation of peptides

Gusev *et al.*, (1996) had previously shown that combinations of DHB, ferulic acid, 5-methoxysalicylic acid and fucose could be used as multi-component matrices for the quantitation of peptides using MALDI TOF MS. The main problem in the quantitative analyses using MALDI TOF MS is the shot-to-shot reproducibility, signal degradation and crystal inhomogeneity. Preliminary experiments using fucose in conjunction with α–cyano-4-hydroxycinnamic acid had shown that significant improvements may be achieved using this matrix. Therefore to test the reproducibility and quantitation of peptides using MALDI TOF MS with this matrix, various amounts of angiotensin I and ACTH (clip 18-39, 10 fmol→5 pmol) were analysed with a constant amount of substance P (1 pmol) as an internal standard. The ratio of the peak height of angiotensin I (1296.8 m/z) and ACTH (clip 18-39, 2465.2 m/z) were then plotted against the peak height of substance P (1347.7 m/z). Figure 3.18 shows the graph of a calibration curve of the two peptides analysed using substance P as an internal standard.

Figure 3.18 Calibration curve of angiotensin I and ACTH (clip 18-39) using substance P as an internal standard.



A linear relationship was observed with increasing concentration of each peptide, with correlation coefficients of 0.996 and 0.994 for the peptides angiotensin I and ACTH (clip 18-39), respectively. This indicated that substance P could be used as a suitable internal standard for the approximate quantitation of both peptides. However, it should be noted that for each quantitation a new calibration curve should be performed because responses, and hence the slope of the line, could vary significantly from run to run. These data indicated that α–cyano-4-hydroxycinnamic acid in conjunction with fucose can be used as a co-matrix for quantitation as in the quantitation studies of Gusev and colleagues (1996).

3.7.3 Conclusions

The most beneficial aspect of adding fucose to the original α–cyano-4-hydroxycinnamic acid preparation, was the concentration of the sample into the centre of the target during analyses. This allowed the positioning on the target of a peptide signal or 'sweet spot' quickly and efficiently, without having to 'search' the entire target area for ions. Figure 3.13 shows the spectra and the different crystal formations of each of the matrices evaluated. Each matrix crystallisation can be extremely varied and it often takes several minutes to scan the target surface for an appropriate crystal that produces peptide ions. This was particularly relevant during the analyses carried out using the TOF Spec E, which was fitted with a single axis probe that only allowed the movement of the laser positioning in the y-axis. Thus if crystallisation of the matrix occurred off centre, it was often difficult to locate a signal and in some instances required a degree of manual repositioning of the laser beam.

The reason for the accompanying increase in sensitivity and amino acid sequence coverage obtained during the analysis of peptide mixtures is unclear. This is probably due to fucose selectively precipitating around the edge of the target and drawing with it any salts or contaminants out of the α–cyano-4-hydroxycinnamic acid matrix and improving the MALDI process. Although MALDI TOF MS is often described as being the most salt tolerant of mass spectrometry techniques, it is also apparent that better quality spectra can be obtained in contaminant-free peptide mixtures. However,

it is important to note that the correct concentrations of matrix and fucose must be maintained for this selective precipitation to occur. At concentrations of fucose above and below 50 mM, the α–cyano-4-hydroxycinnamic acid failed to precipitate in the centre of the target but instead crystallised homogenously over the entire surface of the target. At concentrations of α–cyano-4-hydroxycinnamic acid greater than 10 mg/ml, matrix ions with higher mass were detected. This required the matrix suppression to be set higher (800 amu) and limited the mass range of the analysis. At concentrations of α cyano-4-hydroxy cinnamic acid less than 10mg/ml, 'burn times' were significantly decreased. The commercial preparation of α–cyano-4-hydroxycinnamic acid had to be recrystallised prior to mass spectral analysis (see Materials and Methods) for reliable data.

The lowest energy setting to obtain peptide signals, 20%, was obtained using fucose as a co-matrix, which often resulted in increased resolution of the peptides and hence increased mass accuracy for database searching. The use of fucose also permitted the quantitative analysis of peptides as described by Gusev and colleagues (1995, 1996). Although Billeci and Stults (1993) had shown that the matrix combination of fucose and DHB produced glycopeptide ions in the analysis of fetuin tryptic peptides, no such ions were detected when using fucose as a co-matrix with DHB or α–cyano-4-hydroxycinnamic acid in this project.

3.8 Summary

Every step that was considered important in the analysis of proteins separated by PAGE and subsequent peptide mass mapping by tryptic in-gel digestion and MALDI TOF MS, was optimised prior to commencing the analysis of any patient material. The most important findings were that the sensitivity of each analysis could be increased by using *bis*-acrylamide instead of PDA during PAGE and that the removal of contaminants prior to MALDI TOF MS significantly improved the quality of the mass spectra obtained.

Chapter 4

The development of a strategy for the identification of site-specific glycosylation and amino acid substitutions using pure proteins

Chapter 4 - The development of a strategy for the identification of site-specific glycosylation and amino acid substitutions using pure proteins.

Contents

4.1 Introduction.

The principle aim of this work was to develop sensitive and reliable procedures for the investigation of co- and post-translational modifications and amino acid substitutions in serum glycoproteins from patients with genetic disorders and in a therapeutic recombinant protein. This chapter describes the optimisation and evaluation of procedures for the analysis of pure glycoproteins in solution. Two commercially purified human serum glycoproteins, α_1-antitrypsin and transferrin and a recombinant β-glucosylceramidase (Cerezyme) were used as standard reference glycoproteins.

4.1.1 Experimental and theoretical rationale

The experiments described in the previous chapter (chapter III) identified two major problems with MALDI TOF MS analyses of tryptic peptides and glycopeptides generated by tryptic digestion of α_1-antitrypsin and transferrin.

(i) The coverage of the amino acid sequence attained by peptide mass fingerprinting by MALDI TOF MS analysis was less than 50% and 40% for α_1-antitrypsin and transferrin, respectively.

(ii) The lack of detection of molecular masses corresponding to N-glycopeptides generated from both glycoproteins after tryptic digestion.

The failure to cover less than 50% of the amino acid sequence demonstrates clearly serious limitations to the application of this procedure to the diagnosis or screening of patients who express mutations at the protein level. The extent of the coverage of protein sequence following tryptic digestion has been related to the number of peptides containing arginine residues being generated (Krause *et al.*, 1999; Hale *et al.*, 2000). Krause and colleagues (1999) have demonstrated that arginine-containing peptides are detected with a 4-18 fold increase in the intensity compared with lysine-containing peptides during MALDI TOF MS analysis. Therefore digestion with other proteases,

chymotrypsin and V8 (DE) was investigated to see whether the coverage of the amino acid sequence could be improved.

To investigate the N-glycosylation of proteins, N- linked glycans can be released by the action of peptide N-glycanase (PNGase F) or by endoglycosidases and analysed by a variety of techniques (Lejeune *et al.*, 1989; Reason *et al.*, 1991; Packer *et al.*, Charlwood *et al.*, 1998; Jachymek *et al.*, 1999). Although this approach can identify the types and relative proportions of glycans, it does not assign the glycans to the possible glycosylation sites. This can be achieved by proteolytic digestion of the glycoprotein, separation of the resultant mixture of peptides and glycopeptides followed by the release and analysis of the glycans from each glycopeptide (Stahl *et al.*, 1994; Rahbek-Nielson, 1997; Miele *et al.*, 1997; Schmitt *et al.*, 1999). This is both a time-consuming and difficult procedure. Therefore an alternative strategy was developed for identifying the occupancy and type of N-linked glycan at each potential glycosylation site which does not necessitate separation of glycopeptides after proteolysis.

It was found that an array of digestion protocols using 3 proteases and specific endoglycosidase/peptide N-glycanases, all glycosylation sites in the 3 standard proteins 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 non-glycosylated 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 (Tarentino and Plummer, 1994). 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 the deglycosylated peptide will be 203.1 m/z greater than that observed from the same non-glycosylated peptide (or 349.2 m/z greater with a core fucose). Complex glycans will not be released and hence a peptide mass will not be detected. Release of a complex glycan with endoglycosidase F3 will leave a single N-acetylglucosamine or a fucosylated N-acetylglucosamine attached to

the asparagine, depending on whether the glycan is core α–6-fucosylated or not. The m/z of such a peptide will also be 203.1 (or 349.2 with fucose) greater than that derived for the same non-glycosylated peptide. A summary of the specificities and the actions of the different classes of N-linked glycans by the various endo- and exoglycosidases used in this study is shown in Figure 4.01. The theoretical changes in the masses of the peptides and glycopeptides that would be expected are summarised in Table 4.1.

The cleavage of the glycan to leave a single N-acetylglucosamine residue (or a fucosylated N-acetylglucosamine), removes a significant portion of the glycan structure from the glycopeptide and would theoretically, result in a smaller glycopeptide less susceptible to metastable decay during the ionisation process as seen in uv-MALDI TOF MS. In effect the glycopeptide behaves more like a peptide, increasing the sensitivity of the analysis. Figure 4.02 is a schematic protocol for determining the site-specific glycosylation of N-linked glycoproteins using endoglycosidases and glycanases.

Figure 4.01

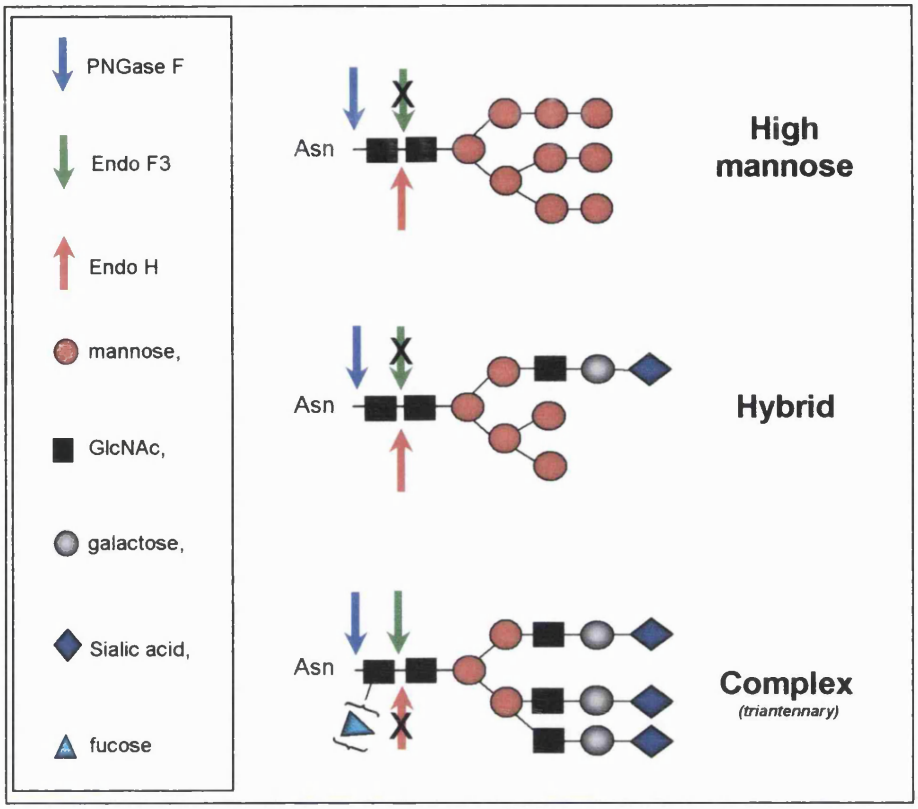


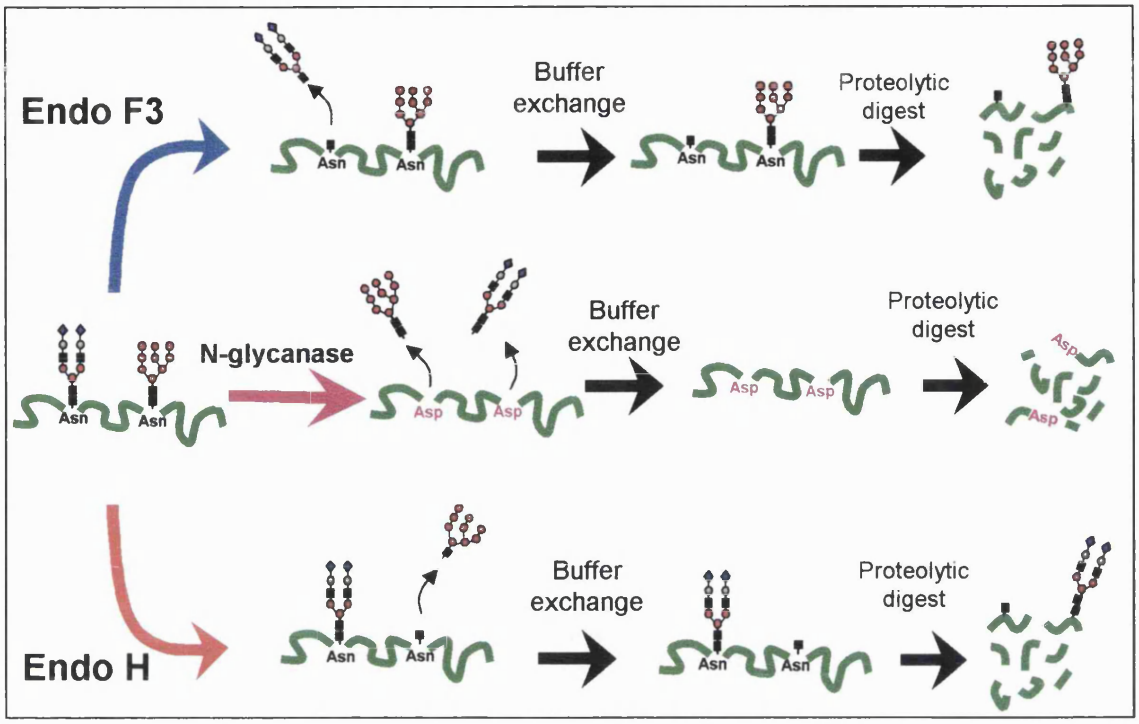
Figure 4.01 **Diagram showing the specificities of the endoglycosidases H and F3, and PNGase F on the three classes of N-linked glycans investigated in this study.**

Table 4.1

Digestion	Occupied		Unoccupied
	Complex	High Mannose/ Hybrid	
Protease alone	-	-	**Asparagine**
Protease and PNGase F	**Aspartic acid (+1Da)**	**Aspartic acid (+1Da)**	**Asparagine**
Protease and Endo F3	**GlcNAc(±Fuc) (203.1 ±146.1 Da)**	-	**Asparagine**
Protease and Endo H	-	**GlcNAc(±Fuc) (203.1 ±146.1 Da)**	**Asparagine**

Table 4.1 **Summary of the theoretical changes in the masses of the peptides and glycopeptides expected after digestion with glycosidases and proteases.**

Figure 4.02



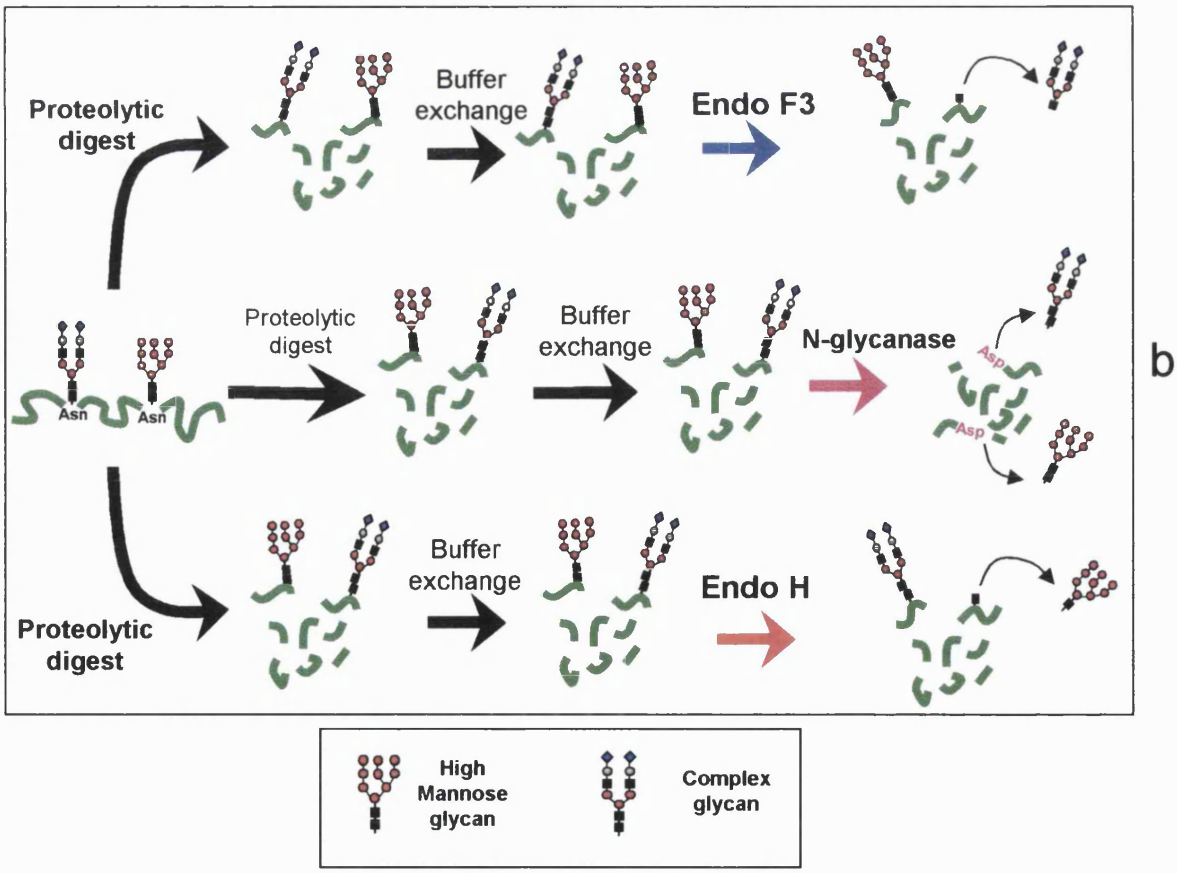


Figure 4.02 **Strategy employed to deduce site-specific glycosylation of glycoproteins using proteolytic and glycosidase digestions with N-glycanase, Endo F3 & Endo H and analysis by MALDI TOF MS.**

(a) The use of exo / endoglycosidases followed by proteolysis.

(b) The use of proteases followed by exo / endoglycosidases.

4.2 Materials and methods.

To test the analytical strategy described, three commercially purified glycoproteins whose structures and patterns of glycosylation are well established, were analysed. The two human serum glycoproteins α_1-antitrypsin and transferrin contain complex glycans and are known to be affected in human diseases. The third protein, recombinant β-glucosylceramidase or Cerezyme was chosen because not all of its potential glycosylation sites are occupied and all the glycans are of the high mannose type. Cerezyme is also known to contain a single amino acid substitution near its C-terminus that is not present in the native form of the human protein.

4.2.1 α_1-Antitrypsin

The amino acid sequence and glycosylation sites of α_1-antitrypsin are shown in Figure 4.03a. Alpha-1-antitrypsin has three glycosylation sites at asparagine residues 46, 83 and 247, which are occupied predominantly by a mixture of fucosylated and non-fucosylated biantennary N-linked glycans. However, some fucosylated and non-fucosylated triantennary structures are also present (for a more extensive description see Chapter I).

4.2.2 Transferrin

The amino acid sequence and glycosylation sites of transferrin are shown in Figure 4.03b. Transferrin has two N-linked glycosylation sites at asparagines 413 and 611, which are occupied predominantly by fucosylated and non-fucosylated biantennary glycans. For a more extensive description on the structure of transferrin see Chapter 1.

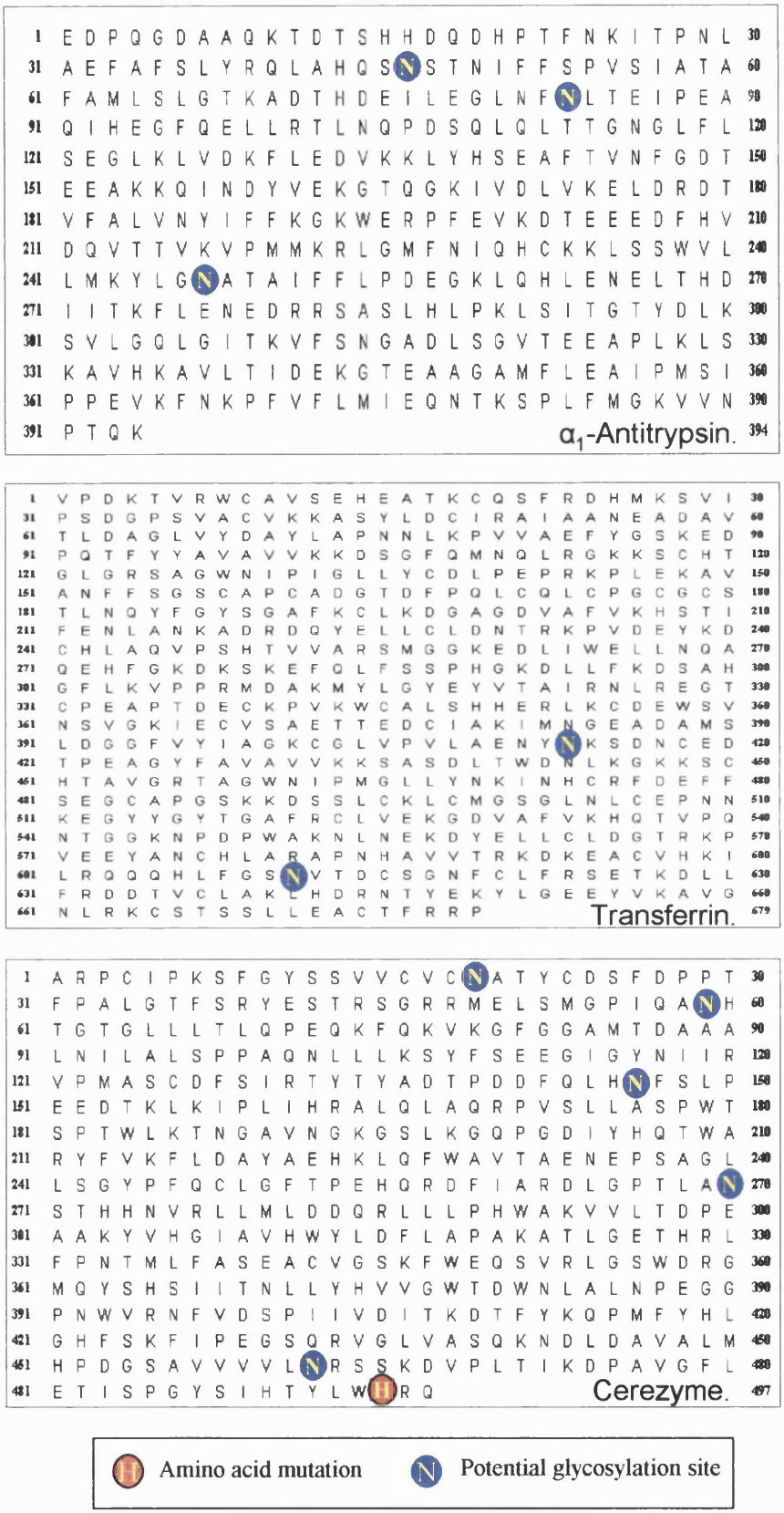


Figure 4.03 The amino acid sequence and the glycosylation sites of the glycoproteins used in this study.

(a) Amino acid sequence of α_1-antitrypsin.
(b) Amino acid sequence of transferrin.
(c) Amino acid sequence of Cerezyme.

4.2.3 β-glucosylceramidase

Recombinant β-glucosylceramidase (Cerezyme) contains five potential glycosylation sites at asparagines 19, 59, 146, 270 and 462 (Figure 4.03c). However, glycosylation site 462 is not occupied by a glycan. The glycan structures present on the molecule are of a mannose-terminated or high-mannose type which are susceptible to removal by endoglycosidase H but not by endoglycosidase F3 (Figure 4.04). Cerezyme also differs from the native human form of β-glucosylceramidase (Figure 4.03b) in that arginine 495 has been replaced by a histidine residue. This substitution is interesting for testing the capability of the procedure to detect amino acid substitutions.

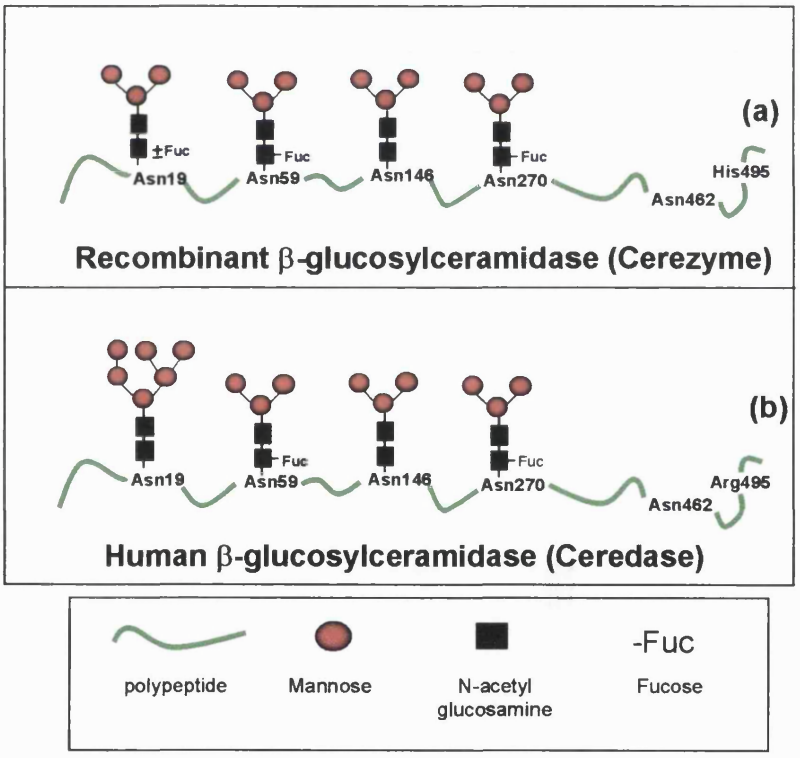


Figure 4.04. Schematic representation of the micro- and macroheterogeneity of Cerezyme and Ceredase.

4.3 Results.

4.3.1 Transferrin (Tables 4.2a and 4.3a).

4.3.1.1 Proteolytic digestion.

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 (DE), respectively. Using the three proteases 469 out of the total 679 (69.1%) amino acids in the sequence were detected (Table 4.2a). Different regions of the core polypeptide 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.

4.3.1.2 Proteolysis after predigestion with endoglycosidase F3

Transferrin was pre-digested with endoglycosidase F3 before digestion with the 3 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 observed 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, Table 4.2a).

4.3.1.3 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% (Table 4.2a).

Table 4.2 **Fraction of amino acid sequence detected by MALDI TOF MS after different digestions**

Summary of results for transferrin (Table 4.2a)

Transferrin	Protease coverage (%)			Complete coverage of the protein (%)
	trypsin	chymotrypsin	V8 (DE)	
With glycans intact	331 / 679 48.7%	328 / 679 48.3%	288 / 679 42.4%	469/679 69.1%
After Endo F_3 then protease	145/ 679 21.4%	148 / 679 21.8%	213 / 679 31.4%	387 / 679 57.0%
After protease then Endo F_3	331 / 679 48.7%	328 / 679 48.3%	288 / 679 42.4%	469/679 69.1%
After PNGase F then protease	407 / 679 59.9%	384 / 679 56.6%	283 / 679 41.7%	604 / 679 88.9%
After protease then PNGase F	360 / 679 53.0%	375 / 679 55.2%	301 / 679 44.3%	519 / 679 76.4%

Summary of results for α1-antitrypsin (Table 4.2b)

α1-antitrypsin	Protease coverage (%)			Complete coverage of the protein (%)
	trypsin	chymotrypsin	V8 (DE)	
With glycans intact	146 / 394 37.1%	139 / 394 35.3%	186 / 679 47.2%	284 / 394 71.8%
After Endo F_3 then protease	236 / 394 59.9%	228 / 394 57.9%	235 / 679 59.6%	353 / 394 89.6%
After protease then Endo F_3	161 / 394 40.9%	164 / 394 41.6%	198 / 394 50.3%	330 / 394 83.8%
After PNGase F then protease	173 / 394 43.9%	248 / 394 62.9%	125 / 394 31.7%	341 / 394 86.5%
After protease then PNGase F	146 / 394 37.1%	139/ 394 35.3%	186 / 394 47.2%	284 / 394 71.8%

Summary of results for Cerezyme® (Table 4.2c)

Cerezyme®	Protease coverage (%)			Complete coverage of the protein (%)
	Trypsin	chymotrypsin	V8 (DE)	
With glycans intact	324 / 497 65.2%	306 / 497 61.6%	259 / 497 52.1%	418 / 497 84.1%
After Endo H then protease	157 / 497 31.6%	350 / 497 70.4%	230 / 497 46.3%	444 / 497 89.3%
After protease then Endo H	392 / 497 78.9%	333 / 497 67.0%	294 / 497 59.2%	479 / 497 96.4%
After PNGase F then protease	277 / 497 55.7%	335 / 497 67.4%	285 / 497 57.3%	440 / 497 88.5%
After protease then PNGase F	338 / 497 68.0%	336 / 497 67.6%	258 / 497 51.9%	435 / 497 87.5%

Table 4.3a. 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.			

Table 4.3b Summary for α_1-antitrypsin 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. α_1-antitrypsin glycosylation sites are 46, 83 and 247.			

Table 4.3c. 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. Cerezyme© glycosylation sites are 19, 59, 146 and 270.					

Protease used	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)	$X_{19}X_{59}X_{146}X_{270}$	$X_{19}X_{59}X_{146}X_{270}$

Key:

$\checkmark_n$	-indicates a peptide which covered a sequon was observed
X_n	-indicates 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	

4.3.1.4 Proteolysis after predigestion with PNGase F.

The glycosylation site, asparagine 611, occurs in the tryptic peptide containing amino acids 603-623. The theoretical m/z of this glycopeptide is 4605.9 when a complex biantennary glycan is attached and 2498.1 m/z if this sequon is not glycosylated. No peptide with an m/z of 2498.1 was detected by MALDI-TOF analysis after tryptic digestion of transferrin (Figure 4.05a) but when the transferrin was pre-digested with PNGase F and then digested with trypsin a peptide of 2499.1 m/z was observed (Figure 4.05b). 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 (DE) produced two peptides that covered the first glycosylation site, asparagine 413, with masses 1259.5 and 2572.3 m/z corresponding to amino acids 411-420 and 393-416 +1 Da, respectively. Chymotryptic digestion of transferrin, that had been predigested with PNGase F, produced masses of 1747.7 and 1725.7 m/z , which correspond to the amino acid sequences, 413-427 and 608-622 +1 Da, containing the two glycosylation asparagines, 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).

4.3.1.5 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 (DE) 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, Table 4.2).

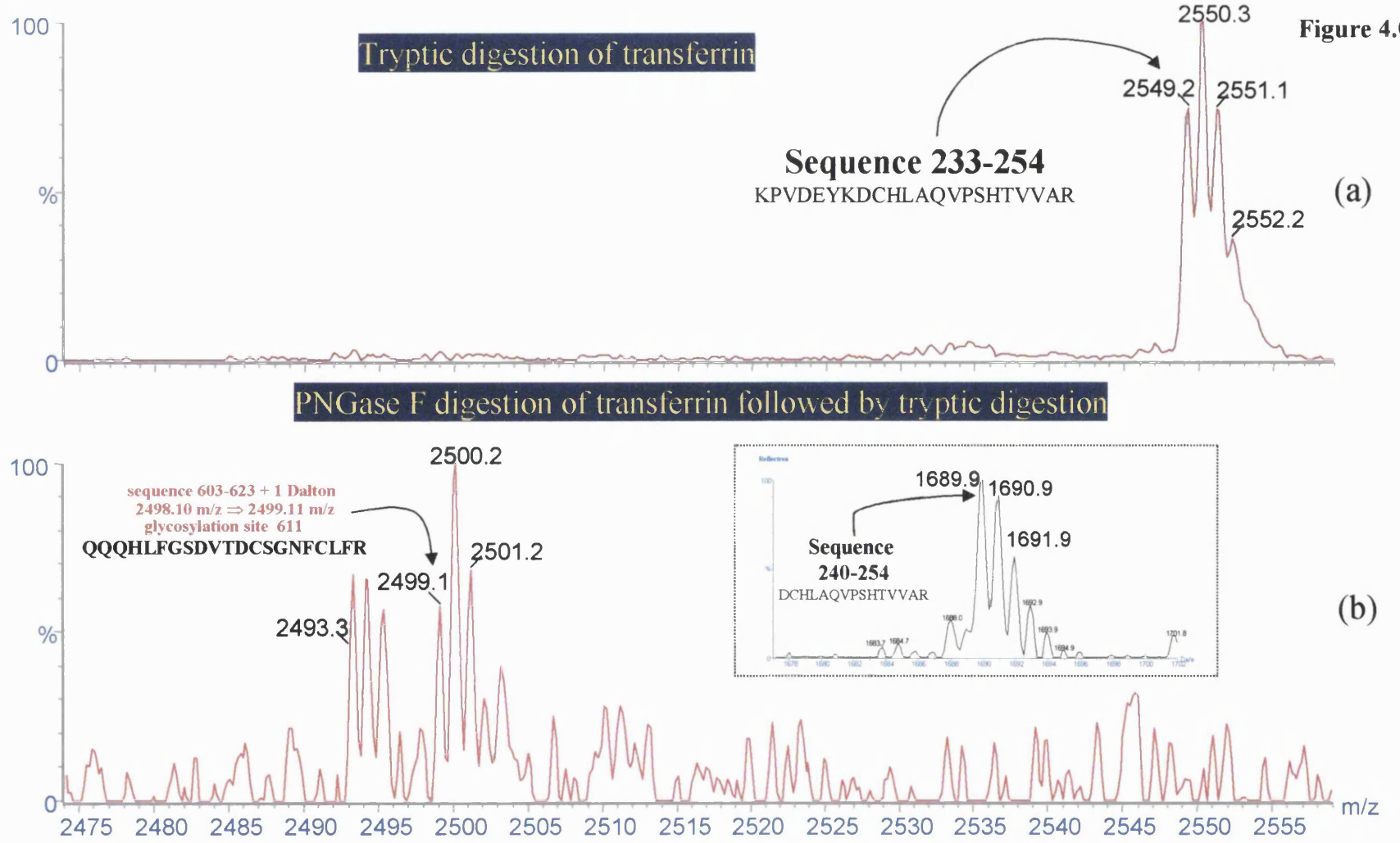


Figure 4.05 Mass spectra showing changes in the peptide masses observed of human transferrin (a) before and (b) after enzymatic removal of the glycans by PNGase F, focusing on peptide 603-623 which contains glycosylation site 611 and peptide 233/240-254. Removal of the glycan structure results in the appearance of a new peptide covering the amino acid sequence 603-623 with an aspartic acid residue at the glycosylation motif. In addition peptide 233-254 has disappeared and is replaced by a peptide of mass 1689.9 m/z, corresponding to the sequence generated by cleavage at a missed cleavage site in peptide 233-254.

4.3.2 α_1-Antitrypsin (Tables 4.2b and 4.3b)

4.3.2.1 Proteolytic digestion alone.

No glycopeptides were detected in the mass spectrum after α_1-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 (DE) covered 37.1%, 35.3% and 47.2% of the amino acid sequence, respectively (Table 4.2b, Figure 4.07a).

4.3.2.2 Proteolysis after predigestion with endoglycosidase F3

A glycopeptide with mass 1958.9 m/z corresponding to the amino acid sequence 244-259 + one N-acetylglucosaminyl unit was detected after tryptic digestion of the deglycosylated protein (Figure 4.06b). It was not detected after digestion with trypsin alone (Figure 4.06a). Such truncated glycopeptides were not detected for the other two glycosylation sites. Glycopeptides with an N-acetylglucosaminyl residue attached to asparagine corresponding to all 3 glycosylation sites were detected when α_1-antitrypsin was pre-digested with endoglycosidase F3 prior to chymotryptic digestion. No truncated glycopeptides were detected after successive digestion with endoglycosidase F3 and protease V8 (DE). 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 4.2b, Figure 4.07b).

4.3.2.3 Digestion with endoglycosidase F3 after proteolysis.

The use of this digestion protocol 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 α_1-antitrypsin resulted in truncated glycopeptides

being detected that covered asparagine 247 only (Figure 4.06). Chymotryptic digestion produced truncated glycopeptides that covered the asparagines 46 and 247 but not 83, whilst digestion with protease V8 (DE) produced truncated glycopeptides containing asparagine 83 only. 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 4.2b,).

4.3.2.4 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 N-glycosylation sequons of the protein. Proteolytic digestion with trypsin 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 (DE) 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 3 proteases increased to 86.5% (341 / 394) (Table 4.2b, Figure 4.07c).

4.3.2.5 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 4.2b).

Figure 4.06

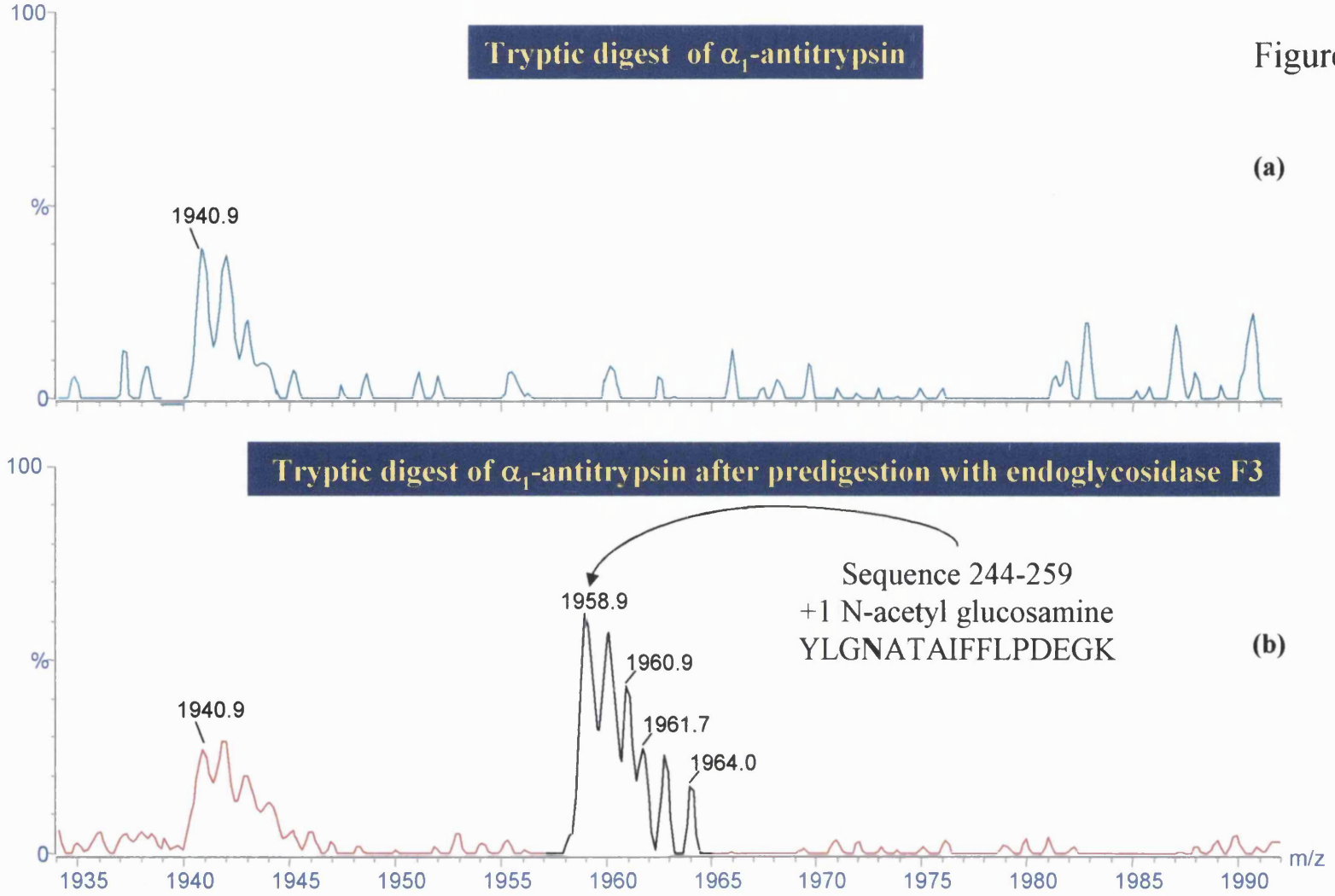


Figure 4.06 Mass spectra of a tryptic digestion of α_1-antitrypsin before (a) and after (b) enzymatic truncation of the glycans by endo F3. The mass 1958.9 m/z corresponds to the peptide sequence 244-259 + 1 GlcNAc residue (+203.1 Da).

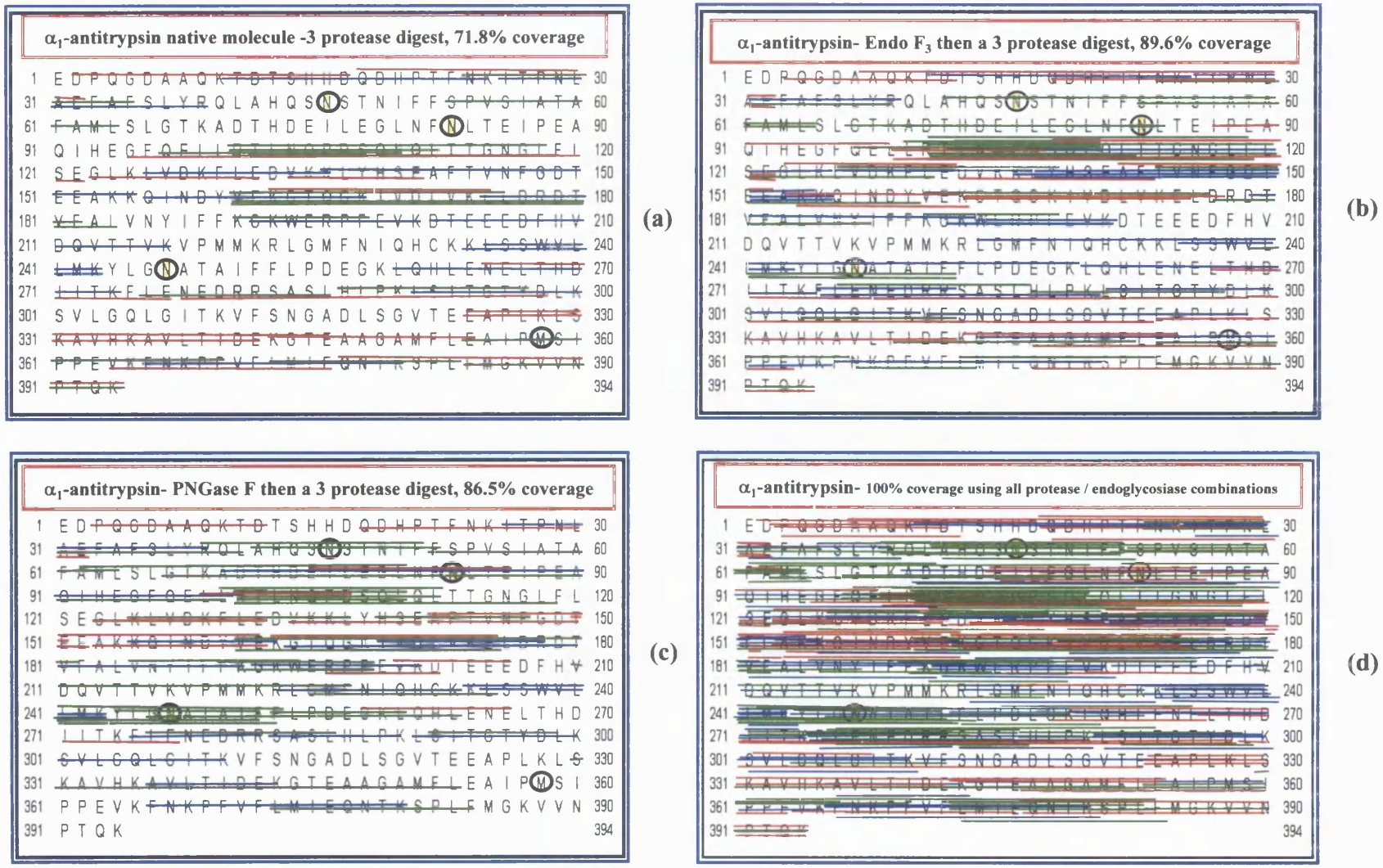


Figure 4.07 Distribution of peptides over the amino acid sequence of α_1-antitrypsin after digestion by different proteases.

4.3.3 β-glucosylceramidase (Tables 4.2c and 4.3c).

4.3.3.1 Proteolytic digestion alone

Masses corresponding to the glycopeptides with $Man_3GlcNAc_2$- and $Man_3GlcNAc_2$-Fuc attached to asparagines 19 and 270, respectively, were observed after tryptic digestion of human placental β-glucosylceramidase (Ceredase). 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. No glycopeptides were however detected after digestion with the other two proteases.

The mass spectra of the glycopeptides from the tryptic digest of the recombinant protein Cerezyme and the human placental form of β-glucosylceramidase, Ceredase, are shown in Figure 4.08a and b, respectively. In both analyses a mass of 4501.9 m/z was observed which corresponds to a glycopeptide covering the amino acid sequence 8-39 and containing a $Man_3GlcNAc_2$ glycan (Figure 4.08a and b). However, two extra masses of 4664.3 m/z and 4826.0 m/z were detected in the Ceredase but not the Cerezyme analyses. These correspond to glycopeptides covering the same sequence of amino acids but containing $Man_4GlcNAc_2$ and $Man_6GlcNAc_2$ glycans respectively (Figure 4.08b). It is concluded that the recombinant form of β-glucosylceramidase (Cerezyme) contains predominantly a $Man_3GlcNAc_2$ glycan at asparagine 19 whereas Ceredase has more heterogeneous glycosylation as has been reported by Edmunds and Zhang (1999; http://www.cerezyme.com/global/pi.pdf). The mass 4664.3 m/z which corresponds to a $Man_4GlcNAc_2$ structure was not reported previously by Edmunds and Zhang (1999).

Several peptides, covering the potential glycosylation site, asparagine 462, were detected and all contained an unmodified asparagine residue, indicating that it was not glycosylated. Collectively from the results obtained from experiments using all 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 glycosylation sites 59 and 146 (Figure 4.09).

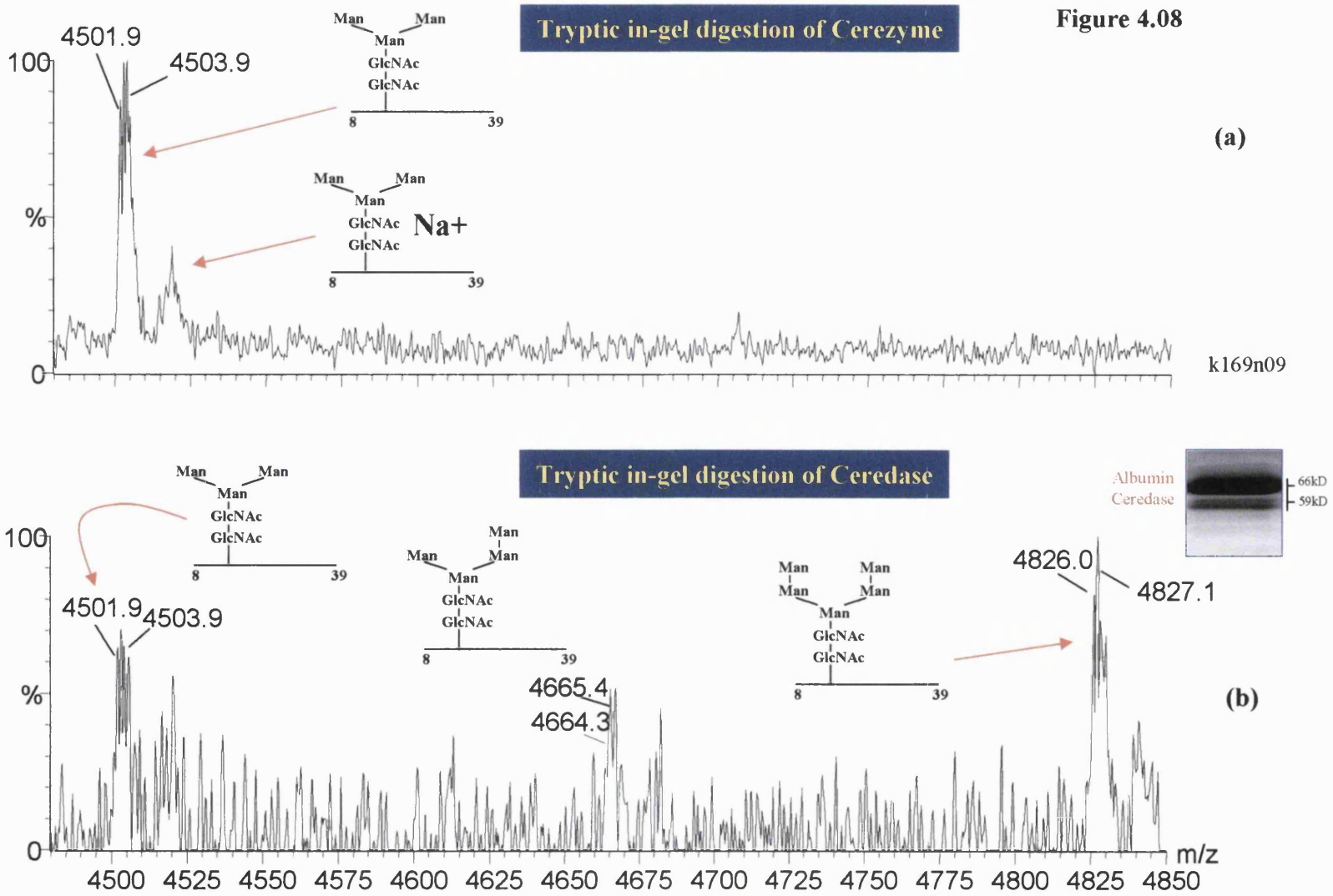


Figure 4.08 Mass spectra of the glycopeptides obtained after the in-gel tryptic digestion of (a) Cerezyme and (b) Ceredase.

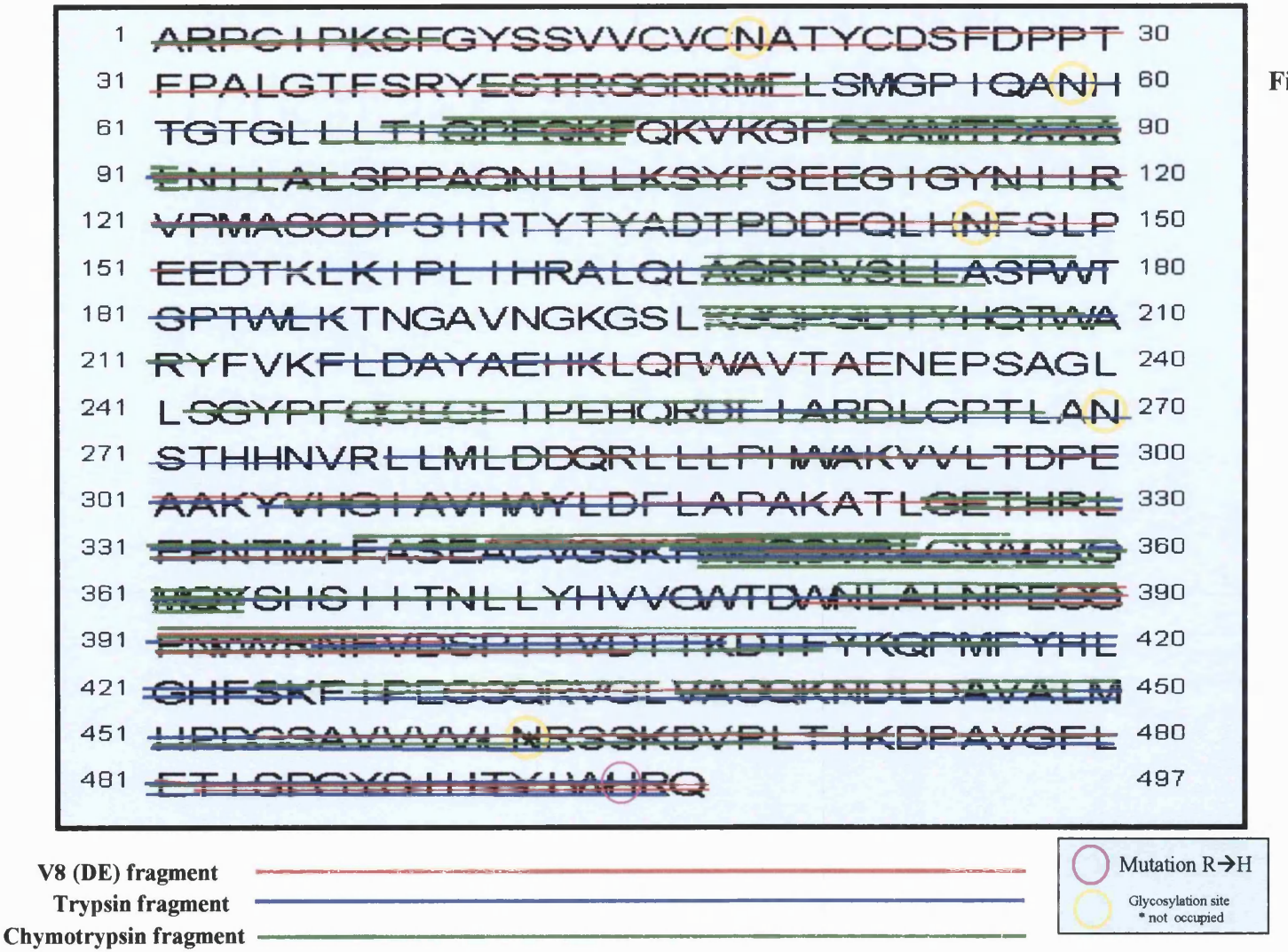


Figure 4.09 Distribution of peptides showing the combined amino acid sequence coverage of Cerezyme after digestion by different protocols (96%).

Figure 4.09

4.3.3.2 Detection of the substitution, arginine 495 to histidine, in Cerezyme

Comparison of the tryptic digestions of the human placental (Ceredase) and recombinant (Cerezyme) enzymes confirmed that arginine 495 in the placental enzyme is replaced by a histidine residue in the recombinant form (Figure 4.10). 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. Figure 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 m/z covering amino acids 474-495 is observed with Ceredase (Figure 4.10a) but not in Cerezyme (Figure 4.10b). Conversely, a peptide of 2659.3 m/z, is present in Cerezyme (Figure 4.10b) but is absent in Ceredase (Figure 4.10a). This peptide corresponds to amino acids 474-496 in the sequence of Cerezyme, in which arginine 495 has been substituted by histidine with the loss of a tryptic cleavage site at arginine 495-arginine 496, 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.

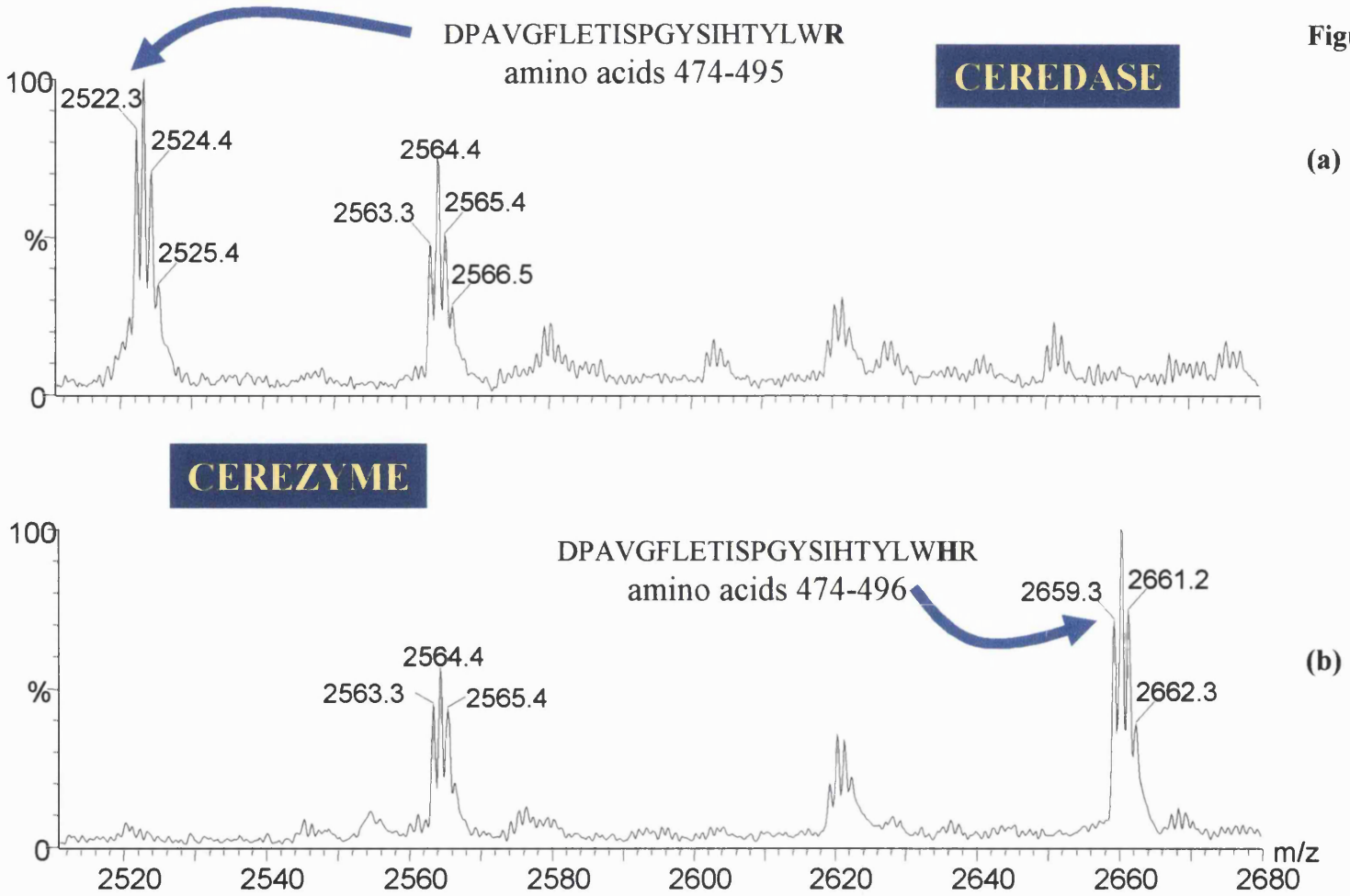


Figure 4.10 Mass spectra of the in-gel tryptic digestion of purified (a) Ceredase and (b) Cerezyme, confirming the substitution of an arginine to a histidine at residue 495 in the amino acid sequence in Cerezyme.

4.3.3.3 Proteolysis after predigestion with PNGase F.

The use of PNGase F to remove the glycans prior to proteolysis confirmed that asparagine residues 59, 146 and 270 were glycosylated. The peptides covering these sites all contained aspartic acid residues and were obtained with all 3 proteases (Table 4.3c). No masses of peptides containing asparagine residue 19 were detected following proteolysis with any of the proteases. This digestion protocol increased coverage of the amino acid sequence to 88.5 % (440 / 497 amino acids, Table 4.2c).

4.3.3.4 Proteolysis after predigestion with endoglycosidase H.

The use of this digestion protocol revealed glycopeptides with N-acetylglucosamine attached to asparagine residues 19, 146 and 270 (Table 4.3c). Peptides, corresponding to amino acids 263-277 with N-acetylglucosamine (1834.9 m/z) or fucosyl N-acetylglucosamine (1981.0 m/z) attached to asparagine 270, were detected in the tryptic digests (Figure 4.11). The masses of peptides containing N-acetylglucosamine structures at glycosylation sites 19 and, 146 and 270, were also detected in the chymotryptic and V8 (DE) digests, respectively (Table 4.3c). 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 arginine 495 to histidine was not detected by this protocol.

4.3.3.5 Proteolysis followed by digestion with endoglycosidase H.

Information about all four glycosylation sites was obtained with this digestion protocol (Table 4.3c). Peptides containing asparagine residues with N-acetylglucosamine and peptides containing fucosyl N-acetylglucosamine moieties attached were observed for all four N-glycosylation sites, showing that the microheterogeneity observed at the asparagine residue site 270, occurs at all four N-glycosylation sites. The overall coverage of the amino acid sequence increased to 96.4% (479 / 497 amino acids, Table 4.2c).

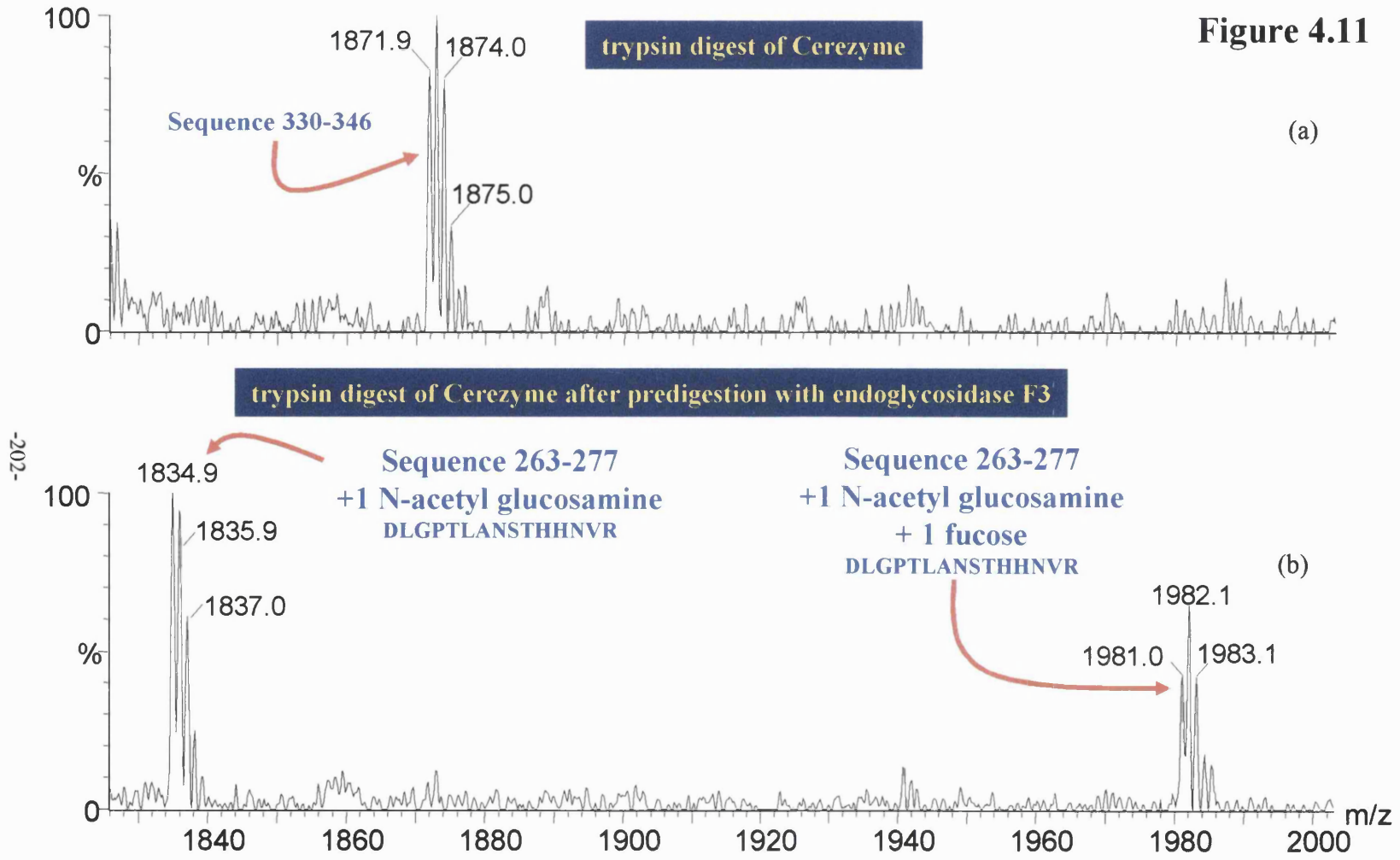


Figure 4.11 Tryptic digestion of Cerezyme (a) without and (b) with prior digestion with endoglycosidase H. The mass 1834.9 m/z corresponds to the peptide sequence 263-277 + 1 GlcNAc residue (+203.1 Da) and the mass 1981.0 m/z corresponds To the same peptide but containing one GlcNAc and one fucose residue (+ 349.2 Da).

4.3.3.6 Proteolysis followed by PNGase F digestion

The use of PNGase F after proteolysis with the 3 proteases, only allowed detection of peptides covering two of the occupied asparagine residues, 146 and 270 (Table 4.3c). 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 containing the glycosylation site at asparagine residue 59 was not detected using this protocol.

In summary the array of digestion protocols used showed that four (asparagine16, 59, 146 and 270) out of five of the potential glycosylation sites in Cerezyme are occupied and that the fifth site, asparagine 462, is not glycosylated (Figure 4.09). The glycans at all four sites are either core α1-6 fucosylated or non-fucosylated forms. Truncated oligomannose glycans, $Man_3GlcNAc_2$- and $Man_3GlcNAc_2$-Fuc, were detected directly at asparagines residues, 19 and 270, respectively, and it is probable that a mixture of both structures occurs at all four glycosylation sites.

4.4 Discussion

The aim of the work described in this chapter was to develop a rapid and sensitive technique for identifying site-specific glycosylation and amino acid substitutions in glycoproteins. The strategy of employing several digestion protocols demonstrated a potential use of this method for the identification of the micro- and macroheterogeneity of the glycosylation of glycoproteins. Appropriate manipulation in the proteolytic digestion before and after digestion with endo / exoglycosidases of the glycoprotein appears to have significant importance, since it generates potential peptide / glycopeptide fragments of 750-2500 m/z mass range. This in turn improves the capability of the MALDI TOF MS in terms of sensitivity, mass accuracy and mass resolution.

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 3 proteins studied. Over 99% of the amino acid sequence of human β-glucosylceramidase and α_1-antitrypsin could be observed using different permutations of the proteases and deglycosylating enzymes (Figure 4.07). 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 were not compatible with analysis by MALDI TOF MS. It is not clear why some peptides are not suitable for MALDI TOF MS analysis but it is likely that this is due to either poor ionisation or signal suppression by co-analytes.

Although DHB was the better matrix for analysis of intact neutral glycopeptides as observed in the analysis of Cerezyme, the glycopeptides tended to be too large (>3000 Da/e) for optimal analysis using the MALDI TOF MS operating in reflectron mode. No intact sialic acid-containing glycopeptides from transferrin or α_1-antitrypsin were detected using α-cyano, DHB or THAP: ammonium citrate matrices. 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 a signal.

The proportion of the amino acid sequence detected after proteolysis alone varied with the three proteases and for the different glycoproteins (Table 4.3a-c). The complete sequence was never observed without deglycosylation and the only glycopeptides detected directly by proteolysis alone were those containing neutral truncated glycans on asparagines 19 and 270 of β-glucosylceramidase. When the peptides produced by proteolysis were deglycosylated (Table 4.3a-c), there was little or no difference in coverage compared to proteolysis alone, although in some cases deglycosylated peptides became more apparent (Table 4.2a-c). When the proteins were deglycosylated prior to proteolysis there was, 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 4.2 and Figure 4.07). 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 increased the assay sensitivity but might also have implications for glycan analysis, because PNGase F is often used to release the glycans for analysis after proteolysis. The present observations indicate that enzymatic release of glycans by PNGase F was more efficient on the intact denatured proteins. It has been demonstrated that PNGase F can release glycans from short peptides but detailed information on the rate of release is not available (Fan and Lee, 1997). In contrast endoglycosidases F3 and H tend to cleave the glycans more efficiently after proteolysis, indicating a preference for peptides over polypeptides

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 (Figure 4.05a) but not after predigestion with PNGase F. However, a peptide with m/z of 1689.9 (Figure 4.05b), 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 produced smaller peptides (Figure 4.07c) and a more even distribution of peptides over the protein sequence (Figure 4.07b) compared to proteolysis alone or proteolysis followed by deglycosylation. This finding 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 more amenable to analysis.

These results indicate that intact glycans on glycoproteins provided some protection against proteolysis around the sequons and at other sites (Figure 4.07a). The truncated glycans generated by the endoglycosidases still seemed to affect proteolysis as incomplete digestion is obtained in their presence (Figure 4.07b). These observations support the theory that one of the functions of glycosylation is protection against proteolysis *in vivo* (Varki, 1993). The presence of glycans tends to direct proteolysis to

particular regions of the molecule. 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 analysis 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 by changes in the pattern of proteolytic digestion (Figure 4.10a and b).

The sites of glycosylation were 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 was not occupied (see section 4.3.3) then the predicted mass of the unmodified amino acid sequence of the peptide was observed.

In summary procedures developed in the present work provide 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 if 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.

Chapter 5

The identification of α_1-antitrypsin variants using proteomic technology

Chapter 5 - The identification of α_1-antitrypsin variants using proteomic technology.

Contents

5.0 The identification of α_1-antitrypsin variants using proteomic technology.

5.1 Genetic variation in the primary structure of α_1-antitrypsin and human disease.

Following the evidence on the association between a lack of a α_1-globulin fraction or $\alpha_{1\text{-}3,5}$-glycoprotein in human serum and early onset of emphysema by Laurell and Erikson (1963), Fagerhol and Laurell (1967) developed a starch gel electrophoretic method for the detection of α_1-antitrypsin variants in human sera. The initial variants detected using this method were designated by a letter depending on their mobility during electrophoresis i.e. F (fast), M (medium), S (slow) and Z (the most cathodal) (Fagerhol, 1967). However, the use of starch gel electrophoresis was eventually superseded by high resolution isoelectric focusing and a new nomenclature was introduced to classify each variant or polymorphism (Cox *et al.*, 1980). The position of α_1-antitrypsin from normal controls after IEF is designated M. Variants of α_1-antitrypsin with amino acid substitutions that decrease the number of positively charged residues (and/or increase the number of negatively charged residues) migrate closer to the anode and are designated by the letters A-L. For example, protease inhibitor (PI) type I has the amino acid substitution Arg39Cys which has one less positive charge than PI M. Variants of α_1-antitrypsin with amino acid substitutions that increase the number of positively charged residues (and/or decrease the number of negatively charged residues) migrate closer to the cathode and are designated by the letters N-Z. For example, PI type Z has the amino acid substitution Glu342Lys which has one less negative charge and one more positive charge than PI M (Cox *et al.*, 1980).

The allele which codes for a particular protein variant is designated in italics and with an asterisk e.g. *PI*Z* encodes the protein with the Glu342Lys amino acid change (PI Z). The nomenclature has been extended for rare variants by including the birthplace of the first described patient with that allele e.g. $PI^*Z_{Bristol}$ (Cox, 2001). An allele which produces no detectable α_1-antitrypsin is designated *PI*null* (Cox, 2001). The two alleles for α_1-antitrypsin are co-

dominant. Thus an individual with *PI*M* and *PI*Z* alleles will produce both M and Z types of α_1-antitrypsin (both bands on the gel) and is described as having the PI MZ phenotype. If only one band is present the phenotype can be referred to as PI M or PI Z (unless genotyping or the family tree allows the phenotype to be described more accurately e.g. PI ZZ or PI Z.

The above description suggests that one α_1-antitrypsin allele gives rise to a single protein i.e. one band on isoelectric focussing. However this is far from the truth. α_1-Antitrypsin has three glycosylation sites and these can be occupied by glycans with 2 or 3 negatively charged sialic acid residues. Thus an individual with the *PI*M / PI*M* genotype has two major bands on an isoelectric focusing gel - one has three biantennary glycans (6 sialic acid residues) the other has two biantennary glycans and one triantennary glycan (7 sialic acid residues) (Cox, 2001).

The nomenclature of different types of α_1-antitrypsin has become somewhat confused by the use of terms like PI M2 and isoform M2. PI M2 is used by some authors to describe an α_1-antitrypsin molecule with an amino acid change that does not alter electrophoretic mobility (Arg101His) or affect function (a polymorphism) (Cox, 2001). Isoform M2 is an α_1-antitrypsin isoform (glycoform) with the usual PI M amino acid sequence (sometimes referred to as PI M1) but with two triantennary and one biantennary glycans (8 sialic acid residues). Furthermore, the known isoforms have been labelled as M2, M4, M6, M7 and M8 (Fagerhol and Laurell, 1967). To avoid confusion with other work in the literature, this convention is retained.

Alpha$_1$-antitrypsin is highly polymorphic with over 60 variants described (Cox, 2001). However, not all polymorphic changes are disease-causing. The commonest non-disease-causing variant is a neutral amino acid substitution, resulting from of the replacement of a valine residue by an alanine, at position 213 in the sequence. This polymorphic variant is sometimes referred to as PI M1ala and constitutes as much as 34% of total α_1-antitrypsin producing a band in the PI M region (Nukiwa *et al.*, 1987).

5.1.1 Molecular basis of α_1-antitrypsin deficiencies

Mutations that lead to reduced levels of plasma α_1-antitrypsin become disease-causing when the levels of α_1-antitrypsin fall below 20% of normal (Cox, 2001). The commonest disease-causing mutation that results in a deficiency in plasma α_1-antitrypsin is the *PI*Z* mutation (Cox, 2001). The *PI*Z* mutation causes the substitution of a lysine for a glutamic acid residue at amino acid 342 in the sequence (E342K) (Jeppsson, 1976; Yoshida *et al.*, 1976). The phenotype caused by *PI*Z/PI*Z* homozygosity is commonly referred to as PI Z (because of cathodal position of the PI Z α_1-antitrypsin obtained after isoelectric focusing, as discussed earlier). In this phenotype, the concentration of plasma α_1-antitrypsin is reduced to about 15-20% of normal levels (Svenger, 1985). This deficiency of α_1-antitrypsin in the circulation leads to a fall in the protease / inhibitor ratio in the lung below a critical level and predisposes the patient to the development of emphysema as early as the fourth decade of life. However, patients with PI Z α_1-antitrypsin deficiency who avoid environmental pollutants such as smoking can often survive until the sixth or seventh decade of life. Smoking in patients with α_1-antitrypsin deficiencies has been shown to be the single most important factor in the development of early onset emphysema. The mean onset of dyspnoea (difficulty of breathing) in PI Z-positive patients, who do not smoke is 51 years, while those that do it is 32 years (Janus *et al.*, 1985).

The *PI*Z* mutation itself is a disease-causing mutation mainly in Caucasians and it has been estimated that the carrier incidence of the *PI*Z* allele is approximately 4% in North Europeans, with a frequency of approximately 1 in 6700 (Carrell and Lomas, 2002). Although PI MZ heterozygotes do not develop any disease symptoms, when the *PI*Z* allele is combined with another mutation *PI*S*, to produce a PI SZ phenotype, individuals can have plasma α_1-antitrypsin levels that fall below the critical level. Interestingly PI SS homozygotes do not show any disease symptoms. Although the level of α_1-antitrypsin is reduced to 40% of normal, it is not sufficiently low to be considered as disease-causing. The PI SS phenotype is caused by the substitution of a glutamic acid residue by a valine residue at amino acid position 264 in the sequence of α_1-antitrypsin (D264V) (Brantly *et al.*, 1988).

PI Z α_1-antitrypsin has a normal rate of synthesis, a similar half-life to that of normal PI M α_1-antitrypsin (5.2 days) and a similar association constant for neutrophil elastase to normal PI M α_1-antitrypsin (Errington *et al.*, 1982; Bathurst *et al.*, 1984). However, the secretion of PI Z α_1-antitrypsin from the hepatocytes into the circulation is decreased, which is consistent with the accumulation of PI Z α_1-antitrypsin in hepatocyte inclusions (Sharp, 1971). The accumulation of PI Z α_1-antitrypsin in the liver cells has been attributed to the mutation, E342K, being situated in a critical area in the hinge of the reactive loop centre that allows spontaneous opening of the main sheet of the molecule (Carrell and Lomas, 1997; Cox *et al.*, 1986). This facilitates the insertion of the reactive loop of one α_1-antitrypsin molecule into the open sheet of another molecule, which in turn links to another molecule forming a polymer of α_1-antitrypsin, which accumulates in the hepatocytes. The sub-cellular location of the α_1-antitrypsin aggregates has been shown subsequently to be the endoplasmic reticulum. Hercz and colleagues (1978) showed that the carbohydrate structures present on the accumulating PI Z α_1-antitrypsin were high-mannose glycans. This observation was confirmed by Bathurst *et al* (Bathurst *et al.*, 1984) who demonstrated that the carbohydrate chains were unprocessed and thus had not been transported to the Golgi apparatus for final processing and secretion. In addition to the lower rates of secretion of α_1-antitrypsin, Cox and colleagues (1986) have shown that effect of the low plasma concentrations in PI Z variants is also compounded by the fact that the PI Z α_1-antitrypsin is prone to aggregation in the circulation. PI Z α_1-antitrypsin deficiency has been included in the group of diseases referred to as the conformational diseases, which include familial encephalopathy, familial Parkinson's disease with Lewy bodies, Creutzfeld-Jacobs disease, bovine spongiform encephalitis, Alzheimer's disease, Pick's disease and Huntington's disease (Carrell and Lomas, 1997).

The PI Z variant is also associated with a liver disease in neonates (Sharp *et al.*, 1969) but not all patients with the PI Z α_1-antitrypsin deficiency go on to develop cirrhosis. Only approximately 12-15% of this genotype develop significant liver disease which may require liver transplantation (Wu *et al.*, 1994). Currently, PI Z α_1-antitrypsin deficiency is the commonest metabolic

defect that requires orthotopic liver transplantation. The reasons why some patients with the *PI ZZ* genotype develop end-stage liver failure and others do not, remains contentious. Wu and colleagues (1994) have shown that when fibroblasts were transduced to over-express α_1-antitrypsin, there were considerable differences in the degradation of the PI Z α_1-antitrypsin from the endoplasmic reticulum between the two groups of patients. The PI Z patients who developed significant liver disease showed a marked lag in the degradation of the PI Z α_1-antitrypsin from the ER. Qu and colleagues (1996) demonstrated that PI Z α_1-antitrypsin binds to calnexin through repeated cycles in the ER until it achieves its correct conformation. If the molecule fails to fold properly it is transferred to the proteasomes for degradation. Although the proteasome is on the cytosolic side of the ER membrane, the incorrectly folded PI Z α_1-antitrypsin on the luminal side of the ER can be directed to the proteasome, by polyubiquitination of the cytosolic portion of the membrane-bound protein calnexin to which it is bound. The method by which the PI Z α_1-antitrypsin-calnexin complex is removed from the ER to the proteasomes is not fully understood but the whole complex is degraded by the proteasome. In the PI Z variants that developed liver disease, there was a poor association between the PI Z α_1-antitrypsin and calnexin, which allows the PI Z-α_1-antitrypsin molecule to escape the degradation pathway because of inefficient polyubiquitination. Other variants were shown to have normal interaction between PI Z α_1-antitrypsin and calnexin but slower rates of ubiquitination of the complex than PI Z variants which did not develop liver disease. These experiments indicate that the 12-15% of those PI Z α_1-antitrypsin-deficient patients, who develop significant liver disease, may be susceptible to conformational diseases because of inefficient quality control of folding in the ER (Wu *et al.*, 1994).

5.1.2 Experimental and theoretical rationale.

The experiments described in Chapters 3 and 4, established the procedures for the analysis of glycoproteins separated by PAGE, including new approaches for

the detection of post-translational modifications and amino acid mutations. In this chapter these techniques have been applied to the detection of:-

(i) changes in the amino acid sequence of plasma α_1-antitrypsin due to mutations in the α_1-antitrypsin gene and
(ii) the effect of these mutations on N-glycosylation.

To do this known neutral and charged amino acid substitutions in α_1-antitrypsin and an amino acid mutation, which causes the elimination of a N-glycosylation sequon, were studied.

5.1.3 Molecular variants of α_1-antitrypsin studied.

Patient I:
Plasma was obtained from a patient homozygous for a valine to alanine substitution at amino acid 213 in α_1-antitrypsin, leading to a neutral amino acid substitution (V213A).

Patient II:
Plasma was obtained from a patient homozygous for a glutamate to lysine substitution at amino acid 342 in α_1-antitrypsin (E342K), leading to a charged amino acid substitution (PI Z).

Patient III and IV:
Plasma was obtained from two patients heterozygous for a C→T transition at codon 255 which results in a change from threonine 85 (ACG) to methionine (ATG) and abolishes the N-glycosylation sequon (NLT → NLM) at amino acids 83-85, preventing glycosylation of asparagine 83 (PI $MZ_{Bristol}$, Lovegrove *et al.*, 1997).

5.2 Characterisation of normal isoforms of α_1-antitrypsin separated using 2D-PAGE.

5.2.1 Identification of α_1-antitrypsin isoforms in control plasma by in-gel enzymatic digestion and mass mapping.

The approximate pI and M_r co-ordinates of α_1-antitrypsin isoforms were obtained using the on-line Swiss-Prot database (http://www.expasy.ch/cgi-bin/nice2dpage.pl/def?P01009). According to this information the pI values and Mr values of all major isoforms of α_1-antitrypsin fall within the pH range 4.9 - 5.10 and mass range of 48-55 kDa, respectively. Therefore the control plasma samples were focused in the first dimension over the pH range 4.5-5.5 and separated by SDS-PAGE in a 10% resolving gel. Proteins were visualised using silver staining (Figure 5.01a).

The major protein spots observed on the gel were excised, digested in-gel with trypsin and desalted as described previously (Chapter 2). Each polypeptide digest was then analysed by MALDI TOF MS. The resulting spectra obtained from the protein digests were used to identify each protein by mass mapping using 'MS FIT', the on-line 'Protein Prospector' database searching tool (http://prospector.ucsf.edu/).

Figure 5.01a shows the identification of the major protein spots excised from the 2D-PAGE over pH range 4.5-5.5. It was possible to identify most of the proteins separated by 2D-PAGE with as little as 15 µl of plasma. Using the identity of the proteins obtained by mass mapping, in conjunction with the on-line plasma 2D-PAGE database Geneva, Switzerland (http://www.expasy.ch/cgi bin/map2/def?PLASMA_HUMAN) and the pI / molecular weight), it was possible to create an accurate pI / molecular weight co-ordinate system for the entire 2D-PAGE gel (Figure 5.01a). On the basis of these co-ordinates, the pI and molecular weights of the seven α_1-antitrypsin isoforms of α_1-antitrypsin that were identified using mass mapping were deduced (Figure 5.01b). For the purposes of identification, the α_1-antitrypsin isoforms were designated the letter a to g depending on the pI of the isoform i.e. isoform (a) having the lowest pI value and isoform (g) having the highest pI value. The pI values of the isoforms

were 4.86, 4.91, 4.95, 5.00, 5.05, 5.10 and 5.16 in this order. The approximate molecular weights of the isoforms ranged from 51.0 -53.1 kDa.

Figure 5.01

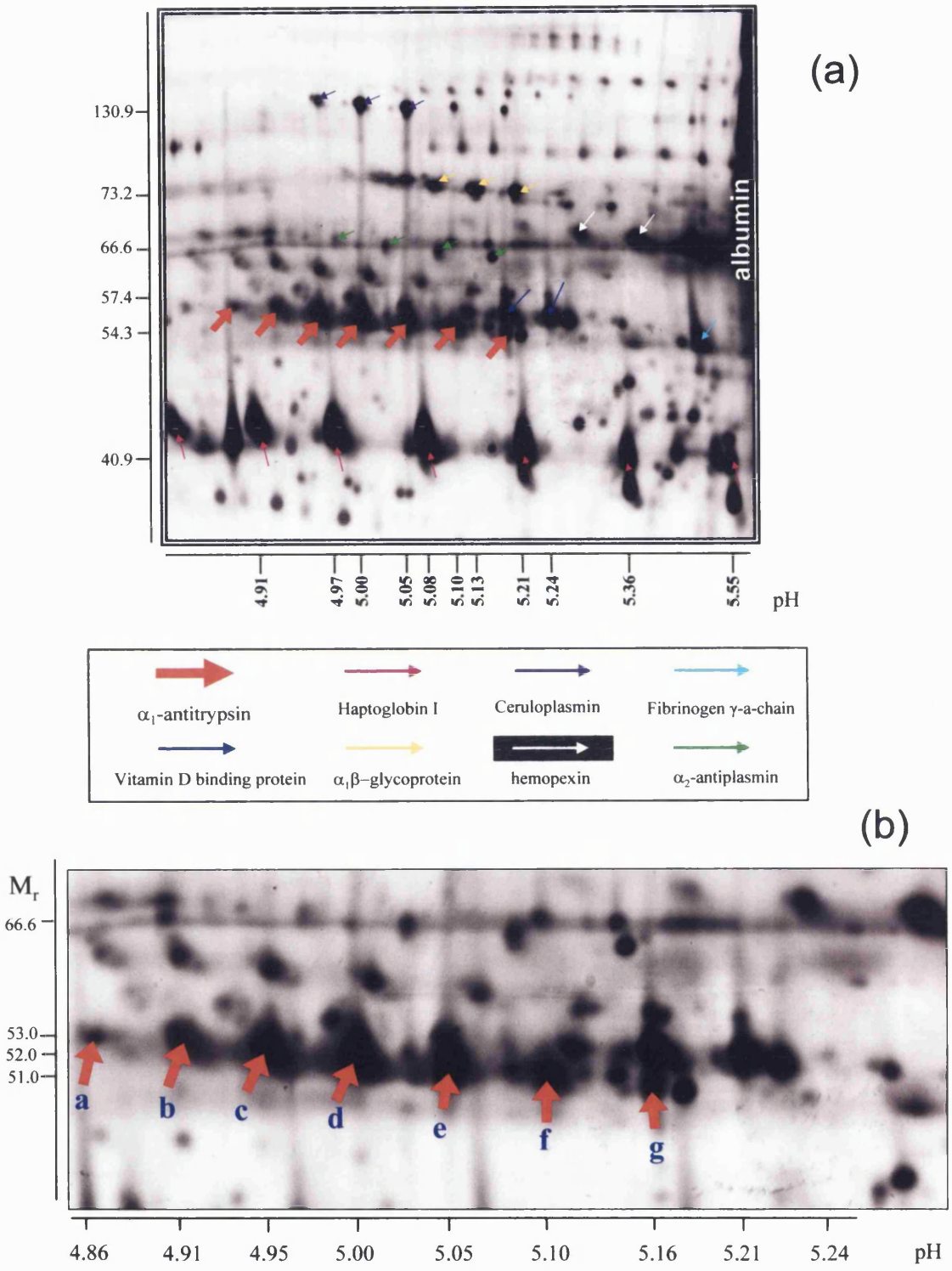


Figure 5.01 2D-PAGE of control plasma (pH 4.5-5.5)

(a) The major proteins identified after separation by 2D-PAGE and mass mapping.

(b) Enlarged section of 2D-PAGE showing seven α_1-antitrypsin isoforms detected by MALDI TOF MS.

5.2.1.1 Assignment of α_1-antitrypsin isoforms in control plasma to the conventional M series notation.

The α_1-antitrypsin proteins isoforms separated by IEF are conventionally designated as the M series (M1-M8) according to their pI values (anodal-low pH to cathodal-high pH, see Figure 1.24b). The difference in the pI value depends primarily on the number of sialic acid residues on the N-linked glycans. However, two minor cathodal forms of α_1-antitrypsin (M7 and M8) were found to contain the same carbohydrate structures as the M4 and M6 isoforms but they lack the first five N-terminal amino acids (Jeppsson *et al.*, 1985).

5.1.1.2 Identification of the M7 and M8 isoforms of α_1-antitrypsin.

The presence or absence of a peptide covering the N-terminal amino acids on the α_1-antitrypsin isoforms (f) and (g), was used to identify them as the M7 and M8 isoforms and to distinguish them from the M1-M6 isoforms. The polypeptide mass spectral data for all isoforms were analysed for peptide masses that corresponded to the presence or absence of N-terminal amino acids. An additional proteolytic digestion step, using chymotrypsin, was incorporated into the analyses to confirm fully the identity of each isoform. Figures 5.02a and b, show typical mass spectra obtained from the α_1-antitrypsin isoforms (d) and (e), after in-gel digestion with trypsin and chymotrypsin, respectively. The presence of peptides of masses 2819.2 m/z and 3704.7 m/z (Figure 5.02a and b) corresponding to peptides containing the amino acids 1-25 and 1-33, respectively, in the α_1–antitrypsin polypeptide indicate that these isoforms do contain the first 5 amino acids. Both peptides contain missed cleavage sites at lysine 10 and phenylalanine 23. No ions corresponding to peptides containing the first five amino acids were observed in the analysis of isoform f (Figure 5.02c) and isoform g (Figure 5.02d), supporting their identification as M7 and M8, respectively. However, a tryptic peptide mass (1779.8 m/z) corresponding to residues 11-26 was detected. It appears that tryptic cleavage after lysine 10 occurs more readily in the truncated protein than in the intact protein, which prevents the appearance of a peptide corresponding to residues 6-25 (2293.0

m/z) from the full length protein. On this basis, the two most anodal isoforms identified on the 2D-PAGE, isoforms (f) and (g), were tentatively identified as isoforms M7 and M8, respectively.

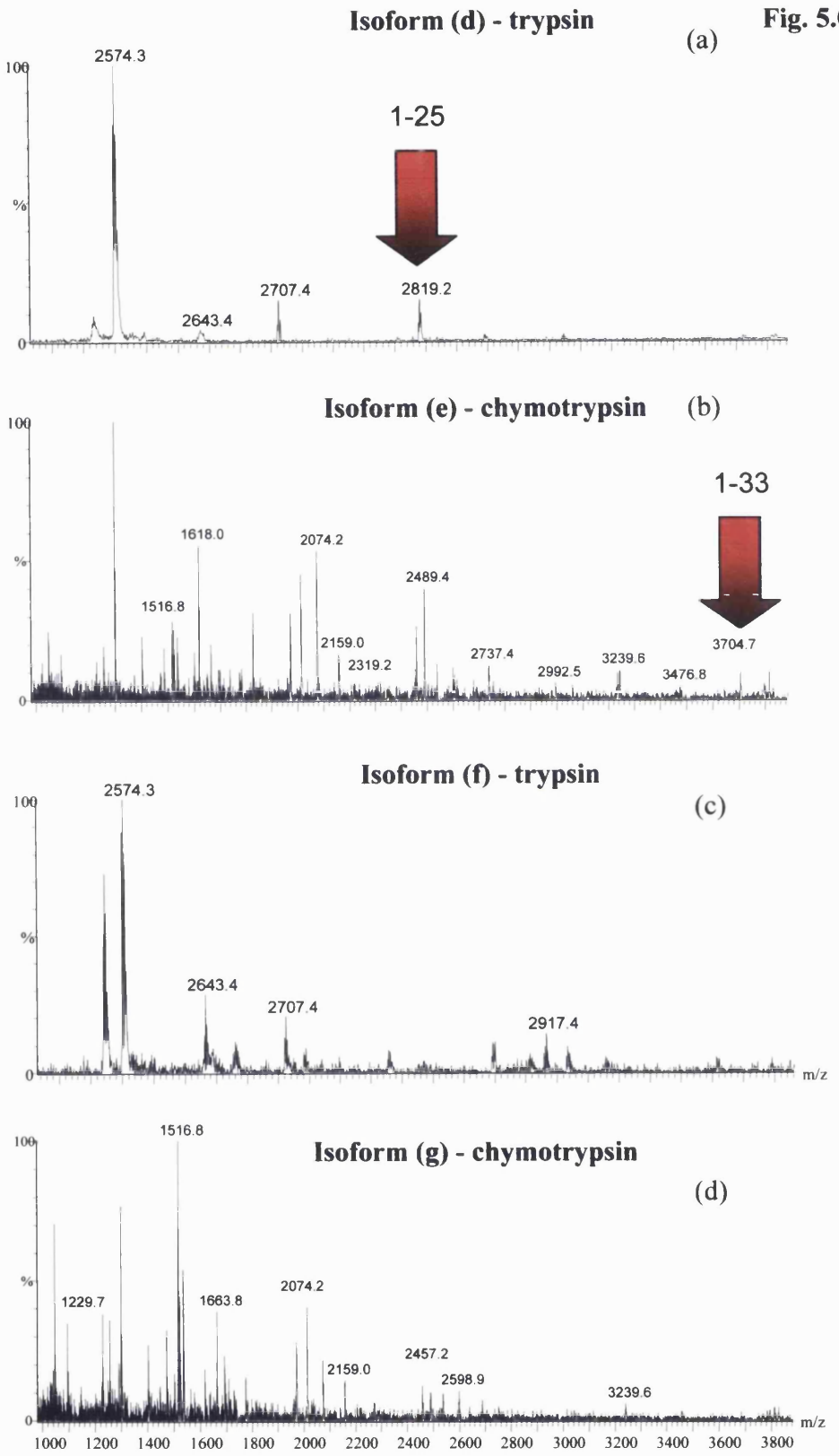


Figure 5.02 MALDI TOF mass spectra of tryptic and chymotryptic digests of α_1-antitrypsin isoforms in normal plasma.

(a) Tryptic digest of isoform (d) showing a peptide of mass 2819.2 m/z corresponding to amino acids 1-25.
(b) Chymotryptic digest of isoform (e) showing a peptide of mass 3704.7 m/z corresponding to amino acids 1-33.
(c) Tryptic digest of isoform (f) showing the absence of peptides corresponding to amino acids 1-25 in the sequence.
(d) Chymotryptic digest of isoform (g) showing no extra peptides corresponding to amino acids 1-33 in the sequence.

5.2.1.3 Identification of the M4 and M6 isoforms of α_1-antitrypsin.

The main α_1-antitrypsin isoforms present in control plasma are the M4 and M6 isoforms, which are known to constitute approximately 40% and 34% of the total α_1-antitrypsin, respectively (Cox, 2001). The most abundant α_1-antitrypsin isoforms detected on the gel were isoforms (d) and (e) (Figure 5.01b), implying that they were the M4 and M6 isoforms, respectively. To confirm the identity of isoforms (d) and (e) the glycans were released with treatment with endoglycosidase N-glycanase and analysed using MALDI TOF MS (see section 2.10.2). Figure 5.03a shows the glycan analysis of the isoform (d). Two major glycans, disialo-biantennary (m/z, 2223.4) and trisialo-triantennary (m/z, 2880.1) were detected in a ratio of approximately 2:1. The next most abundant peak, 1932.1 m/z, is probably due to in-source loss of sialic acid from the major bi-antennary glycan. Masses corresponding to fucosylated bi- (m/z 2369.5) and tri-antennary (m/z 3026.3) complex glycans were also observed and constituted 5% and 7% of the total glycan signal. The glycan analysis of isoform (e) or the predicted M6 isoform, showed the presence of only one major species corresponding to a disialo-biantennary complex glycan structure (Figure 5.03b). As with the analysis of the M4 spot, a mass corresponding to a fucosylated bi-antennary structure was also observed, which constituted approximately 8% of the total glycan signal. From these data it was deduced that the isoforms designated (d) and (e) separated by 2D-PAGE (Figure 5.01b) were the M4 and M6 isoforms, respectively.

These results are in agreement with previously published data for the microheterogeneity of glycosylation of the major isoforms of α_1–antitrypsin in human plasma (Vaughan *et al.*, 1982; Fagerhol and Laurell, 1967; Jeppsson *et al.*, 1985). The M4 isoform of α_1–antitrypsin has been reported to contain two bi-antennary and one tri-antennary complex glycan, which is in agreement with the present observation of a ratio of 2:1 of bi / tri-antennary glycans in isoform (d). Similarly the detection of only bi-antennary glycans on isoform (e) agrees with the data published by Jeppsson and colleagues (1985) for the glycan structures present on the M4 α_1-antitrypsin isoform. Naitoh and colleagues (1999) have reported that the M4 and M6 α_1–antitrypsin isoforms contain small

amounts of fucosylated glycans, which again is in agreement with the present observations. The glycosylation analysis of the putative M7 (isoform f) and M8 isoforms (isoform g), were the same as those found for the M4 and M6 isoforms, respectively, confirming that their different pI's were due to the loss of the five N-terminal amino acids and not due to a difference in glycosylation.

Figure 5.03

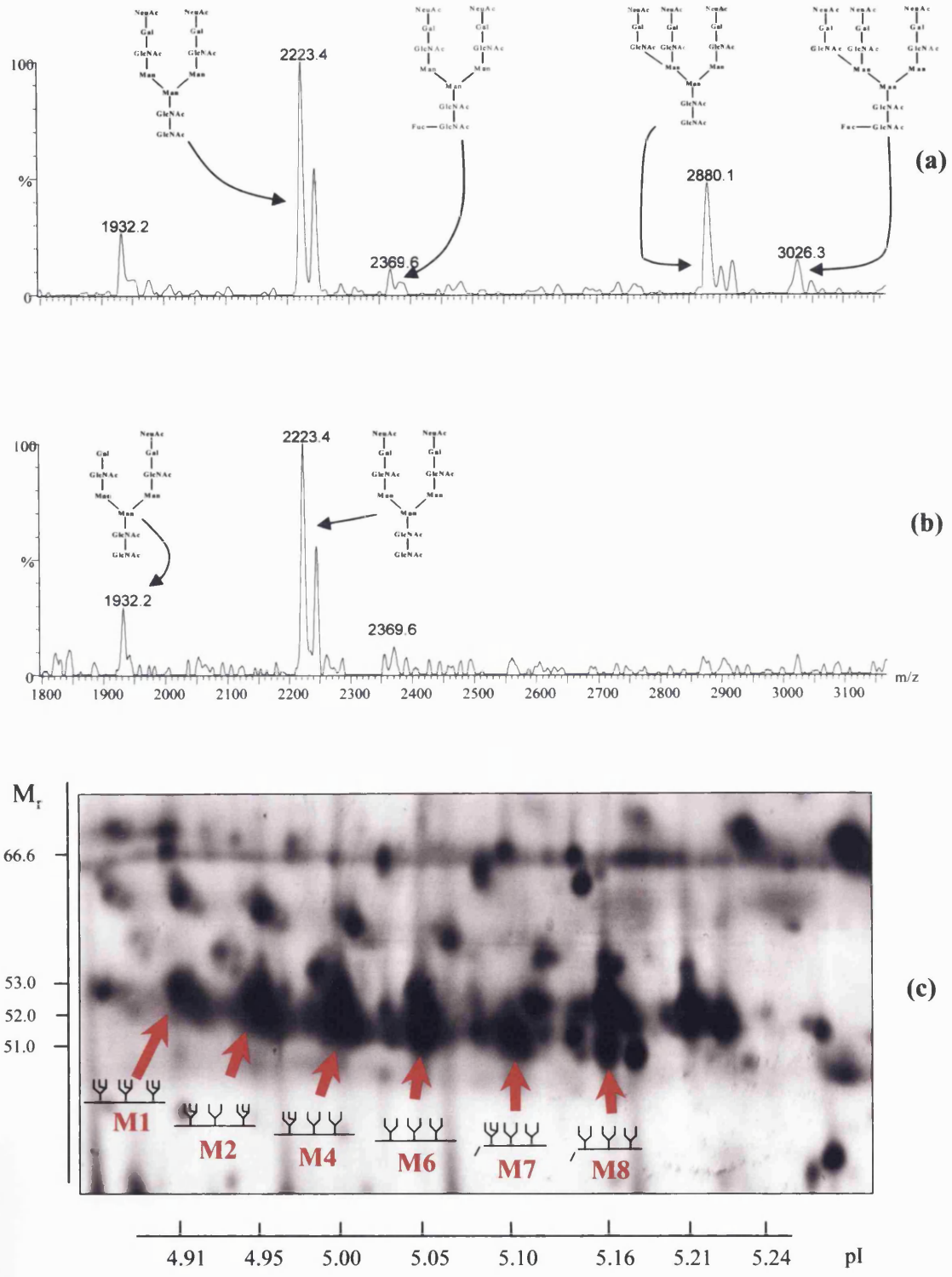


Figure 5.03 Glycan analysis of α_1-antitrypsin isoforms separated by 2D-PAGE.

(a) Mass spectra of the glycans obtained from the in-gel N-glycanase digestion of the M4 isoform of α_1-antitrypsin.

(b) Mass spectra of the glycans obtained from the in-gel N-glycanase digestion of the M6 isoform of α_1-antitrypsin.

(c) Proposed nomenclature of the M series of α_1-antitrypsin in control plasma after separation by 2D-PAGE.

5.2.1.4 Identification of the M1 and M2 isoforms of α_1-antitrypsin.

The 2D-PAGE data showed a difference in pI of 0.05 units between each of the M4, M6, M7 and M8 isoforms of α_1-antitrypsin. Under the conditions used here a difference of one sialic acid residue or the loss of the five N-terminal amino acids on an α_1-antitrypsin isoform results in an increase in pI of 0.05 and 0.1 pH units, respectively. The results of the in-gel tryptic digestion (section 5.2.1.1) showed that the first five N-terminal amino acids were present in the isoforms (a)-(e). Therefore the difference in pI between each isoform observed on the 2D-PAGE, must be due to the carbohydrate moieties and not due to changes in the polypeptide backbone. Unfortunately, mass spectral signal of the glycans released from α_1-antitrypsin isoforms (a)-(c) were extremely weak and their identity could not be confirmed by glycan analysis as for the M4, M6, M7 and M8 isoforms. However, on the basis of relative estimated pI values, the isoforms (b) and (c) have pIs which are 0.05 units less than those of the M4 and M6, respectively. This would be consistent with α_1-antitrypsin isoform c (Figure 5.01b) containing one more sialic residue than the M4 isoform, suggesting that it contains two tri- and one biantennary glycans and is the M2 α_1-antitrypsin isoform. The pI difference observed on the 2D-PAGE of ~0.1 pI units between isoform (b) and the M4 α_1-antitrypsin isoform indicates that it contains two additional sialic acid residues. This suggests that isoform (b) is the M1 isoform of α_1-antitrypsin, which contains a triantennary glycan at each of the three glycosylation sites. The α_1-antitrypsin isoform (a), which was present at low concentrations and only detected during preparative 2D-PAGE analyses, is probably a minor isoform containing one tetra-antennary and two triantennary glycan structures.

5.2.2 Summary of α_1-antitrypsin isoforms separated by 2D-PAGE.

The final proposed nomenclature for the M series of normal control α_1-antitrypsin isoforms separated by 2D-PAGE is shown in Figure 5.03c. The use

of high resolution isoelectric focusing over the pH range 4.5-5.5 allowed complete separation of the isoforms of α_1–antitrypsin, which would not be possible using wider pH ranges of ampholines. The ability to 'focus' on particular pH ranges allows greater resolution of individual proteins and significantly increases the probability of resolving the protein of interest from other proteins of similar pI and mass for analysis by mass spectrometry. The combination of glycan analysis to determine the microheterogeneity of glycosylation and peptide mass fingerprinting allowed the accurate identification of the abundant isoforms, M4, M6, M7 and M8 and a reasonable prediction of the structure of the M1 and M2 isoforms of α_1–antitrypsin separated by 2D-PAGE.

The mass mapping identification of the plasma proteins on the 2D-PAGE (Figure 5.02a) and their correlation with their pI and molecular weights contained in the Swiss-Prot protein database, allowed the construction of a series of highly accurate 'anchor points' or co-ordinates, that could be used to calculate the pI and molecular weight of other proteins or protein isoforms present on the gel.

5.3 Analysis of a neutral amino acid substitution in the α_1-antitrypsin variant (V213A).

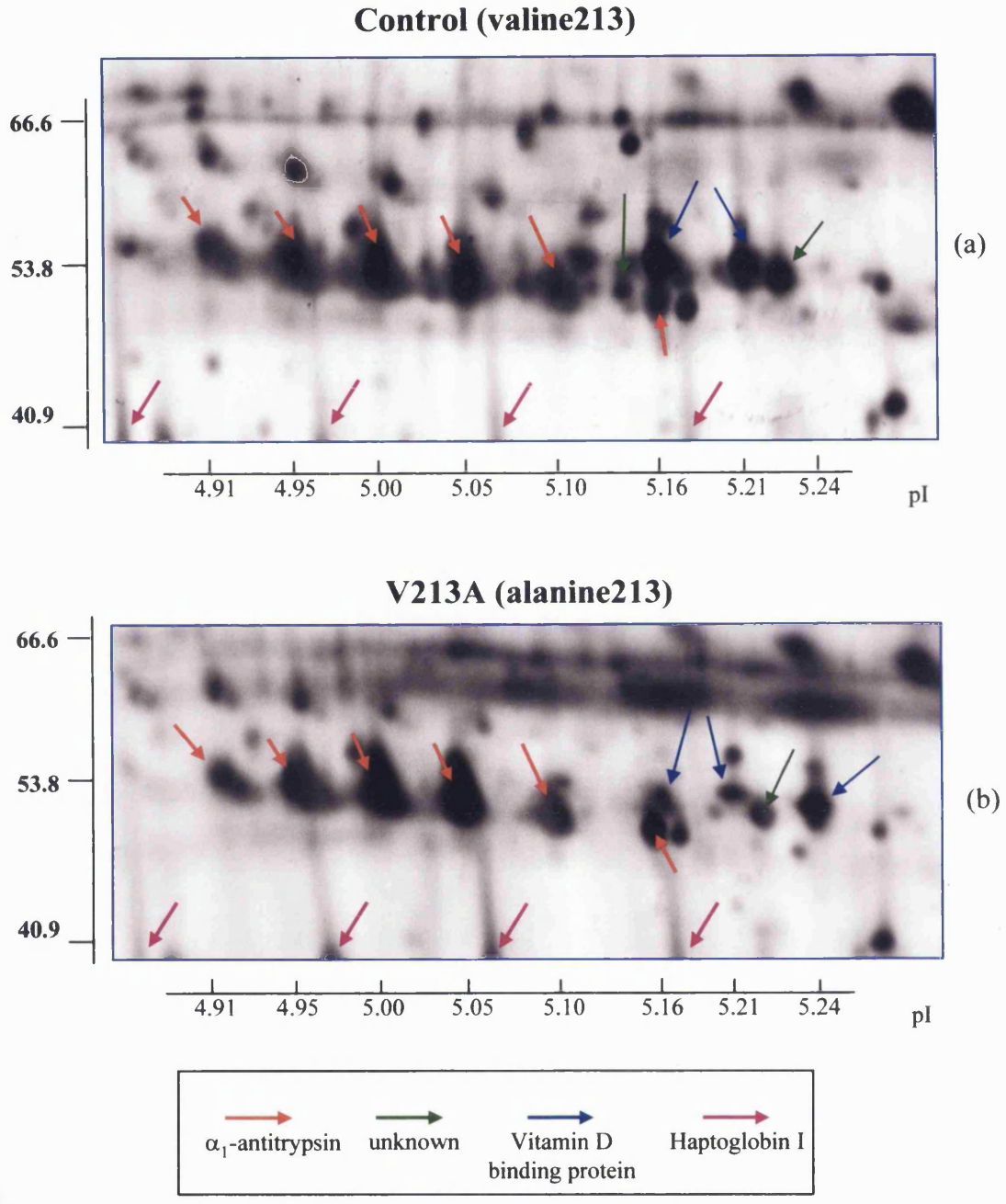
5.3.1 2D-PAGE analysis of plasma from patient V213A.

No marked changes in the protein intensity, pI or molecular weight of α_1-antitrypsin isoforms between the control and V213A variant were observed on the 2D-PAGE (Figure 5.04a and b, respectively). This is consistent with a neutral amino acid substitution, which would not be expected to alter the physical characteristics of the α_1-antitrypsin protein.

5.3.2 In-gel tryptic digestion of α_1-antitrypsin isoforms separated by 2D-PAGE.

Figure 5.05a and b show the MALDI TOF MS spectra of the tryptic digests of the control and the variant (V213A), M4 α_1-antitrypsin isoforms respectively. The mass spectral analyses of the tryptic digests of the control and patient α_1–antitrypsin isoforms were similar except for the fragment corresponding to the tryptic peptide containing amino acids 192-217. Masses of 3148.5 m/z and 3120.5 m/z were observed from the equivalent α_1-antitrypsin isoforms from the normal and V231A variant, respectively (Figure 5.05a and b respectively). These masses differ by 28 m/z and are consistent with a peptide containing a valine to alanine substitution at residue 213 in the variant. The mass of 3148.5 m/z derived from the normal α_1-antitrypsin was not observed in the α_1-antitrypsin from the V213A variant patient and vice versa. This peptide mass discrepancy between the control and variant (V213A) was also observed in each of the M series isoforms of α_1-antitrypsin.

Figure 5.04

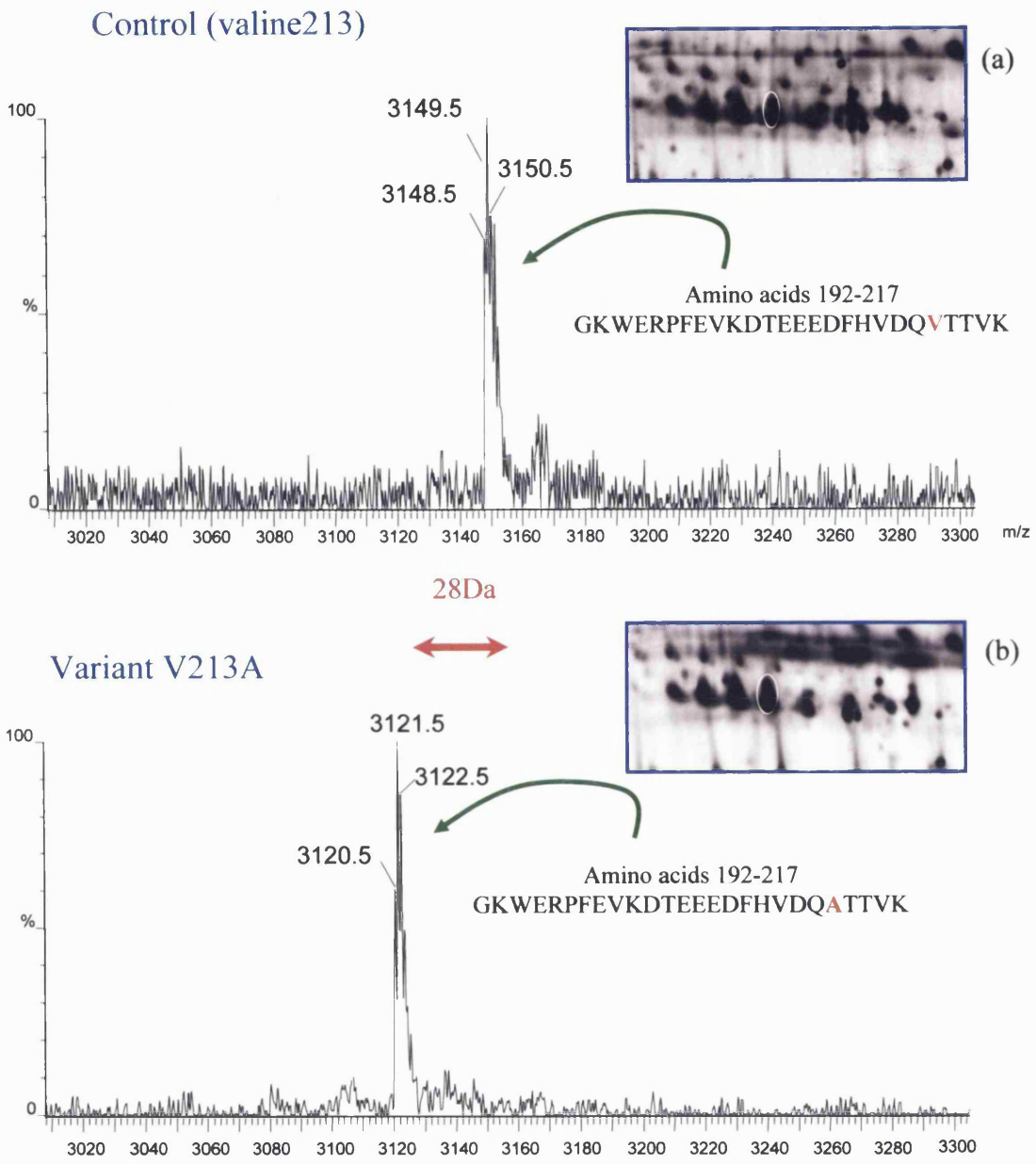


Figures 5.04 2D-PAGE of normal and V213A variant plasma.

(a) Enlarged section of a 2D PAGE of normal plasma α_1-antitrypsin with IEF over the pH range, 4.5 – 5.5

(b) Enlarged section of a 2D PAGE of V213A variant plasma α_1-antitrypsin with IEF over the pH range, 4.5 – 5.5

Figure 5.05



Figures 5.05 MALDI TOF mass spectra of the tryptic in-gel digestion of normal and V213A α_1-antitrypsin.

(a) Tryptic in-gel digest of the control M4 isoform showing a peptide of mass 3148.5m/z corresponding to amino acids 199-217 containing valine at position 213.

(b) Tryptic in-gel digest of the MIA213, M4 isoform showing a peptide of mass 3120.5m/z corresponding to amino acids 199-217 containing alanine at position 213.

5.4 Analysis of a charged amino acid substitution (PI Z variant, E342K).

5.4.1 2D-PAGE analysis of PI Z plasma.

The 2D-PAGE analysis of plasma from a PI Z patient showed changes in the intensity and pI of all α_1-antitrypsin isoforms compared to the normal plasma sample (Figure 5.06a and b). The intensity of the main M4 and M6 isoforms in the PI Z plasma was approximately 20% of the intensities of the equivalent spots in the normal plasma. The level of α_1-antitrypsin in PI Z plasma has been reported by Svenger (1985) to be 17 ± 3% of that in normal plasma. The six α_1-antitrypsin isoforms in the PI Z plasma showed a cathodal shift relative to their counterparts in normal plasma (Table 5.1).

Isoform	Control α_1-antitrypsin	PI Z α_1-antitrypsin
M8	5.15	5.27
M7	5.10	5.22
M6	5.05	5.16
M4	5.00	5.10
M2	4.95	5.04
M1	4.90	4.98

Table 5.1 Estimated pI values of α_1-antitrypsin isoforms in normal control plasma and in plasma from a PI Z patient.

The average cathodal shift in pI between each of the PI Z isoforms and its equivalent normal isoform was 0.1 pH units. This is approximately double the pI difference observed between adjacent isoforms in the same plasma sample (0.05 pH units), which differ by one sialic acid residue or one negative charge. Thus the increase in pI of each PI Z isoform relative to its normal counterpart can be explained by the substitution of a positively charged lysine residue for a negatively charged glutamate residue with a net decrease in negative charge of two, which is equivalent to an increase in pI of approximately 0.1 pH units.

Figure 5.06

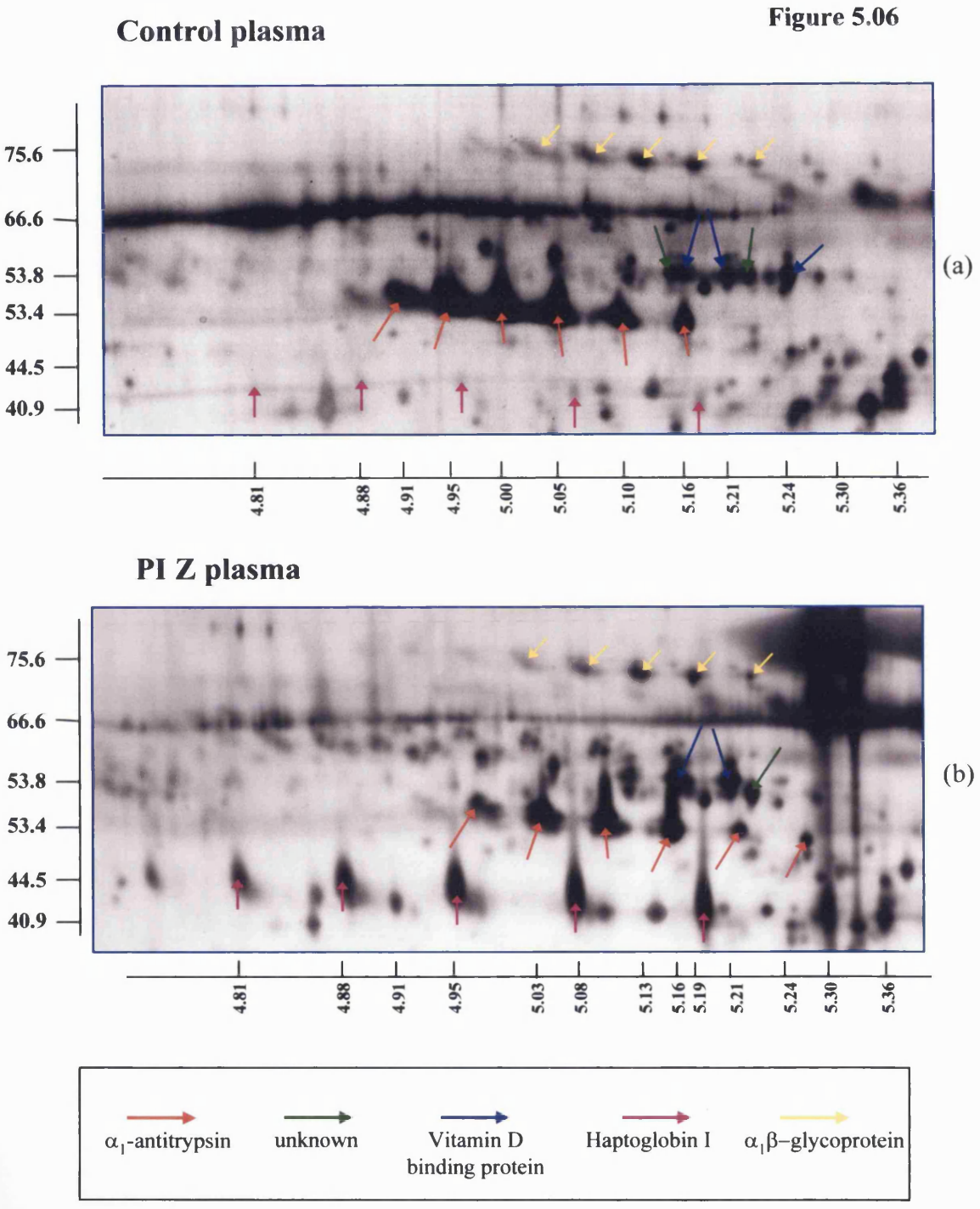


Figure. 5.06 2D-PAGE of normal and PI Z plasma.

(a) Enlarged section of a 2D PAGE of normal plasma α_1-antitrypsin with IEF over the pH range, 4.5 – 5.5.

(b) Enlarged section of a 2D PAGE of PI Z (Glu342Lys) plasma α_1-antitrypsin with IEF over the pH range, 4.5 – 5.5.

5.4.2 In-gel tryptic digestion of α_1-antitrypsin isoforms separated by 2D-PAGE.

All the isoforms of α_1-antitrypsin in the variant plasma separated on 2D-PAGE were identified by in-gel tryptic digestion and peptide mass mapping. Although the quality of the MALDI TOF MS spectra were sufficient for the identification (as α_1-antitrypsin) of the six isoforms of α_1-antitrypsin present in the PI Z variant (Figure 5.06b), no masses corresponding to peptides containing the E342K mutation were detected. Low levels of α_1-antitrypsin were detected in the plasma of the PI Z variant separated by 2D-PAGE (Figure 5.06b) in agreement with the low levels described previously (Brantly *et al.*, 1991). The low levels of PI Z α_1-antitrypsin present in the variant 2D-PAGE resulting in the diminished intensity of the tryptic peptide ions detected during their subsequent MALDI TOF MS analyses.

5.4.3 Purification of plasma PI Z α_1-antitrypsin by affinity chromatography.

To overcome the problem of the low circulatory levels of α_1-antitrypsin in the plasma of the PI Z patient, affinity chromatography was used to isolate and purify α_1-antitrypsin. This permitted the in-gel tryptic analysis of higher amounts of PI Z α_1-antitrypsin than could be isolated using 2D-PAGE. The α_1-antitrypsin in the variant plasma was purified by immuno affinity chromatography as described in section 2.4 and analysed by 1D-PAGE after gel filtration to remove low molecular weight contaminants (Figure 5.07). Two protein components with approximate molecular weights of 53 and 67 kDa were detected in all fractions.

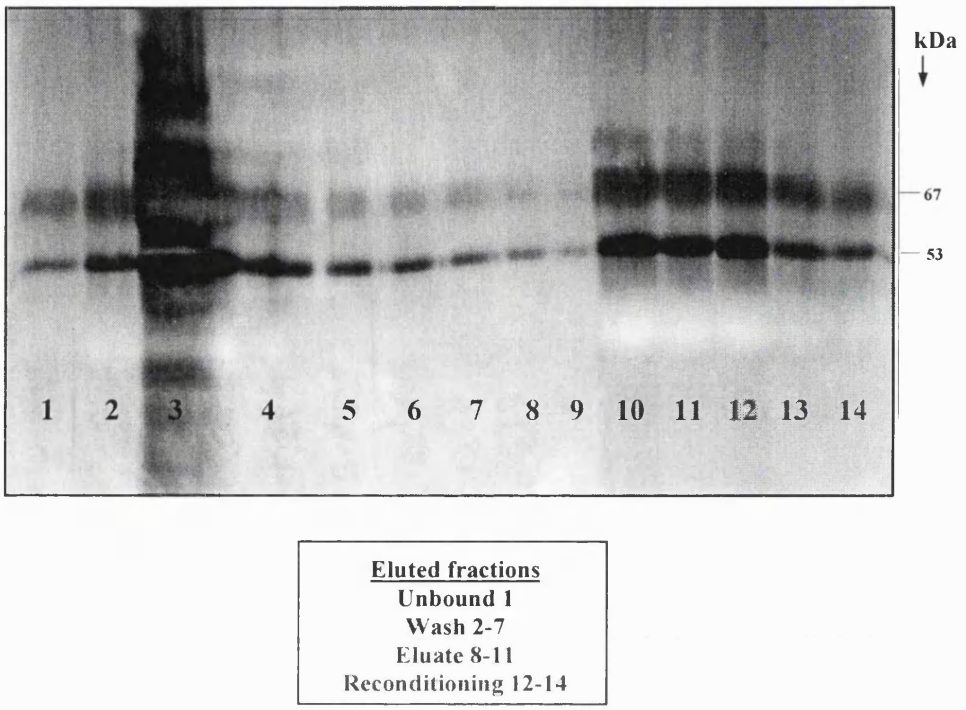


Figure 5.07 1D-PAGE analysis of immunopurified PI Z α_1-antitrypsin.

The 53 kDa and 67 kDa proteins were identified as α_1-antitrypsin and albumin, respectively, using in-gel tryptic digestion and peptide mapping. Approximately equal amounts of albumin and PI Z α_1-antitrypsin were co-immunopurified during the affinity chromatography purification (Figure 5.07, lanes 10-14). However, the use of a 1D-PAGE after the immunopurification permitted the separation of albumin from the PI Z α_1-antitrypsin.

A second preparative immunopurification of PI Z α_1-antitrypsin was performed as described previously, the eluant and reconditioning steps were combined, desalted using ultracentrifugation and separated by 1D-PAGE (Figure 5.08).

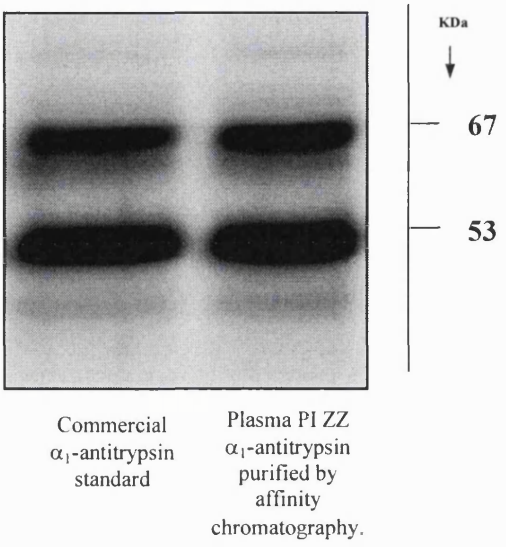


Figure 5.08 1D-PAGE of commercial α_1-antitrypsin standard and PI Z α_1-antitrypsin purified by affinity chromatography.

5.4.3.1 Detection of the E342K mutation in PI Z α_1-antitrypsin.

The purified α_1–antitrypsin band was digested in-gel with trypsin and the resulting peptide mixture analysed by MALDI TOF mass spectrometry. Figures 5.09a and b show the typical spectra observed from the tryptic in-gel digestion of control and PI Z α_1-antitrypsin over the mass range 2020-2640 m/z. Three extra masses, 2387.2 m/z, 2403.2 m/z and 2419.2 m/z, were observed in the PI Z $\alpha_1$antitrypsin spectrum which were not observed in the control analysis. These masses correspond to a peptide covering the amino acids 343-365 in the sequence of α_1–antitrypsin after cleavage between the two lysine residues at

Figure 5.09

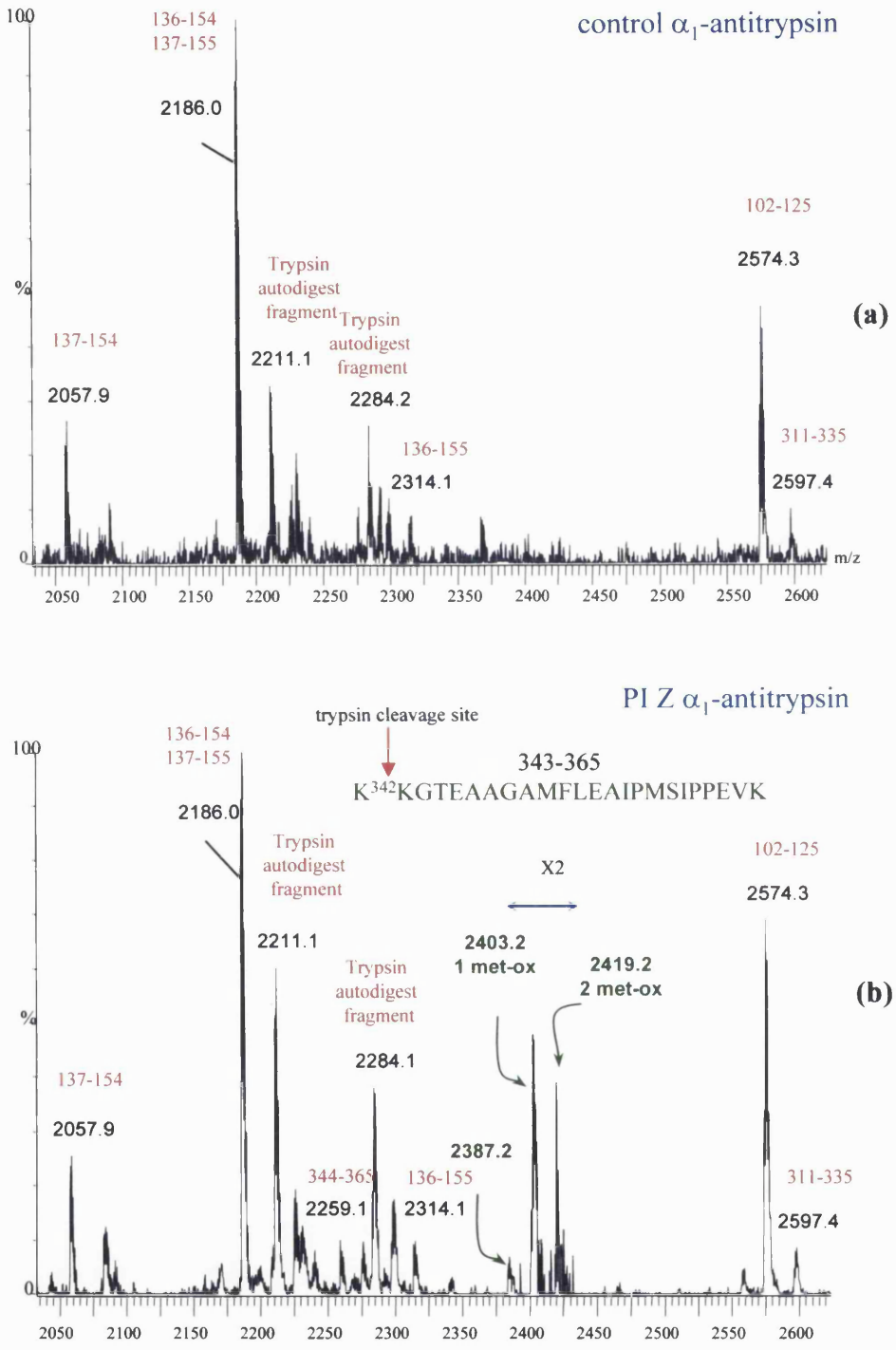


Figure 5.09a Mass spectra of tryptic peptides from control and PI Z α_1-antitrypsin over the mass range 2020 – 2640 m/z.
(a) Control (b) PI Z,
Numbers in red or green correspond to the amino acid sequences in α_1-antitrypsin.

positions 342 and 343. The detection of three masses for a single peptide is explained by the presence of the same peptide in three forms i.e. native (2387.2 m/z) and with one (2403.2 m/z) or two (2419.2 m/z) oxidised methionine residues. This confirmed that the PI Z patient was homozygous for the E342K mutation. No peptide covering the amino acids 343-365 in the amino acid sequence was detected in the analysis of the normal plasma. A possible explanation for the detection of a peptide covering the mutation region in the PI Z sample but not in the control could be the nature of the amino acid substitution. The substitution of a lysine for a glutamate residue adjacent to another lysine residue at amino acid 343 creates a highly positively charged dilysine motif in the variant. This would be expected to enhance tryptic cleavage between the 2 lysines. In contrast, the negatively charged glutamic acid at position 342 in the normal sequence would be expected to inhibit cleavage after lysine 343. In fact this area of the sequence is resistant to in-gel tryptic digestion even after prior deglycosylation (See Figure 3.07, Chapter 3).

After detection of the E342K mutation in the PI Z α_1-antitrypsin that had been purified by affinity chromatography, a second 2D-PAGE of PI Z plasma proteins (30 µl) was performed using double the amount of plasma required normally for peptide mass mapping studies. The analysis of the tryptic in-gel digestion of each isoform of PI Z α_1-antitrypsin also showed the presence of three extra masses of 2387.2 m/z, 2403.2 m/z and 2419.2 m/z in the variant but not in the control (Figure 5.09a and b, respectively). Only peptides containing a lysine residue at position 342 as opposed to a glutamate were detected in the analyses.

5.4.3.2 Detection of a nitrosylated cysteine 232.

The possibility of nitrosylation of α_1-antitrypsin was examined by comparing the observed masses of all the peptides with the calculated masses for the peptides containing a nitrosylated or unmodified cysteine residue 232 in normal and variant α_1-antitrypsin. Analysis of the mass spectrum of the PI Z α_1-antitrypsin showed the presence of an extra peptide of mass 1962.1 m/z (Figure 5.10b), which was not detected in the control sample (Figure 5.10a) or any of the other variants analysed. The detected mass corresponded to a calculated mass

for the peptide, 218-233 in the amino sequence with a mass increment of 29 m/z, equivalent to the addition of a nitrosyl group. However, analysis of the mass spectra showed that a carboamidomethylated peptide containing cysteine 232 (2150.1 m/z) was also present in both the control and PI Z patient samples. No masses were detected for a peptide containing an unmodified cysteine residue, indicating that the carboamidomethylation of cysteine residues in the 2D-PAGE analyses is extremely efficient. The presence of a NO group on the cysteine residue 232 would block the addition of an acetamido group during this carbamidomethylation. These data indicate that cysteine residue 232 in the PI Z α_1-antitrypsin molecule is partially nitrosylated.

The detection of an ion of mass 1962.1 m/z strongly suggests that the peptide 218-233 is nitrosylated. However, to confirm this modification, this peptide should ideally be sequenced. The significance or benefit, of this post-translational modification of the cysteine group in the single PI Z variant analysed in this study is unclear. It has been suggested that α_1-antitrypsin can act as a reservoir for nitric oxide and that this nitrosylation can bestow antibacterial properties upon the molecule (Miyamoto *et al.*, 2000). It is also known that nitrosylation of proteins is stimulated during periods of inflammation and during infection (Ikebe *et al.*, 2000). However, the patient studied in this work had not suffered any recent bouts of infection prior to the plasma sample being taken making it unlikely that the nitrosylation of α_1-antitrypsin occurred for this reason. Another possible explanation is that cysteine 232 in the PI Z α_1-antitrypsin molecule is more readily nitrosylated or that it infers a greater stability and resistance to degradation than that of the wild type protein. The discovery that detectable amounts of [NO]-α_1-antitrypsin are found in the plasma of the PI Z variant is interesting but this finding needs further study to determine whether this phenomenon has any clinical relevance.

Figure 5.10

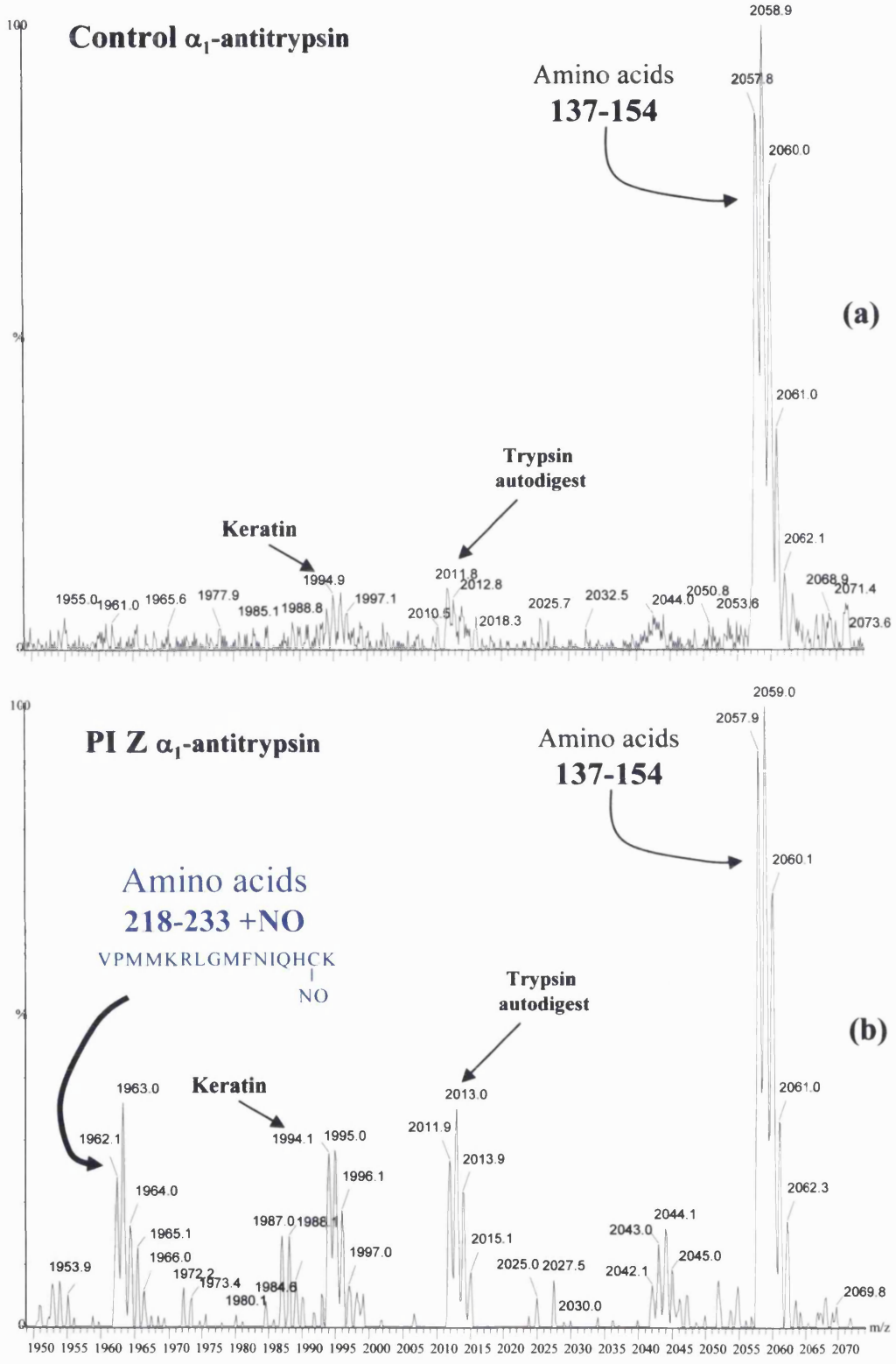


Figure 5.10 **Mass spectra of tryptic peptides from control and PI Z α_1-antitrypsin over the mass range 1950 – 2070 m/z.**
(a) Control (b) PI Z.

5.5 Analysis of PI $MZ_{Bristol}$ heterozygote variant (T85M).

5.5.1 2D-PAGE analysis of heterozygote plasma.

The 2D-PAGE analysis of the plasma from two PI $MZ_{Bristol}$ heterozygotes (Figure 5.11b and c, respectively) revealed the presence of 4 additional putative α_1–antitrypsin isoforms that are not present in the normal control plasma (Figure 5.11a). All 4 of these additional spots were identified as α_1–antitrypsin by mass mapping studies. Spots 1, 2, 3 and 4 had pI values of 5.07, 5.10, 5.16 and 5.36, respectively, and were approximately 2 kDa lighter than the normal M series of α_1–antitrypsin isoforms in the normal plasma. These changes in pI and mass are consistent with the loss of a complex biantennary glycan. The substitution of a methionine for threonine at amino acid position 85 in PI $MZ_{Bristol}$ α_1–antitrypsin results in the loss of a sequon for N-linked glycosylation i.e. $NLT^{85}E \rightarrow NLM^{85}E$ (Lovegrove *et al.*, 1997). The loss of this sequon would be expected to lead to isoforms of α_1–antitrypsin that are not glycosylated at asparagine 83 and have only two out the three-glycosylation sites occupied. The observed changes in pI and molecular weight suggest that the 4 novel isoforms contain the PI $Z_{Bristol}$ mutation. The presence of two sets of isoforms in the 2D-PAGE shows that both the normal (M) and variant alleles of α_1–antitrypsin are expressed in the plasma of PI $MZ_{Bristol}$ heterozygotes and can be detected by using the present procedure.

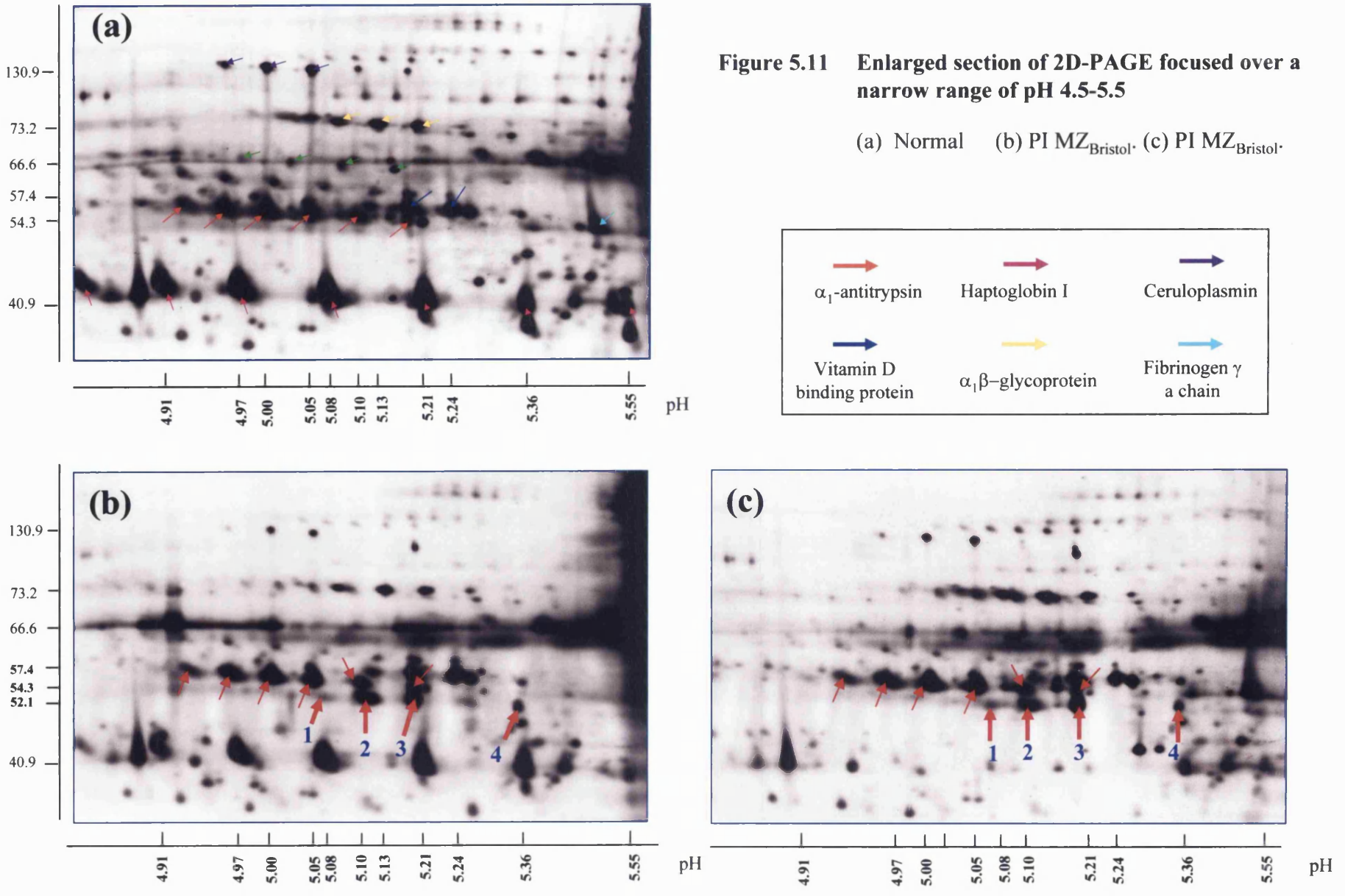


Figure 5.11 Enlarged section of 2D-PAGE focused over a narrow range of pH 4.5-5.5

(a) Normal (b) PI $MZ_{Bristol}$. (c) PI $MZ_{Bristol}$.

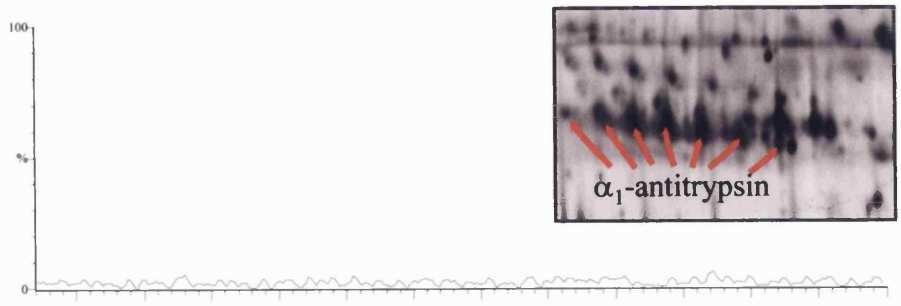
5.5.2 In-gel tryptic digestion of α_1-antitrypsin isoforms separated by 2D-PAGE.

To prove that the novel isoforms, 1-4, were due to the *PI* $MZ_{Bristol}$ allele, they were subjected to in-gel digestion with trypsin followed by mass spectrometry of the resultant peptide mixture. An additional mass of 3737.8 m/z was detected in each of the isoforms (Figure 5.12c), which was not present in the normal M series of α_1–antitrypsin in either the PI $MZ_{Bristol}$ heterozygote or the normal control (Figure 5.12a). A peptide of mass / charge ratio 3737.8 can be accounted for by a peptide containing the amino acids 70-101 with methionine at position 85 and asparagine at position 83 i.e. it is derived from the *PI* $MZ_{Bristol}$ allele with a non-glycosylated asparagine 83. A glycopeptide covering the sequon 83 was not observed in any of the M series of α_1–antitrypsin isoforms from normal or mutant samples. This is due to the difficulty of analysing complex sialic acid-containing glycopeptides using MALDI TOF MS. Further evidence that a peptide containing amino acids 70-101 can be produced by tryptic digestion and detected by MALDI TOF MS if it is not glycosylated, was obtained by using N-glycanase to remove glycans prior to mass mapping studies. A peptide of mass 3692.8 m/z was detected in the mass spectrum after in-gel tryptic digestion of the main M4 isoform after prior enzymatic removal of the glycans with N-glycanase F (Figure 5.12b). This mass corresponds to a peptide covering the amino acids 70-101 in the sequence but with an aspartic acid residue at position 83.

Further characterisation of the novel isoforms 1, 2, 3 and 4 was carried out by analysis of their N-terminal peptides and glycans. The major mutant isoforms, spots 2 and 3 (Figure 5.11b and c), both had the same glycan profile as the normal M4 and M6 isoforms (Section 5.2.1, Figure 5.03a and b, respectively). The peptide analysis of spots 2 and 3 also contained masses corresponding to amino acids 1-20 in the amino acid sequence. Therefore spots 2 and 3 are the equivalents of the normal major M4 and M6 isoforms lacking glycosylation at asparagine 83. The combined glycan and peptide analysis of spot 4 showed that it was the equivalent of the normal M8 isoform of α_1–antitrypsin lacking glycosylation at asparagine 83. It was not possible to carry out glycan

Figure 5.12

α_1-antitrypsin control with glycans present

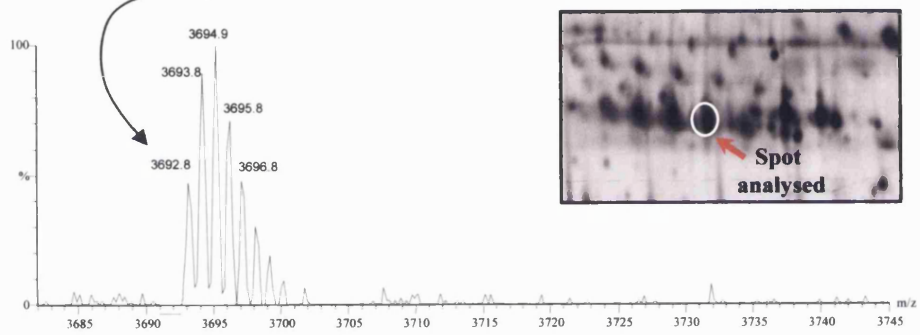


(a)

α_1-antitrypsin control, after deglycosylation using PNGase F

ADTHDEILEGLNFDLTEIPEAQIHEGFQELLR

Peptide sequence 70-101

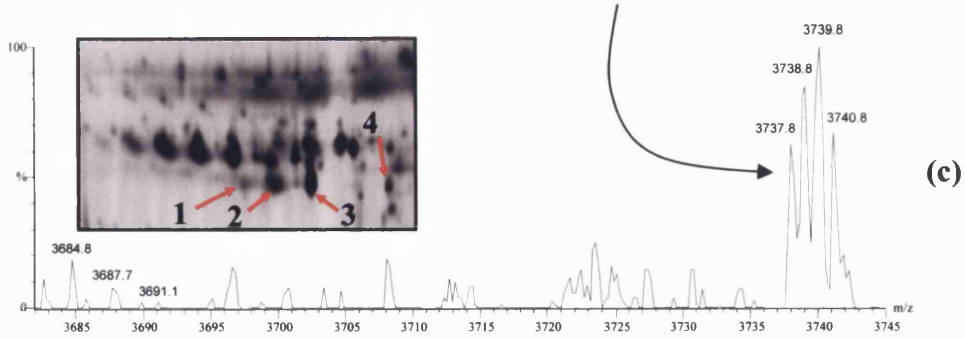


(b)

α_1-antitrypsin –Heterozygote PI $MZ_{Bristol}$ (T85M)

ADTHDEILEGLNFNLMEIPEAQIHEGFQELLR (1 oxidised methionine)

Peptide sequence 70-101



(c)

Figure 5.12 2D-PAGE and mass spectra of tryptic peptides covering amino acid sequence 70- 101 of α_1-antitrypsin from control and PI $MZ_{Bristol}$ plasma.

(a) 2D-PAGE of normal plasma α_1-antitrypsin and mass spectra of the tryptic in-gel digestion of the M6 isoform

(b) 2D-PAGE of normal plasma α_1-antitrypsin and mass spectra of the tryptic in-gel digestion of the M6 isoform after prior enzymatic removal of the glycans.

(c) 2D-PAGE of PI $MZ_{Bristol}$ plasma α_1-antitrypsin and mass spectra of the tryptic in-gel digestion of spot 3.

analysis of spot 1 but its pI and mass suggest that it is the equivalent of the normal M2 isoform lacking glycosylation at asparagine 83. Thus the mutant spots 1, 2, 3 and 4 observed in the PI $MZ_{Bristol}$ plasma, are the underglycosylated (at asparagine 83) equivalents of the M2, M4, M6 and M8 isoforms of α_1–antitrypsin. The reason for the absence of an equivalent to the M7 α_1–antitrypsin isoform in the PI $MZ_{Bristol}$ plasma is not clear. It could be that it is not secreted from the liver as in the case of PI Z phenotype where decreased amounts of mutant α_1–antitrypsin are observed in the plasma. However, only the mutant M7 isoform was absent from the plasmas of the PI $MZ_{Bristol}$ heterozygotes, whereas all other isoforms were detected, albeit in decreased amounts.

5.6 Discussion

The work described in this chapter demonstrates clearly the usefulness and power of proteomics as a tool in the detection and further understanding of genetic metabolic diseases. However, it is important to recognise the limitations of the technique. This study showed that it is possible to detect known alterations in the amino acid sequence of a protein by using MALDI TOF MS to detect the presence or absence of a peptide which is not present in the digestion of the normal protein. Reliable detection of unknown amino acid substitutions will require sequencing of all the peptides generated by proteolytic digestion of the protein. Although MALDI TOF MS has the capability of sequencing peptides by post-source decay analysis, sequencing can be carried out more efficiently and sensitively with a quadrupole-time of flight mass spectrometer such as the Q-TOF or ion-trap mass spectrometer.

The data generated in this study were essentially qualitative. To derive the maximum information from proteomic analysis it will be important to quantitate the level of expression of all the important proteins. Methods need to be developed that ensure that all peptides are sequenced and measured quantitatively for the reliable detection of mutations and measurement of the level of expression of proteins in a metabolic pathway.

Chapter 6

The identification and analysis of underglycosylated glycoproteins in the plasma of patients with CDG-I syndrome

Chapter 6 - The detection, identification and analysis of underglycosylated glycoproteins in the plasma of patients with CDG-I syndrome.

Contents

6.1 Introduction – the congenital disorders of glycosylation type-I (CDG-I).

Over the last two decades a number of inherited human diseases have been reported which cause impaired N-glycosylation of proteins. These include galactosaemia, hereditary fructose intolerance, leukocyte adhesion deficiency type II (LAD II), hereditary erythroblastic multinuclearity with positive acidifed serum lysis test (HEMPAS) and the congenital disorders of glycosylation (CDG), which are a clinically and genetically heterogeneous group of multi-system disorders. CDG was first described by Jaeken *et al.*, (1980) as a case of decreased serum thyroxine-binding globulin and increased arylsulphatase A activity in two patients with familial pyschomotor retardation. In 1984, Jaeken and colleagues showed a sialic acid deficiency in serum and cerebrospinal fluid transferrin, and proposed for the first time that the CDG syndrome was due to defective N-glycosylation (Jaeken *et al.*, 1984). Subsequently it was shown that the abnormal IEF pattern found in serum transferrin and other serum glycoproteins was due to the absence of N-linked glycan(s) hence causing a decrease in the negative charge of the molecule due to the absence of sialic acid residues (Stibler and Jaeken, 1990). Consequently, the disease was designated carbohydrate deficient glycoprotein syndrome. Van Schaftingen and Jaeken (1995) were first to report the exact biochemical cause of the syndrome described by Jaeken in 1984 was a deficiency of phosphomannomutase, an enzyme which converts mannose-6-phosphate to mannose-1-phosphate. The mannose-1-phosphate is a precursor of GDP-mannose synthesis, which in turn is required for the synthesis of the lipid-linked oligosaccharide (LLO, $Glc_3Man_9GlcNAc_2$-PP-dolichol). The decrease in the synthesis of LLO leads to underglycosylation of proteins, the hallmark of CDG-I. This group of diseases is now designated as the congenital disorders of glycosylation disorders Type I and some 7 distinct genetic disorders (types 1a-g) have been described to date, all of which effect the synthesis of LLO. Other variants of CDG which involve defects in the processing of protein N-linked oligosaccharides in the ER and Golgi and are called CDG-II.

At present some 300 Type I patients have been diagnosed with these defects worldwide (Grünewald *et al.*, 2002, Schachter 2001b, Chantret *et al* 2002; Grubenmann *et al* 2002 Table 6.1).

N-glycan assembly (CDG-I)	Nomenclature
Phosphomannomutase deficiency	a
Phosphomannose isomerase deficiency	b
Glucosyltransferase I deficiency	c
Mannosyltransferase VI deficiency	d
Dolichol-phosphate-mannose synthase-1 deficiency	e
Mannose-P-dolichol utilisation defect	f
Dolichol-phosphate-mannose : $Man_7GlcNAc_2$-PP-mannosyltransferase deficiency	g
N-glycan processing (CDG-II)	**Nomenclature**
N-acetylglucosaminyltransferase II deficiency	Ia
Glucosidase I deficiency	Ib
GDP-fucose transporter deficiency	Ic
UDP-galatose:N-acetylglucosamine β-1,4-galactosyltransferase I	Id

Table 6.1 Table of the known metabolic defects causing CDG.

All CDG-I disorders are defects in the synthesis of the lipid linked oligosaccharide and lead to a reduced supply of LLOs to meet the N-glycosylation requirement of the cell. This results in the underglycosylation of N-glycoproteins. The underglycosylation of serum N-glycoproteins and of transferrin in particular is used as the basis for the primary screen in the detection of CDG patients. The absence of a N-linked complex glycan structure

results in a change in the pI of the glycoprotein due to a decrease in the number of negatively charged sialic acids. Thus, the absence of a single complex biantennary glycan structure from a glycoprotein would result in the loss of 2 charged sialic acid residues, which can be detected by IEF.

6.1.1 Experimental and theoretical rationale.

The major aims of the work presented in this chapter were:

(i) To investigate the underglycosylation of plasma N-glycoproteins from CDG-Ia, Ic and Ix patients.

(ii) To investigate the site-specific N-glycosylation or lack of N-glycosylation in partially N-glycosylated isoforms of serum transferrin and α_1–antitrypsin from CDG-Ia, Ic and Ix patients.

The initial step in the protocol for the detection of underglycosylation of plasma N-glycoproteins is analysis of shifts in mass and pI values after 2D-PAGE. Many changes in the N-glycosylation of glycoproteins can be readily detected using 2D-PAGE by characteristic losses in mass and increases in pI. This phenomenon was demonstrated in the analysis of the $PIZ_{Bristol}$ heterozygote, whereby a T85M substitution resulted in the removal of a glycosylation sequon in the α_1–antitrypsin molecule (Chapter 5.5). The absence of a single complex glycan from the α_1–antitrypsin molecule led to a decrease in the mass and change in pI of the protein observed on the 2D-PAGE. Therefore protein spots with characteristic changes on 2D-PAGE were excised from the gel and the N-glycoproteins identified by peptide mass-mapping. Investigation of the site-specific glycosylation was performed using the techniques developed in Chapter 4.3. In the case where the protein of interest was not fully separated by 2D-PAGE over a wide focusing-range, another 2D-PAGE analysis will be performed over a narrower range of pH or alternatively the protein was purified before analysis by PAGE.

6.1.2 CDG-I patients analysed in this study.

Plasma was obtained from patients, who had been diagnosed with CDG-I on the basis of an abnormal isoelectric focusing profile for transferrin, ± specific enzyme deficiency or ± DNA analysis (Keir *et al.*, 1999). Patient 1 was diagnosed as CDG-Ia on the basis of a phosphomannomutase deficiency and a defect in the PMM2 gene. Patient 2 was confirmed as CDG-Ic on the basis of a mutation in the human ALG6 gene. Patient 3 had a typical IEF-pattern of serum transferrin as in CDG-I but the biochemical or genetic basis has not been defined and is denoted as CDG-Ix.

6.2 The identification of underglycosylated proteins in CDG-I plasma.

6.2.1 2D-PAGE (pH 4-7) analysis of CDG plasma proteins.

Plasma samples from age-matched controls and CDG-I patients were subjected to 2D-PAGE with focusing over the pH range 4-7 and the proteins visualised using silver-staining (Figure 6.01a, b, c and d). The N-glycoproteins of interest were identified by peptide mass mapping.

Among the many plasma N-glycoproteins that displayed characteristic signs of aberrant glycosylation were α_1–antitrypsin, α_1–antichymotrypsin, fetuin, transferrin, $\alpha_1\beta$-glycoprotein, hemopexin and complement factor B. The amount of the N-glycoprotein, haptoglobin I, was found to be decreased markedly in the plasma sample of the CDG-Ia patient and completely absent in the CDG-Ic and Ix patient as compared with the control sample (Figure 6.01). Haptoglobin I is a N-glycoprotein and it is unclear if the decreased amounts of this protein observed in the CDG plasma are due to defects in its N-glycosylation that subsequently target the nascent misfolded haptoglobin I to the proteasomal degradation pathway. Haptoglobin I is an acute phase protein, which is up-regulated during an inflammatory response, with its main role being haemoglobin-binding to protect against haemoglobin-induced renal damage during haemolysis. The samples obtained from the CDG patients did not show

any signs of haemolysis which could result in the loss of the haptoglobin fraction. 2D-PAGE analysis of the second PI $MZ_{Bristol}$ heterozygote shown in Figure 5.11c, Chapter 5, also showed the absence of the haptoglobin fraction.

The results for plasma glycoproteins (Figure 6.01) clearly demonstrate that 2D-PAGE is an extremely powerful technique for identifying changes in the level of expression and in the post-translational modifications of N-glycoproteins. Changes in the number of N-linked complex glycans have been shown to be particularly amenable to identification by 2D-PAGE. A characteristic increase in pI of 0.1 or 0.15 pI units, respectively, is observed when a bi- or triantennary complex glycan is absent. The absence of a complex glycan also results in a decrease in mass of ~2.2 kDa or 2.8 kDa for a bi- or triantennary complex glycan, respectively.

Figure 6.01

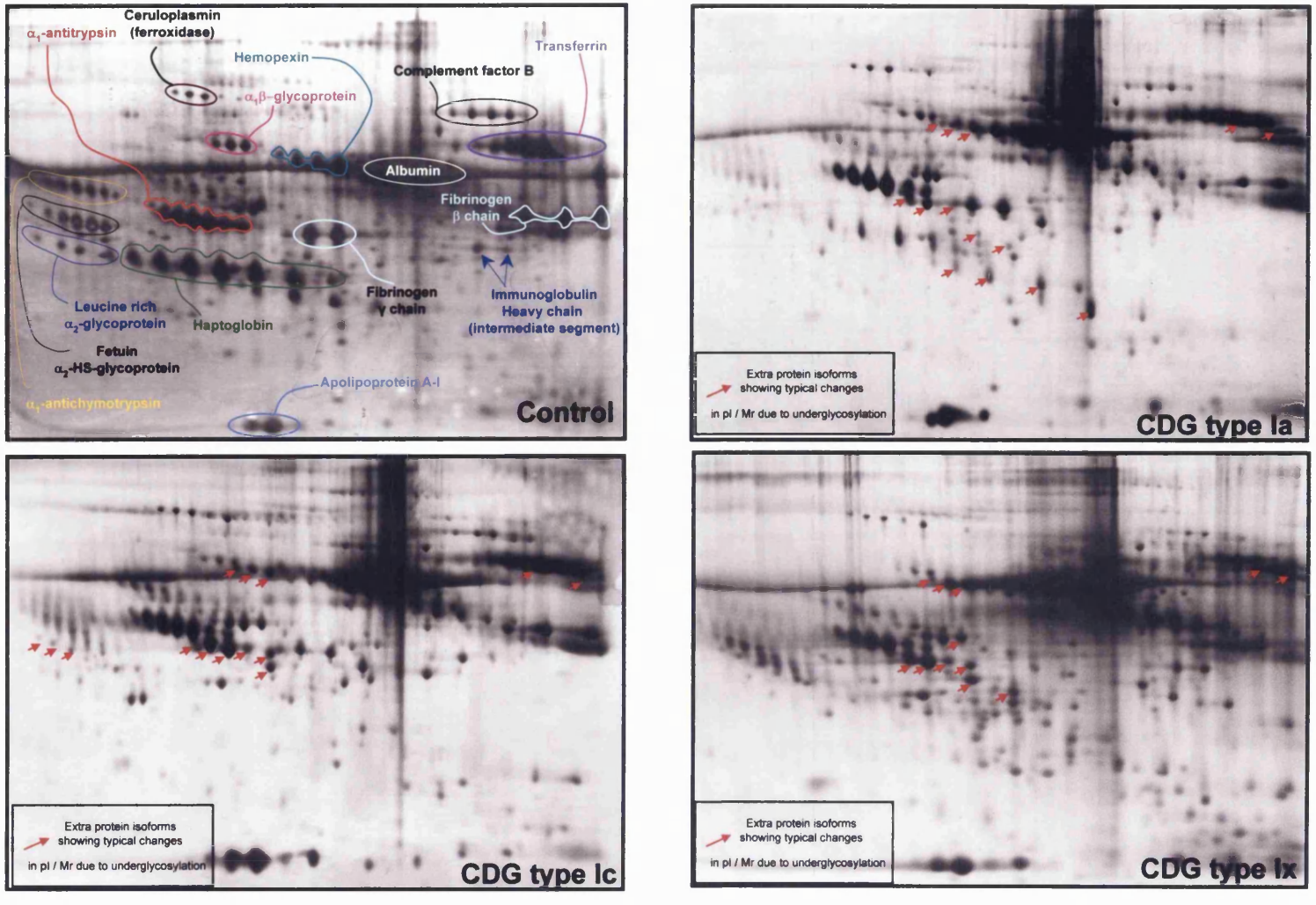


Figure 6.01 2D-PAGE (pH 4-7) of plasma from control and CDG-Ia, Ic and Ix patients

6.3 The determination of the site-specific glycosylation of plasma proteins purified from plasma.

The 2D-PAGE analyses demonstrated that many classes of N-glycoproteins were underglycosylated in the CDG-I syndromes. Although, the data from the peptide mass mapping studies were adequate to confirm the identity of each underglycosylated N-glycoprotein, they could not provide precise information on the site-specific glycosylation of each isoform. It was therefore decided to focus on two N-glycoproteins, namely α_1-antitrypsin and transferrin. These N-glycoproteins were chosen because of their relatively high concentration in plasma and also because their N-glycosylation has been investigated in detail and described extensively in the literature (see Chapter 1). Both of the N-glycoproteins were purified further using specific analytical techniques in order to analyse the site-specific glycosylation of their underglycosylated isoforms in CDG-I.

6.3.1 Reasons for the choice of plasma transferrin from CDG patients.

Transferrin is present in the plasma at a concentration of approximately 2-4 mg/ml of plasma and contains two N-linked glycosylation sites at asparagines 413 and 611 in the polypeptide sequence. The glycans present on these asparagine residues are predominantly biantennary structures although as much as 17% of plasma transferrin can contain one bi- and one triantennary glycan (see Chapter 1).

The site-specific glycosylation of transferrin in CDG-Ia patients has been described previously by several groups (Henry *et al.*, 1999) and has been claimed to be a random phenomenon. It was therefore considered a good benchmark protein for the evaluation of the proteomic techniques developed in the present study to investigate the site-specific glycosylation of plasma N-glycoproteins.

6.3.1.1 2D-PAGE analysis of plasma transferrin.

The analysis of plasma transferrin by 2D-PAGE focused over the pH-range 4-7 in the initial experiments did not completely resolve the fully, partially and non-glycosylated isoforms of transferrin. To overcome this problem, plasma samples from a control and each CDG patient, were subjected to 2D-PAGE over a narrow range of focusing, pH 6-9, to obtain better resolution (Figure 6.02).

Three isoforms of transferrin were observed, of approximate molecular weights of 79 kDa, 77 kDa and 75 kDa (Figure 6.02). The identity of each isoform was confirmed by in-gel tryptic digestion and MALDI TOF MS as being transferrin.

The highest molecular weight transferrin isoforms in the CDG patients had approximately the same molecular weight as the isoforms in control transferrin (79 kDa). This indicated that the high molecular weight transferrin isoforms in the CDG patients were the fully i.e. diglycosylated transferrin isoforms.

The intermediate molecular weight transferrin isoforms observed on the 2D-PAGE, had an Mr of approximately 77 kDa and were approximately 2.5 kDa lighter than the fully glycosylated transferrin series. This difference in mass approximates to a single biantennary glycan structure and indicated that the intermediate molecular weight isoforms are the monoglycosylated forms of transferrin.

The lowest molecular weight series of transferrin isoforms had an Mr of approximately 75 kDa. This isoform was approximately 5.0 kDa and 2.5 kDa lighter in mass than the fully and monoglycosylated transferrin isoforms, respectively and indicated that the lowest molecular weight form was the non-glycosylated transferrin.

Although the three patients representing the different sub-types of CDG-I analysed in this study were shown to have fully, partially and non-glycosylated transferrin isoforms present in the plasma, the relative amounts of the three transferrin isoforms were not similar (also see section 6.3.1.4). The 2D-PAGE analysis of plasma transferrin from the CDG-Ia patient indicated that the fully glycosylated isoform of transferrin was predominant whilst the monoglycosylated and non-glycosylated isoforms were present in decreasing and considerably lower amounts (Figure 6.02 (i)). The CDG-Ic patient had approximately equal amounts of the fully and monoglycosylated isoforms of

transferrin, with almost no non-glycosylated transferrin being detected (Figure 6.02 (ii)). However, the predominant isoform detected in the plasma from the CDG-Ix patient was monoglycosylated transferrin, followed by the non-glycosylated species with the fully glycosylated isoform of transferrin present in the lowest concentration.

Figure 6.02

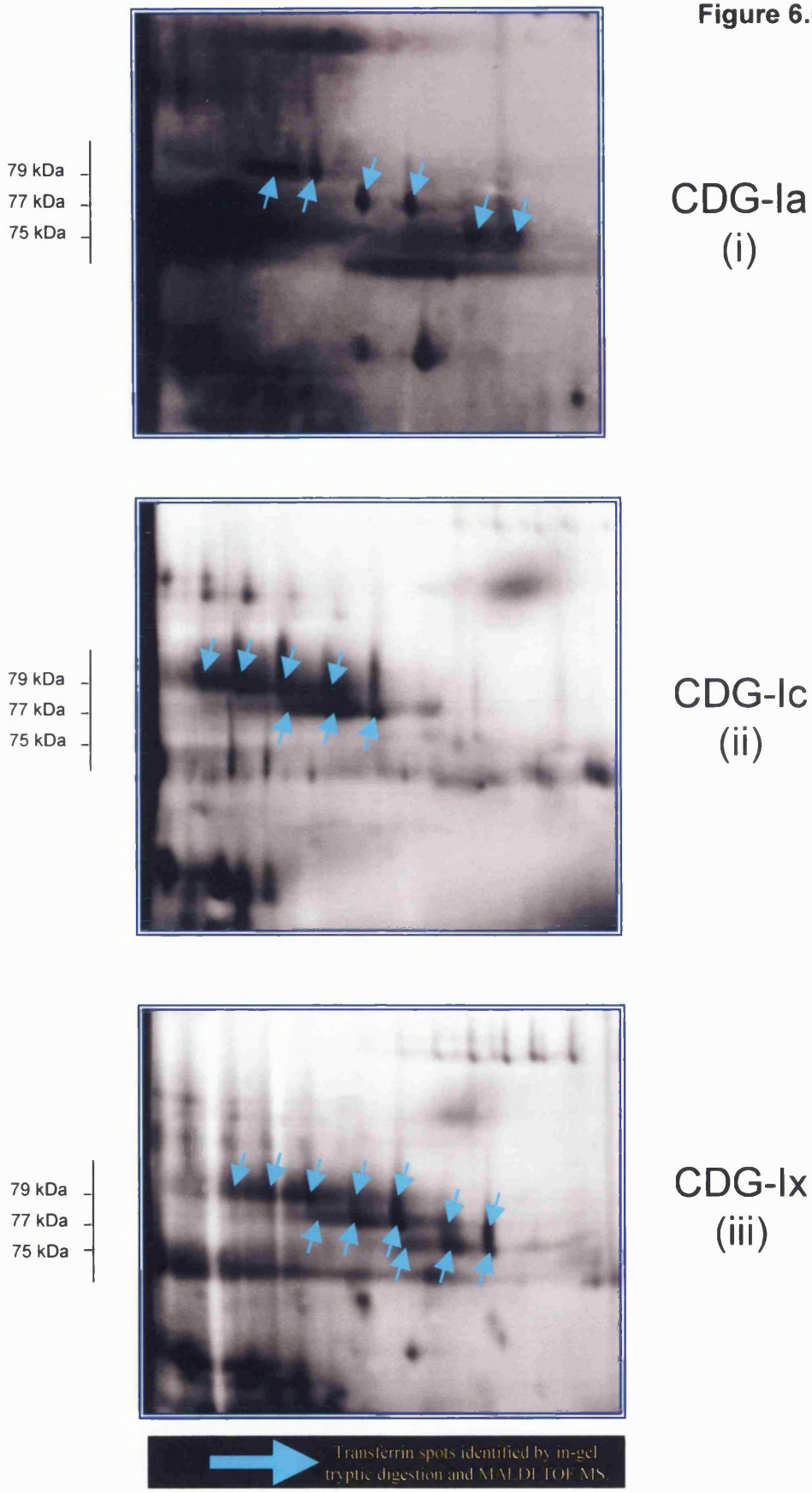


Figure 6.02 High-resolution 2D-PAGE (pH 6-9) of plasma transferrin from; (i) CDG-Ia patient. (ii) CDG-Ic patient. (iii) CDG-Ix patient.

6.3.1.2 Identification and analysis of transferrin isoforms by MALDI TOF MS.

6.3.1.2.1 Fully glycosylated transferrin isoforms (79 kDa).

Mass spectral analysis of the tryptic peptides from the high molecular weight isoforms of transferrin did not reveal any masses corresponding to glycopeptides or peptides covering either of the glycosylation sites at asparagine residues 413 or 611 in any of the CDG-I and control plasma samples.

However, release of the glycan structures using N-glycanase prior to tryptic in-gel digestion resulted in the presence of two extra peptide ions of mass 2499.1 m/z and 2516.1 m/z (Figure 6.03). These masses correspond to the deglycosylated glycopeptide in which asparagine 611 has been converted into aspartic acid by N-glycanase, indicating that this site is occupied by a glycan. The mass 2499.1 m/z is the pyroglutamate derivative of the peptide 603-623 and the mass 2516.1 m/z is the corresponding native peptide mass. Pyroglutamate derivatives, which are often observed during MALDI TOF MS analyses along with the native peptide mass, can result from conversion of the N-terminal glutamine to pyroglutamate during the work up procedures or during the ionisation process (Thiede *et al.*, 2000).

As in the studies of the site-specific glycosylation of transferrin in solution (Chapter 4), no peptides corresponding to a glycosylated or deglycosylated tryptic peptide containing asparagine 413 were detected with or without predigestion with N-glycanase. Further, no peptides covering this region were detected when chymotrypsin was used as the protease. These results differ from those obtained in Chapter 4, which indicated that a combination of the release of the glycan structures using N-glycanase with a subsequent chymotryptic digest resulted in peptides covering the 413 sequon being detected by MALDI TOF MS. The sequon at asparagine residue 413 could not be monitored using the combination of N-glycanase and protease treatment when analysed in-gel whereas it could be after digestion in-solution.

Figure 6.03

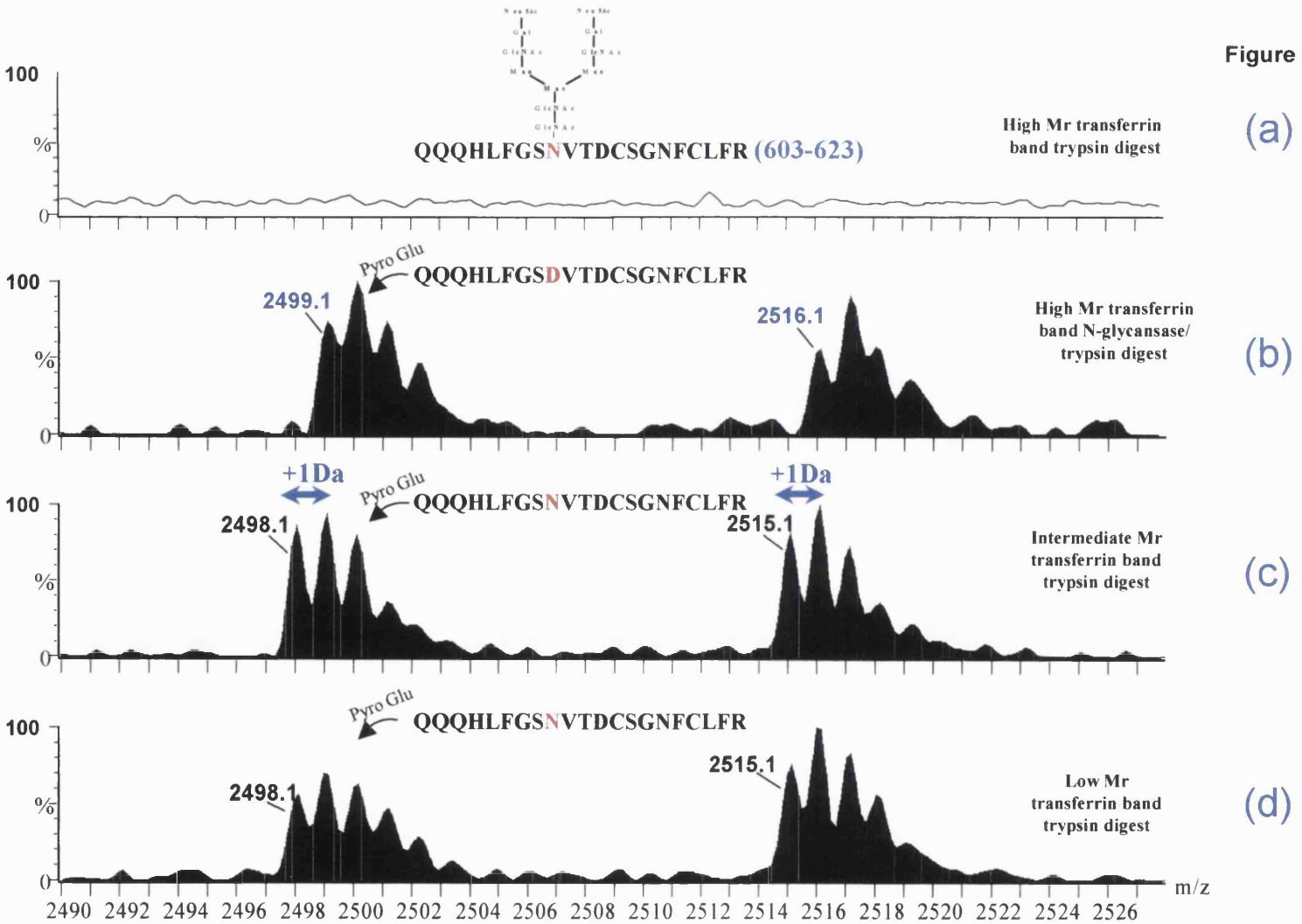


Figure 6.03 Monitoring the glycosylation of Asn 611 in transferrin by mass spectral analysis of tryptic peptides after in-gel digestion of;
(a) Fully glycosylated transferrin (b) deglycosylated diglycosylated transferrin
(c) Deglycosylated monoglycosylated transferrin (d) non-glycosylated transferrin

6.3.1.2.2 Monoglycosylated transferrin isoforms (Intermediate molecular weight - 77 kDa).

Analysis of the tryptic peptides from the in-gel digestion of each of the monoglycosylated transferrin isoforms showed the presence of three extra masses that were not observed in the fully glycosylated transferrin analyses. These ions were:-

i) 1476.8 m/z, This corresponds to peptide 402-414 with asparagine at the glycosylation site. It was also seen in the digests of isoforms from the non-glycosylated transferrin band (Figure 6.04b). Its presence confirms the fact that when there is an asparagine at position 413, a peptide is seen but when (following PNGase F treatment of the glycosylated arginine) there is an aspartic acid at position 413, no peptide is seen.

ii) 2498.1 m/z which correspond to peptide 603-623 with asparagine at glycosylation site 611 (Figure 6.03c).

iii) 2515 m/z which is the same peptide as ii) with the terminal glutamine converted to pyroglutamate (Figure 6.03c).

To investigate whether the glycosylation of each sequon was random and not specific, the monoglycosylated transferrin isoforms were subjected to in-gel tryptic digestion after pre-digestion with N-glycanase to demonstrate the conversion of the glycosylated sequon asparagine to aspartic acid residue using mass spectrometry. If Asn 611 was glycosylated, one could expect an increase in the ratio of 2499.1 / 2498.1 after PNGase F treatment (Figure 6.03). If Asn 413 was glycosylated no change could be expected because the aspartate-containing peptide is not seen.

No change was seen in the ratio of 2499.1 / 2498.1 following PNGase F treatment suggesting that Asn 611 was almost entirely non-glycosylated (a control in-gel analysis using α_1-antitrypsin confirmed that the N-glycanase reaction has worked efficiently). The fact that a significant peak corresponding to unglycosylated peptide 402-414 was seen (with and without PNGase F

treatment) indicated that the monoglycosylated band contained some transferrin with Asn 413 unglycosylated. Thus the conclusion must be that the monoglycosylated band probably contains more transferrin with Asn 611 unglycosylated than transferrin with Asn 413 unglycosylated but both species were present.

Figure 6.04

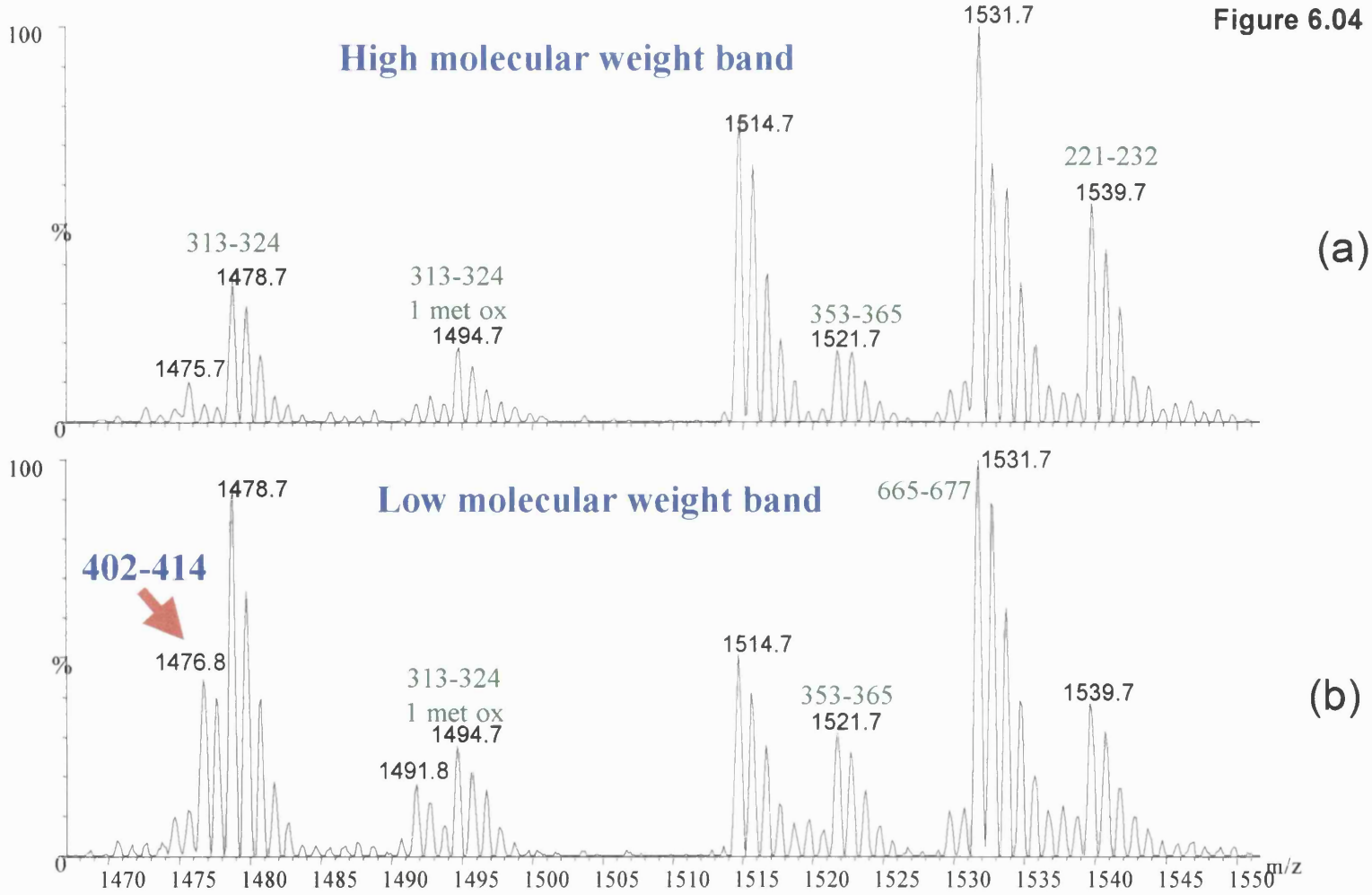


Figure 6.04 Monitoring the glycosylation of Asn 413 in transferrin by mass spectral analysis of tryptic peptides after in-gel digestion of (a) fully glycosylated transferrin (b) non-glycosylated transferrin.

1476.8 m/z corresponding to peptide 402-414 containing asparagine at residue 413

6.3.1.2.3 Analysis of the non-glycosylated transferrin isoforms (low molecular weight - 75 kDa).

As in the analysis of the intermediate molecular weight isoforms of transferrin, three extra peptide masses (1476.8 m/z, 2498.1 m/z and 2515.1 m/z) were observed in all of the non-glycosylated transferrin isoforms in the CDG-I plasma samples (Figures 6.03 and 6.04). The masses correspond to the peptides 402-414 and 603-623 containing asparagine residues at 413 and 611, respectively. The pre-treatment of each transferrin isoform with N-glycanase prior to tryptic in-gel digestion, did not result in any change in the isotopic ratio of the ions for peptide 603-623 and no disappearance of peptide 402-414 as would be predicted for non-glycosylated transferrin.

6.3.1.3 Discussion of findings for plasma transferrin.

The use of high resolution 2D-PAGE improved the separation and increased the resolution of each transferrin isoform detected, compared with focusing over the wider range of pH 4-7. However, this posed more questions as several more isoforms of transferrin were detected than anticipated. In Chapter 5, it was demonstrated that each isoform of α_1–antitrypsin resolved by 2D-PAGE was separated by 0.05 pI units due to increasing amounts of sialic acid. The transferrin analysis also demonstrated approximate pI differences of 0.05 pI units between each isoform, indicating that this could be explained by increases in the branching of the carbohydrate side chains in the fully and monoglycosylated transferrin isoforms as observed in the α_1–antitrypsin analyses (Chapter 5). However, the detection of three distinct isoforms of transferrin for the non-glycosylated transferrin remains difficult to interpret since the separation cannot be accounted for by glycan structures alone. The binding of Fe^{2+} ions to transferrin can affect its isoelectric focusing pattern but under the experimental conditions of the SDS-PAGE which are denaturing, the Fe^{2+} binding capacity of the molecule is almost certainly disrupted. The IEF of

plasma transferrin is routinely used in the diagnosis of CDG but no information has been reported on the presence of extra isoforms of non-glycosylated transferrin during these analyses.

The mass spectral analysis of monoglycosylated series of isoforms indicated that the glycosylation was preferentially on the asparagine 413 residue, the sequon nearer the N-terminus. However, using the methods developed here it was not possible to confirm these preliminary results as the glycosylation of asparagine residue 413 could not be directly monitored. Although a mass that corresponded to a peptide that covered the first sequon (Asparagine 413) in the monoglycosylated and non-glycosylated transferrin series was detected, the removal of the glycans using N-glycanase did not result in the anticipated +1 Da increase in the mass due to the conversion of an asparagine to an aspartic acid. However such a change of 1 Da was detected in the analysis of the fully glycosylated transferrin isoforms, following the removal of a complex glycan from the second glycosylation site, asparagine 611. This discrepancy could be explained by the conversion of the asparagine to an aspartic acid residue in the peptide 402-414, to produce a peptide which is not amenable to analysis by MALDI TOF MS. This hypothesis would explain why the peptide 402-414 containing an asparagine residue at the glycosylation sequon is detected in the intermediate and low molecular weight transferrin analyses. Another possible explanation is that the conversion of an asparagine to an aspartic acid at amino acid 413 in the sequence, results in the insertion of a negatively charged amino acid next to a tryptic cleavage site. This may make the tryptic cleavage between amino acids 414 and serine 415 less favourable as trypsin preferentially cleaves at basic amino acid residues. A failure to cleave after lysine 414 would result in larger peptides consisting of amino acids 402-433 and 402-434, with masses of 3530.9 m/z or 3659.1 m/z, respectively. However, no peptides of these masses were detected in the analyses.

The tryptic digestion of the transferrin molecule followed by removal of the glycan structures using N-glycanase, also did not result in an ion being detected that corresponded to a peptide containing an aspartic acid at residue 413. These results support the hypothesis that the conversion of an asparagine to an aspartic acid at residue 413 produces a peptide that is not amenable to analysis by MALDI TOF MS.

The treatment of the monoglycosylated transferrin isoforms with N-glycanase prior to tryptic digestion, also failed to show any significant increase in the isotopic ratio of ions of the sequon containing peptide 603-623 (Asn 611), which could be used as an indirect marker of the glycosylation status of the first glycosylation site (Asn 413). These data suggest that the monoglycosylated transferrin fraction contains some molecules with Asn 413 glycosylated and some with Asn 611 glycosylated, but that the first site (Asn 413) is preferentially occupied.

6.3.1.4 1D-PAGE analysis of plasma transferrin

6.3.1.4.1 Rationale for the rivanol purification and 1D-PAGE analysis of plasma transferrin.

The results of the 2D-PAGE analyses of plasma transferrin from CDG-I patients were difficult to interpret because it was not possible to explain the different isoforms of the non-glycosylated transferrin. It was therefore decided to purify transferrin from the plasma of each CDG-I patient using the rivanol precipitation procedure (Chapter 2). The transferrin was then separated further into the fully, mono- and non-glycosylated isoforms by 1D-PAGE, to allow in-gel tryptic digestion of the intermediate or monoglycosylated transferrin isoform from each patient. If it was true that the conversion of the asparagine 413 to an aspartic acid residue resulted in the formation of a peptide 402-414 which couldn't be detected by mass spectrometry, then by directly comparing the response of the peptides covering each glycosylation site in the monoglycosylated band with the non-glycosylated band (which are both theoretically unoccupied), it might be possible to determine the relative ratio of site occupancy for transferrin.

6.3.1.4.2 The rivanol purification and 1D-PAGE analysis of plasma transferrin.

Plasma transferrin was purified from each CDG-I patient using the rivanol chemical precipitation technique (Chapter 2). Any residual salts were removed prior to 1D-PAGE using centrifugal filtration as described in Chapter 2. Each

sample was reduced to dryness by freeze-drying and the transferrin isoforms separated according to molecular weight by 1D-PAGE.

Figure 6.05a shows the 1D-PAGE analysis of the rivanol-purified transferrin from each CDG-I patient. The 1D-PAGE analysis allowed the complete separation of the individual isoforms into fully (79 kDa), mono- (77 kDa) and non-glycosylated (75 kDa) protein bands. Although the rivanol-purification of transferrin was developed for the purification of fully glycosylated transferrin, the results presented here demonstrate that the glycosylation status of the molecule had little or no effect on the rivanol-purification of the protein. Comparison of the 1D-PAGE of rivanol-purified transferrin and the 2D-PAGE analysis showed that purification by both methods resulted in similar relative proportions of each isoform (Figure 6.02b).

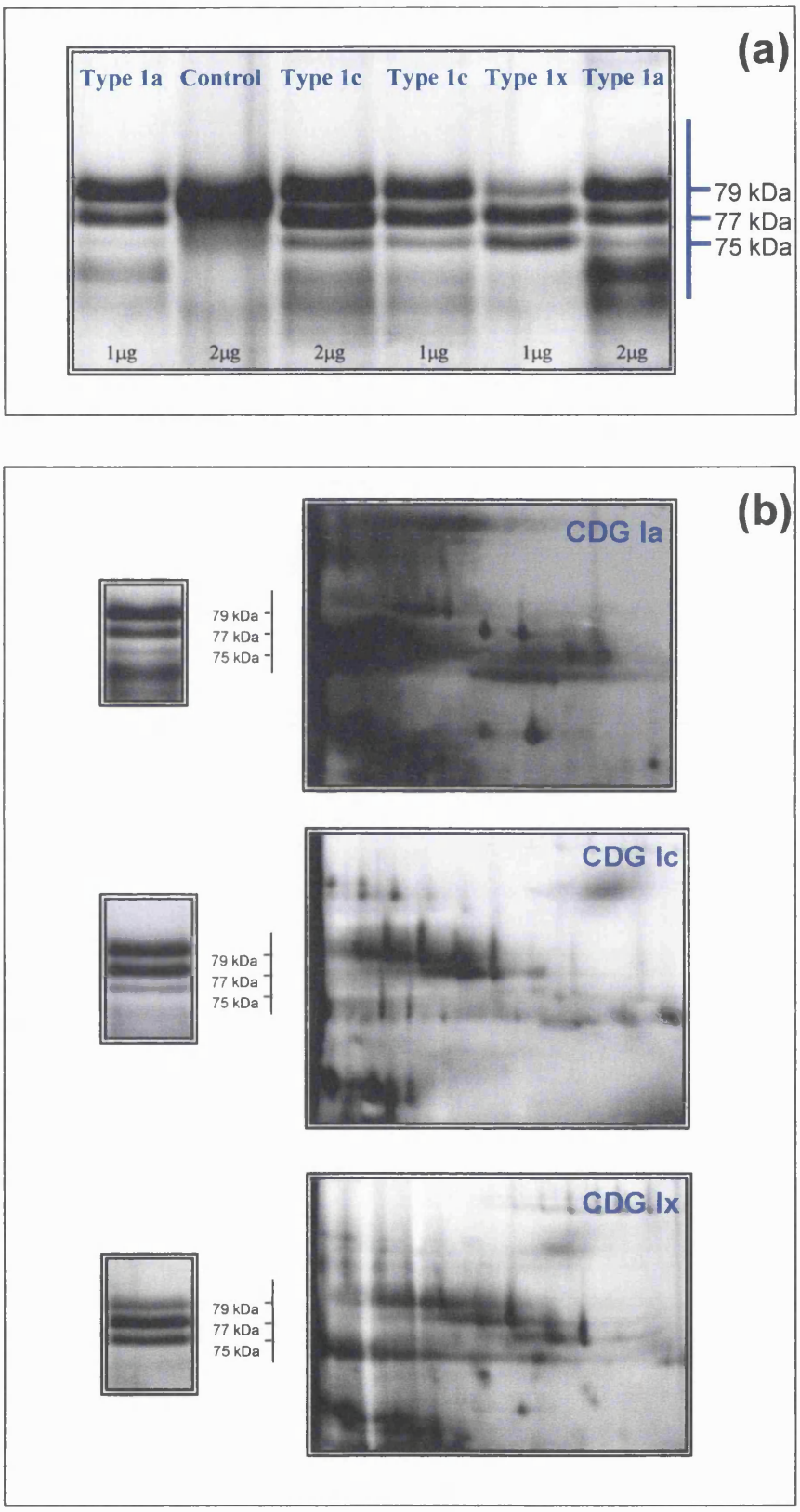


Figure 6.05 (a) 1D-PAGE separation of rivanol-purified plasma CDG transferrin.
(b) Comparison of rivanol-purified CDG transferrin and CDG transferrin purified using 2D-PAGE.

6.3.1.4.3 MALDI TOF MS identification and analysis of transferrin isoforms.

Each band of transferrin from the 1D-PAGE was excised and subjected to in-gel tryptic digestion according to Chapter 2. Figure 6.06 shows typical spectra obtained from the analysis of tryptic peptides from the fully, mono- and non-glycosylated species of transferrin. In agreement with the 2D-PAGE analyses of CDG plasma transferrin, no masses corresponding to peptides or glycopeptides covering either glycosylation sites at asparagine residues 413 or 611 were detected.

Analysis of the tryptic peptides from each of the monoglycosylated transferrin bands from each patient demonstrated the presence of the masses 1476.8 m/z, 2498.1 m/z and 2515.1 m/z that were also observed in the analysis of the monoglycosylated transferrin isoforms isolated by 2D-PAGE. These masses correspond to the peptide 402-414 containing Asn 413 and its pyroglutamate derivative, and the native 603-623 peptide containing Asn 611.
The release of the glycan structures using N-glycanase prior to tryptic digestion did not result in any extra masses or changes in the isotopic ratio of the peptides being detected. This also was in agreement with results obtained for the monoglycosylated isoforms of transferrin separated by 2D-PAGE.

Mass spectral analysis of the tryptic peptides from the non-glycosylated transferrin bands showed the presence of the ions 1476.8 m/z, 2498.1 m/z and 2515.1 m/z which corresponded to the peptides 402-414 and the pyroglutamate and native 603-623 peptide, respectively. However, the response of the ion of mass 1476.8 m/z, which corresponds to the peptide 402-414 with an asparagine at the glycosylation site, was significantly higher in the non-glycosylated band than in the monoglycosylated band analyses for each CDG-I patient analysed. This suggests that significant amounts of the sequon at asparagine 413 are occupied by a glycan in the monoglycosylated species of transferrin.

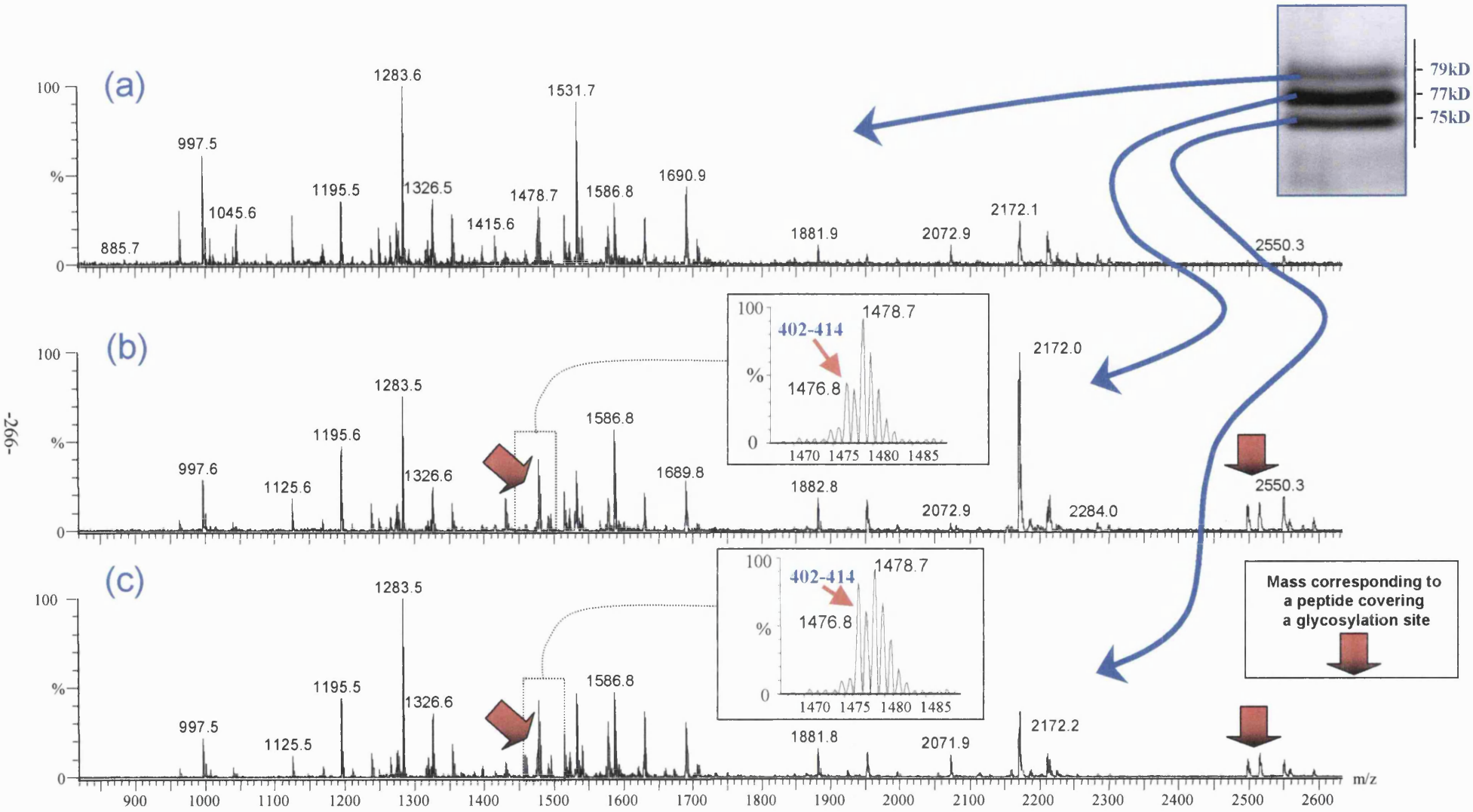


Figure 6.06 Mass spectral analysis of the peptides generated after in-gel tryptic digestion of rivanol purified transferrin from CDG-Ix plasma. (a) fully glycosylated transferrin (b) monoglycosylated transferrin (c) non-glycosylated transferrin

In an attempt to investigate the relative proportion of the site-occupancy between the two sequons in the monoglycosylated transferrin species, the mass spectral response for each peptide was plotted against the amino acid sequence and compared to the results obtained for the non-glycosylated species (Figure 6.07). The non-glycosylated band acts as a control for the analysis, since both sites are non-occupied (Figure 6.07 iv). The use of a fully glycosylated transferrin band that had been deglycosylated in-gel prior to tryptic digestion, was not suitable as a control due to the inability to detect the peptide 402-414 with an aspartic acid residue at the glycosylation sequon. Fortuitously, the total responses of the peptides covering each glycosylation site in the analysis of the non-glycosylated transferrin band were approximately equal. This result is rather serendipitous considering that both glycosylation sites were known not to be occupied by a glycan. Therefore the ratio of the responses for each peptide in the monoglycosylated transferrin should be a measure of the relative occupancy of each site.

The analysis of the tryptic peptides from the monoglycosylated transferrin isoforms in the CDG-Ia, Ic and Ix patients demonstrated that this ratio was not equal in the peptides covering each glycosylation site. The results shown in Figure 6.07 show that although both sites are partially glycosylated, the glycosylation site nearer the N-terminus (asparagine 413) was preferentially occupied. These results show that approximately 87%, 99% and 86% of the glycosylation site asparagine 611 was unoccupied in the monoglycosylated transferrin isoform for the CDG-Ia, Ix and Ic patients, respectively. This contradicts the conclusion of Henry and colleagues (1999) who found that the distribution of glycans between each sequon was approximately equal.

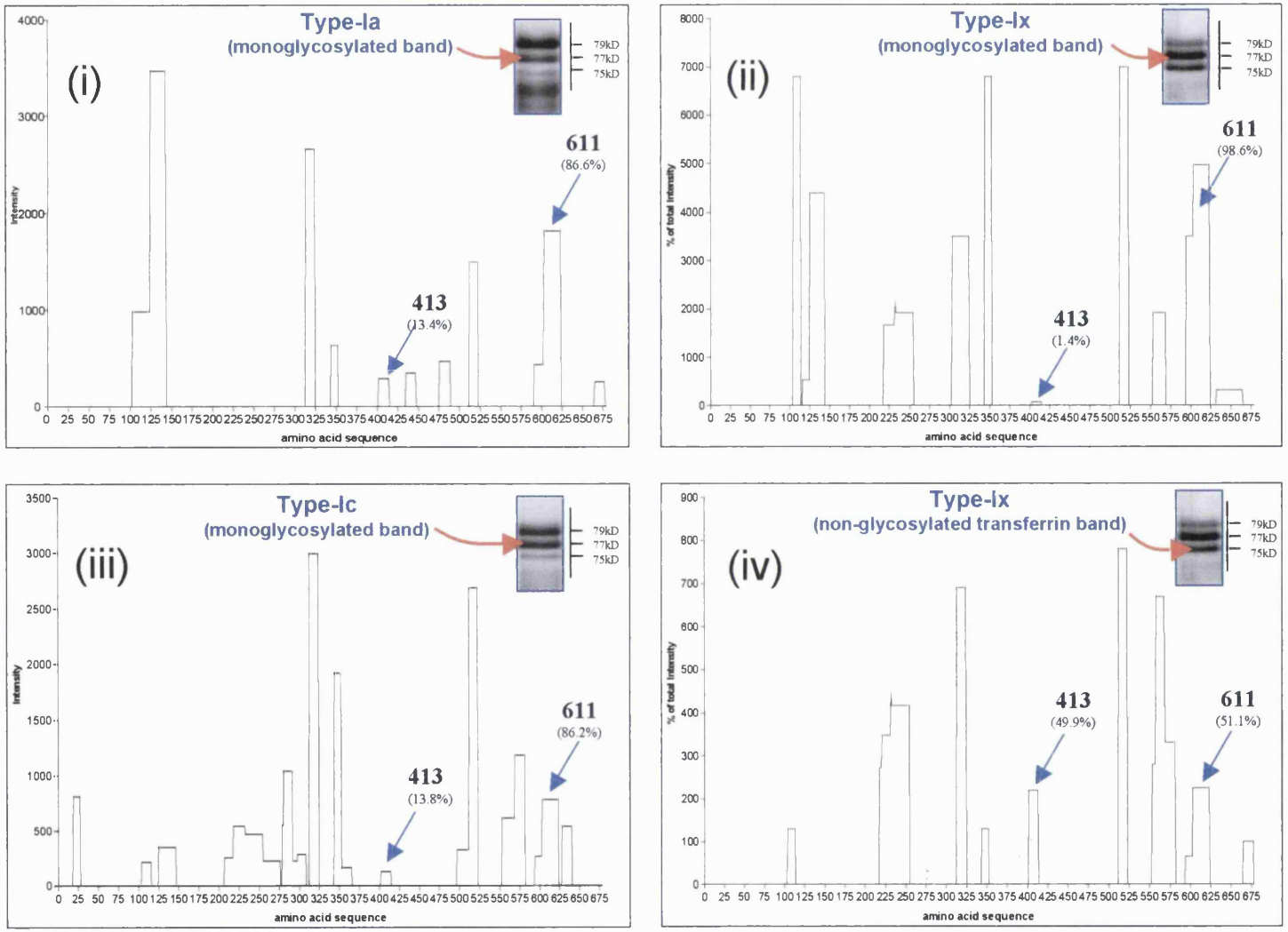


Figure 6.07

Figure 6.07 Sequence coverage and the mass spectral response of the peptides produced by in-gel tryptic digestion of;
(i), (ii) and (iii) monoglycosylated species of transferrin in CDG types Ia, Ix and Ic, respectively.
(iv) non-glycosylated species of transferrin in CDG types Ix.

6.3.2 Conclusions.

As in the 2D-PAGE analysis of transferrin, the N-glycosylation status of the first sequon at asparagine residue 413 could not be directly monitored in any of the CDG-I patients or controls. This was based on the inability of the technique to demonstrate a shift of +1 Da after the conversion of the asparagine to aspartic acid in peptide 402-414. Although a peptide of mass 1476.8 m/z, which corresponds to the mass of peptide 402-414 with an asparagine at the sequon, was detected in the mass spectral analyses of the mono- and non-glycosylated transferrin species, it could not be confirmed using the methods developed in Chapter 4. However, the ratio of the response of the peptides for both sequons in the monoglycosylated transferrin isoforms in each patient indicated that although both sequons were partially glycosylated, the glycosylation site nearer the N-terminus was preferentially occupied. Further confirmation of these findings was the failure to detect any significant changes in the isotopic ratio of the ions 1476.8 / 1477.8 (Sequon 413) and 2498.1 / 2499.1 (sequon 613), after deglycosylation using PNGase F. If the glycosylation of the transferrin was a completely random phenomenon then a change in the isotopic ratio in these ions should be observed after the conversion of asparagine to aspartic acid and the corresponding mass change. Although plotting of the responses of each peptide indicated that the glycosylation site 413 was preferentially occupied (Figure 6.07), it should be noted that these figures are only an approximate determination of the site-occupancy.

Mellquist and colleagues (1998) have shown that the glycosylation sites containing threonine as opposed to serine are more efficiently glycosylated. In the amino acid sequence of transferrin the glycosylation motif at Asn 413 contains a serine residue while the glycosylation motif at Asn 611 contains a threonine residue, indicating that Asn 611 should be more efficiently glycosylated and contradicting our findings. However, the methods developed in this work are better suited to determining whether a sequon is or is not occupied by a glycan and not to determining the proportion of each glycosylation site occupied (due to the nature of the MALDI process and changes in the sequence coverage that can be observed with varying protein

amounts). A more accurate way of determining the extent of glycosylation would be the analysis of the sequon containing glycopeptides or quantitation of the peptides using suitable internal standards and ESI-tandem mass spectrometry.

6.3.3 The analysis of N-glycosylation site-occupancy of plasma α_1–antitrypsin in CDG-I patients.

Transferrin, unlike α_1–antitrypsin, contains 19 disulphide bridges and therefore the glycans probably play a less important role in stabilising the conformation of the molecule than in α_1–antitrypsin which has no disulphide bridges. Furthermore, α_1–antitrypsin contains three glycosylation sites (asparagine residues 46, 83 and 247), which are N-glycosylated in normal plasma samples. In Chapter 5, it was shown how the M nomenclature of control plasma α_1–antitrypsin isoforms was assigned after their separation by 2D-PAGE. Using a combination of glycan and polypeptide analysis, it was possible to identify each isoform unequivocally with its M-series equivalent as described under the conditions of IEF (Jeppsson *et al.*, 1985). Using the techniques developed in Chapter 3 and 4, the underglycosylation of α_1–antitrypsin was investigated to determine whether the site-occupancy was a random or an ordered phenomenon in CDG-I patients.

Previously it had been shown that N-glycanase could be used to determine the site-occupancy of glycosylation sites by monitoring the conversion of an asparagine residue to an aspartic acid. However, the results obtained in section 6.3.1.2 and Chapter 4 showed there was significant difference between the amino acid sequence coverage obtained with the peptide mass-mapping when the proteolytic reaction is carried out in-solution or in-gel. Although it was not possible to monitor the site-occupancy of the first sequon in transferrin, it was possible to monitor all three sequons in α_1–antitrypsin using the strategy developed in this project.

6.3.3.1 2D-PAGE separation of plasma α_1–antitrypsin.

Plasma samples from CDG-I patients were analysed by 2D-PAGE with IEF over a pH-range of 4-7. Three additional series of α_1–antitrypsin were identified by peptide mass fingerprinting in the 2D-PAGE analyses of the plasma from the CDG-I patients (Figure 6.08). The molecular sizes of the additional isoforms

were approximately 2.5, 5 and 7.5 kDa less than the M series. The mass losses and changes in pI observed on the 2D-PAGE suggested that these extra series of α_1–antitrypsin were the diglycosylated, monoglycosylated and non-glycosylated isoforms, respectively (Figure 6.08b, c & d). To facilitate the description of their analysis, each isoform has been designated by a small Roman numeral: diglycosylated (i) to (v), monoglycosylated (vi) and non-glycosylated (vii) (Figure 6.08 and Table 6.1).

Isoform	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)
pI	**5.05**	**5.10**	**5.15**	**5.21**	**5.26**	**5.26**	**5.36**
Estimated Mr	**49.4**	**49.0**	**48.7**	**48.7**	**48.2**	**46.5**	**44.3**
CDG-Ia	**X**	✓	✓	**X**	✓	✓	✓
CDG-Ic	✓	✓	✓	✓	✓	✓	✓
CDG-Ix	**X**	✓	✓	✓	✓	✓	✓
✓ = isoform detected, X = isoform not detected							

Table 6.2 Nomenclature and characteristics of different isoforms of α_1–antitrypsin in CDG-I.

6.3.3.2 Analysis of the fully glycosylated α_1–antitrypsin isoforms by MALDI TOF MS.

The fully glycosylated isoforms of α_1–antitrypsin for each CDG-I patient were identified and designated as the M series of α_1–antitrypsin (Figure 6.08). Using the pI and Mr co-ordinates for each of M series of α_1–antitrypsin isoforms, combined with the pI / Mr coordinates of other proteins identified on the 2D-PAGE by mass mapping, it was possible to create a highly accurate pI /Mr co-ordinate system for the 2D-PAGE. This system was used to detect changes in pI / Mr for comparison between each gel and each patient.

Figure 6.08

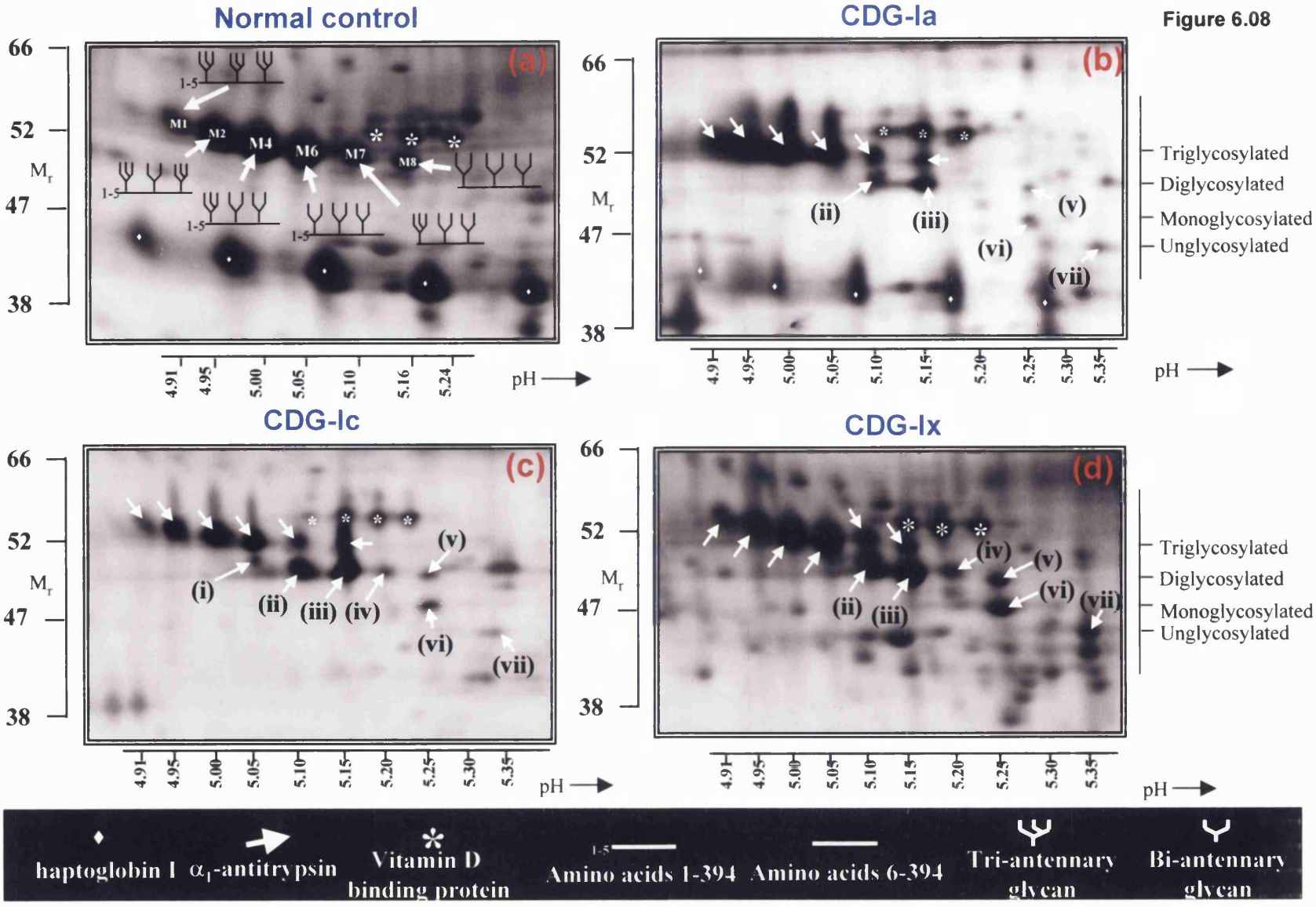


Figure 6.08 The 2D-PAGE separation of plasma α_1-antitrypsin from (a) normal, (b) CDG-Ia, (c) CDG-Ic and (d) CDG-Ix patients

6.3.3.3 Analysis of the diglycosylated α_1–antitrypsin isoforms (spots i to v) in the plasma of a patient with CDG-I.

6.3.3.3.1 Polypeptide analyses

Analysis of the tryptic peptides i to v confirmed that all the isoforms were α_1–antitrypsin molecules (Figures 6.08b, c and d). The reporter sequence for the N-terminus, 2819.2 m/z, was detected in components i, ii and iii but not in iv and v. The mass spectra were analysed for peptides containing the three glycosylation site asparagines, 46, 83 and 247. A mass of 3691.8 m/z was detected in all of the isoforms i to v (Figure 6.09b), which was not observed in the analysis of any of the M series of isoforms (Figure 6.09a). This mass corresponds to the amino acids 70-101 in the sequence of α_1–antitrypsin, with an asparagine residue at glycosylation site 83. This suggests that asparagine 83 was non-glycosylated in the diglycosylated isoforms of α_1–antitrypsin found in all CDG-I plasma. Masses corresponding to peptides containing non-glycosylated asparagines 46 (Fig 6.12c) and 247 (Fig. 6.11c) were not detected.
To confirm this, α_1–antitrypsin isoforms i to v were deglycosylated with N-glycanase prior to tryptic digestion. The resultant peptide maps contained four masses, 1756.9 m/z, 3181.6 m/z, 3198.6 and 3691.8 m/z, which were not observed in the analysis of the M series. This demonstrated that all of the isoforms were deglycosylated efficiently by N-glycanase treatment, prior to the trypsin digest, and that the unusual isoforms were not members of the usual M series of α_1–antitrypsin isoforms. The masses 3181.6 m/z and 3198.6 m/z (Fig 6.12b) correspond to the amino acid sequence 40-69 with oxidised methionine with and without pyroglutamate, respectively, but with aspartic acid at residue 46, indicating that this residue was glycosylated. Similarly the presence of mass 1756.9 (Fig. 5b) indicated that asparagine 247 was also glycosylated. The presence of mass 3691.8 m/z, corresponding to amino acid sequence 70-101 with an asparagine at position 83, confirmed that asparagine 83 was non-glycosylated in all of the diglycosylated α_1–antitrypsin isoforms in CDG-I. No apparent changes in the isotopic ratio of the peptides were observed in the mass spectra after the enzymic removal of the glycans using N-glycanase, which

converts an asparagine to aspartic acid with an increase in m/z of 1. This indicated very strongly that not even small amounts of asparagine 83 were glycosylated.

Figure 6.09

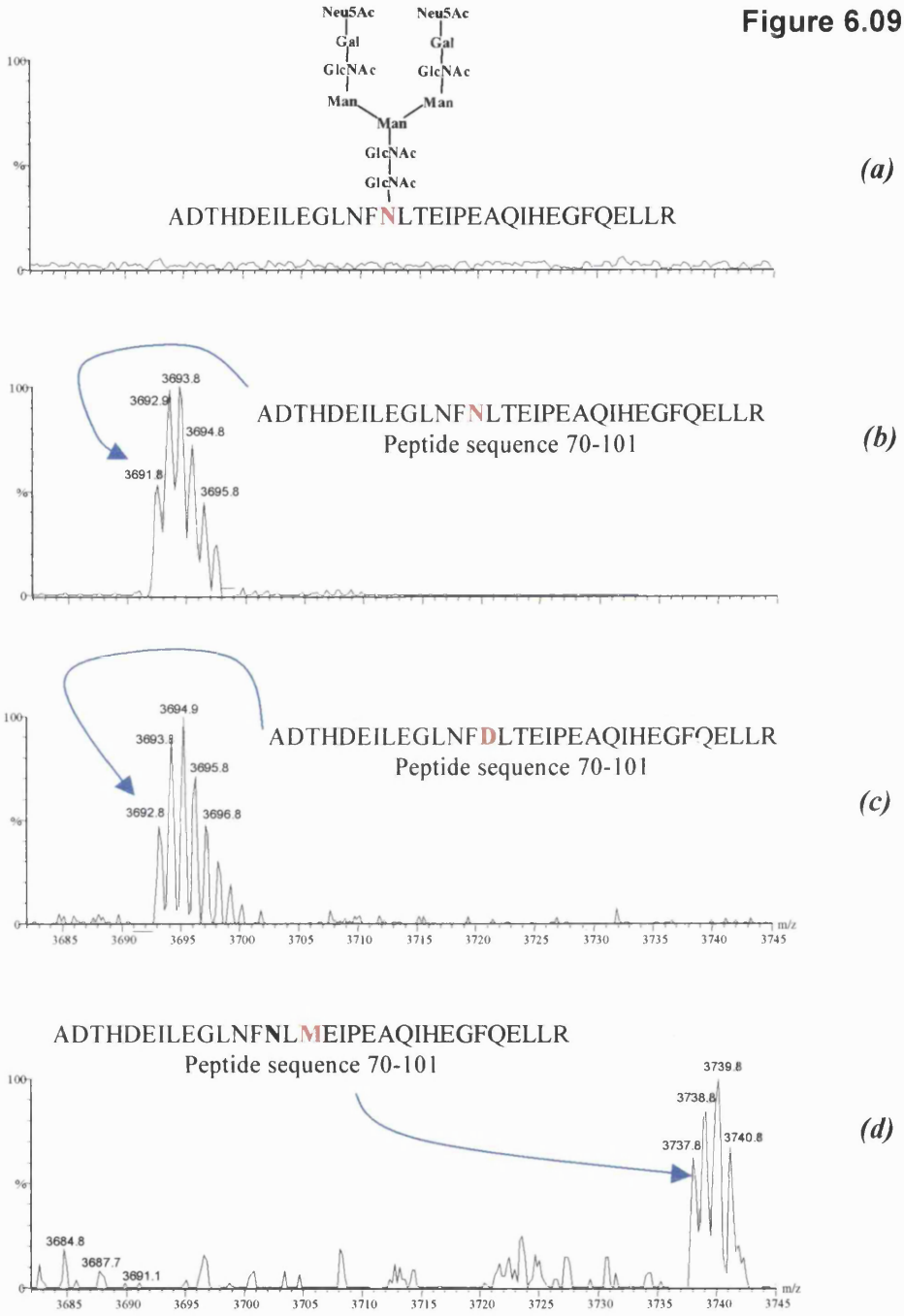


Figure 6.09 Mass spectral analysis of tryptic peptide containing Asn 83 of α_1-antitrypsin;

(a) In-gel digestion of (fully glycosylated) control α_1-antitrypsin.

(b) In-gel digestion of diglycosylated α_1-antitrypsin from CDG-I plasma, showing a peptide of mass 3691.8 m/z corresponding to amino acids 70-101.

(c) In-gel digestion of (fully glycosylated) α_1-antitrypsin after prior enzymatic removal of the glycans. A new peptide of mass 3692.8 m/z corresponding to amino acids 70-101 is observed.

(d) In-gel digestion diglycosylated α_1-antitrypsin isoform from a $PIZ_{Bristol}$ heterozygote patient, showing a peptide of mass 3737.8 m/z corresponding to amino acids 70-101.

Further evidence of the identity of the peptide covering amino acid residues 70-101 can be produced by mass mapping of the tryptic digest of the M series of α_1–antitrypsin after pre-treatment with N-glycanase. A peptide of mass 3692.8 m/z was detected in the mass spectrum after in-gel tryptic digestion of the deglycosylated isoform (Figure 6.09c). This mass corresponds to a peptide covering the amino acids 70-101 but with an aspartic acid residue at position 83. The mass of this peptide (3692.8 m/z) is one mass unit greater than the mass of the non-glycosylated peptide (3691.8 m/z), containing amino acids 70-101 found in α_1-antitrypsin isoforms (i)-(v) from CDG-Ia.

Confirmation of the identity of the peptide of mass 3691.8 m/z was gained from the analysis of the underglycosylated $PIZZ_{Bristol}$ α_1–antitrypsin isoforms as described in Chapter 5. An additional mass of 3737.8 m/z was detected in each of the isoforms (Figure 6.09d), which was not present in the normal M series of α_1–antitrypsin in both the PI $MZ_{Bristol}$ heterozygote and the normal control (Figure 6.09a). A peptide of 3737.8 m/z can be accounted for by a peptide containing the amino acids 70-101 with methionine at position 85 and asparagine at position 83.

6.3.3.3.2 Glycan analyses

Analysis by MALDI TOF MS of the glycans removed from the isoforms (i) to (v) was only possible for the more abundant diglycosylated species (ii) and (iii) (Figure 6.09). The spectrum for the glycans from isoform (ii) contained two ions of masses of 2223.4 m/z and 2880.1 m/z, corresponding to biantennary and triantennary complex glycans, respectively. The carbohydrate content of isoform (ii), combined with the peptide data showing both the presence of the N-terminal amino acids and the absence of a glycan structure at asparagine 83, indicates that this isoform is the diglycosylated equivalent of the M4 isoform of α_1–antitrypsin (Figure 6.09). This is consistent with the mass decrease of approximately 2.2 kDa and the cathodal shift in the pI of 0.1 units observed by 2D-PAGE.

The mass spectral analysis of the glycans released from isoform (iii) contained only one major ion species of mass 2223.4 m/z, corresponding to a biantennary complex glycan (Figure 6.10). The glycan and peptide mass spectral data indicate that isoform (iv) is the diglycosylated equivalent of the M6 isoform of α_1–antitrypsin, specifically lacking a biantennary glycan at asparagine 83. This is consistent with the pI and Mr co-ordinates observed for isoform (iii) on 2D-PAGE (Table 6.1).

Although it was not possible to analyse the carbohydrate composition of isoforms (iv) and (v), it is reasonable to assume from their co-ordinates on 2D-PAGE (Table 6.1) and peptide analysis that they were the diglycosylated equivalents of M7 and M8 isoforms of α_1–antitrypsin, respectively. They would arise in CDG by the N-terminal proteolytic processing of isoforms (ii) and (iii).

Isoform (i), which was only detected in CDG-Ic, has the same pI as isoform M6, which possesses three biantennary glycans. Therefore as isoform (i) retains the N-terminal sequence it probably has two triantennary complex glycans at asparagines 46 and 247.

Thus, the analysis of glycans was consistent with the data suggesting in all the diglycosylated isoforms of α_1–antitrypsin in plasma, from each form of CDG-I, asparagine 83 is preferentially not glycosylated.

Figure 6.10

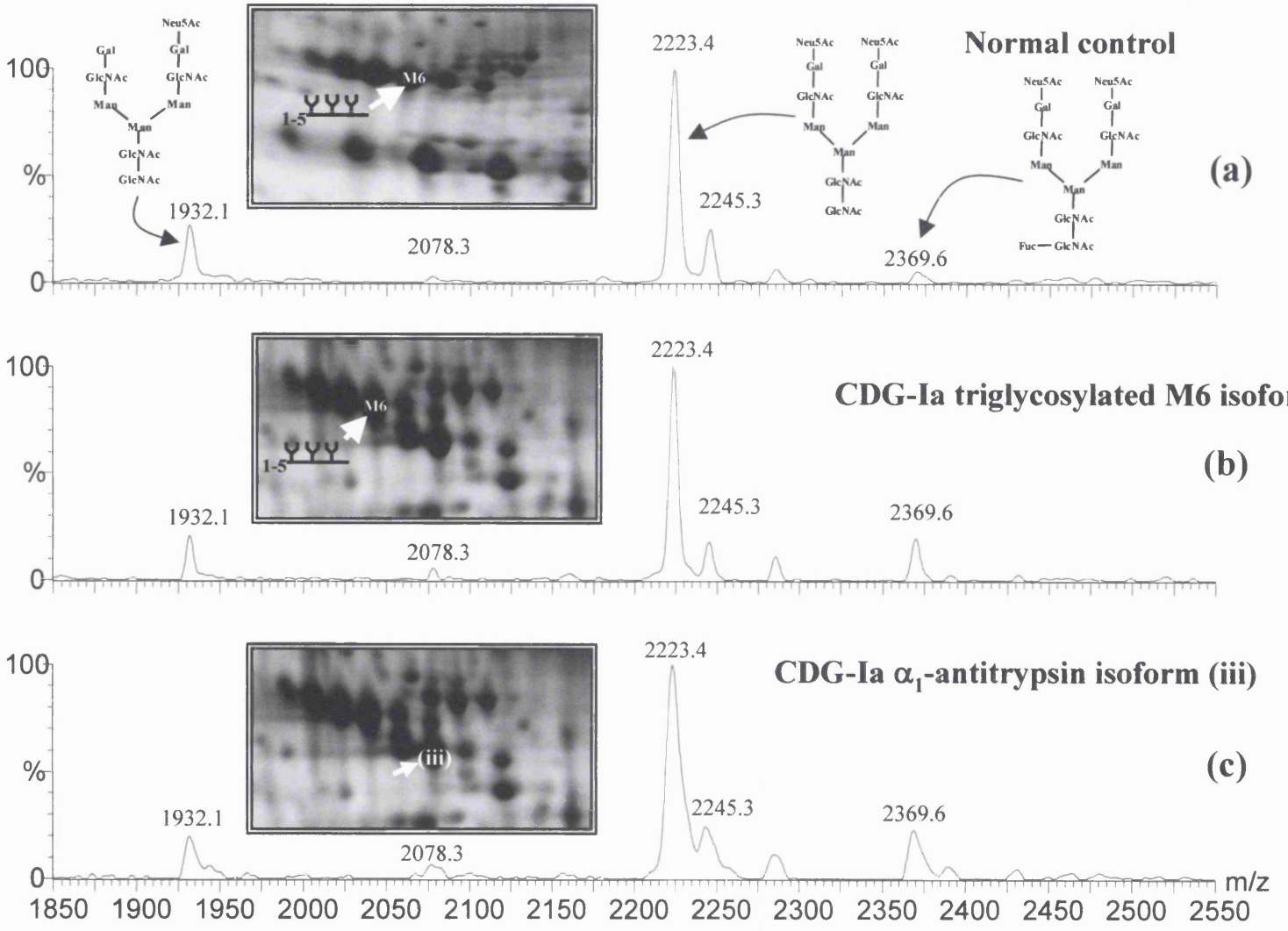


Figure 6.10 **Mass spectra of glycans from;**
(a) control M6 α_1-antitrypsin (b) CDG-I M6 α_1-antitrypsin (c) CDG-I diglycosylated α_1-antitrpysin isoform

6.3.3.4 Analysis of the monoglycosylated α_1–antitrypsin isoform (vi)

Mass spectral analysis of the peptides liberated from isoform (vi) confirmed this protein as α_1–antitrypsin. The detection of the reporter peptide for the N-terminus with a mass of 2819.2 m/z confirmed that the N-terminus was present. The presence of the mass of 3691.8 m/z, which is diagnostic for amino acids 70-101 with an asparagine at position 83, confirmed that this asparagine was also not glycosylated in this isoform (Figure 6.09c). A mass of 1755.9 m/z, which is due to amino acids 244-259 with an asparagine at position 247, was also detected (Figure 6.11c). This showed that the second unoccupied glycosylation site was asparagine 246. These results indicated that glycosylation sites 83 and 247 were not occupied in this isoform of α_1–antitrypsin.

This was confirmed by the mass spectral analysis of the tryptic digestion of deglycosylated isoform (vi) (Figure 6.12b). The appearance of two ions of masses 3181.6 m/z and 3198.6 m/z, corresponding to the amino acid sequence, 40-69, and its oxidised methionine derivative, with both containing an aspartic acid, indicated that asparagine 46 was glycosylated. The diagnostic reporter masses for non-glycosylated asparagines, 83 and 247, were also detected. Therefore, isoform (vi) is monoglycosylated at asparagine 46 and retains the first 5 amino acids.

It was not possible to analyse the glycans attached to asparagine 46 in isoform (vi) by in-gel digestion but theoretically a bi or a tri-antennary glycan could be present. The estimated pI from its position on the 2D gel, 5.26, is consistent with the presence of two sialic acids i.e. a bi-antennary glycan. The fully defined structure, M6, which has three bi-antennary glycans and the N-terminal sequence, has a pI of 5.05. The pI of isoform (vi) is consistent with the loss of two biantennary glycans, which would increase the pI by approximately 0.2 pH units. Therefore, isoform (vi) is probably the monoglycosylated equivalent of M6 with only one biantennary glycan at asparagine 46.

Figure 6.11

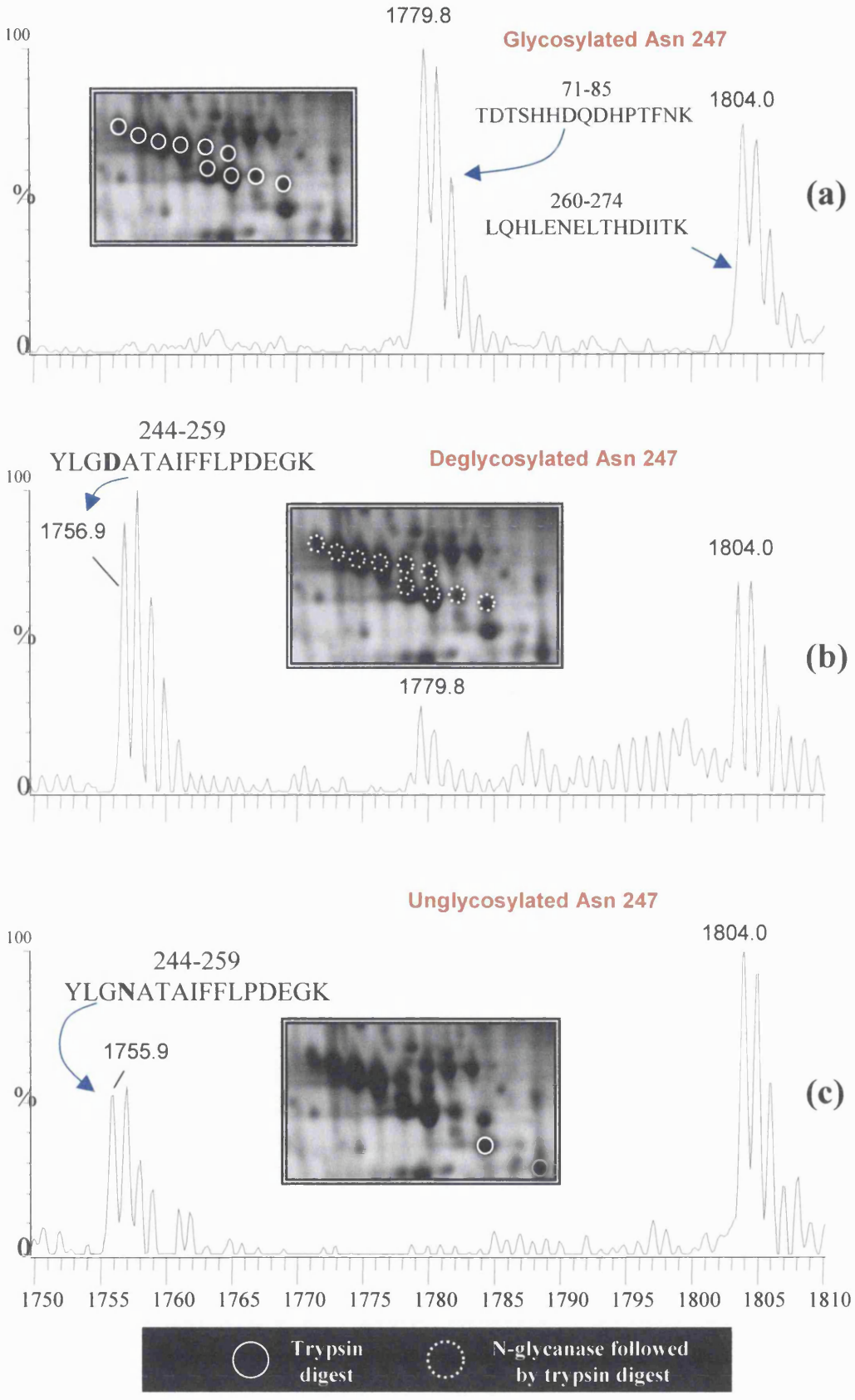


Figure 6.11 Mass spectral analysis of tryptic peptide containing Asn 247 of α_1-antitrypsin from;

(a) In-gel digestion of control α_1-antitrypsin.

(b) In-gel digestion of deglycosylated α_1-antitrypsin from CDG-I showing a peptide of mass 1756.9 m/z corresponding to amino acids 244-259.

(c) In-gel digestion of mono- and unglycosylated α_1-antitrypsin from CDG-I showing a peptide of mass 1755.9 m/z corresponding to amino acids 244-259.

6.3.3.5 Analysis of the non-glycosylated α_1–antitrypsin isoform (vii)

Mass spectral analysis of the peptides liberated from isoform (vii) identified this protein as an isoform of α_1–antitrypsin with an intact N-terminus. The presence of the peptide masses of 3180.6 m/z and 3197.6 m/z (Figure 6.12c), 3691.8 m/z (Figure 6.09c) and 1755.9 m/z (Figure 6.11c) indicated that asparagines 46, 83 and 247, were all non-glycosylated in isoform (vii) of α_1–antitrypsin.

Treatment of isoform (vii) with N-glycanase, prior to tryptic digestion, did not produce any additional masses and confirmed that this was the non-glycosylated isoform of α_1–antitrypsin. Theoretically, on the 2D-PAGE system, the non-glycosylated isoform of α_1–antitrypsin with an intact N-terminus should have a pI of approximately 5.35 units, which is in good agreement with the estimated observed pI of 5.36. The proposed macro- and micro-heterogeneity of the underglycosylated isoforms of α_1–antitrypsin in patients are shown in Figure 6.13.

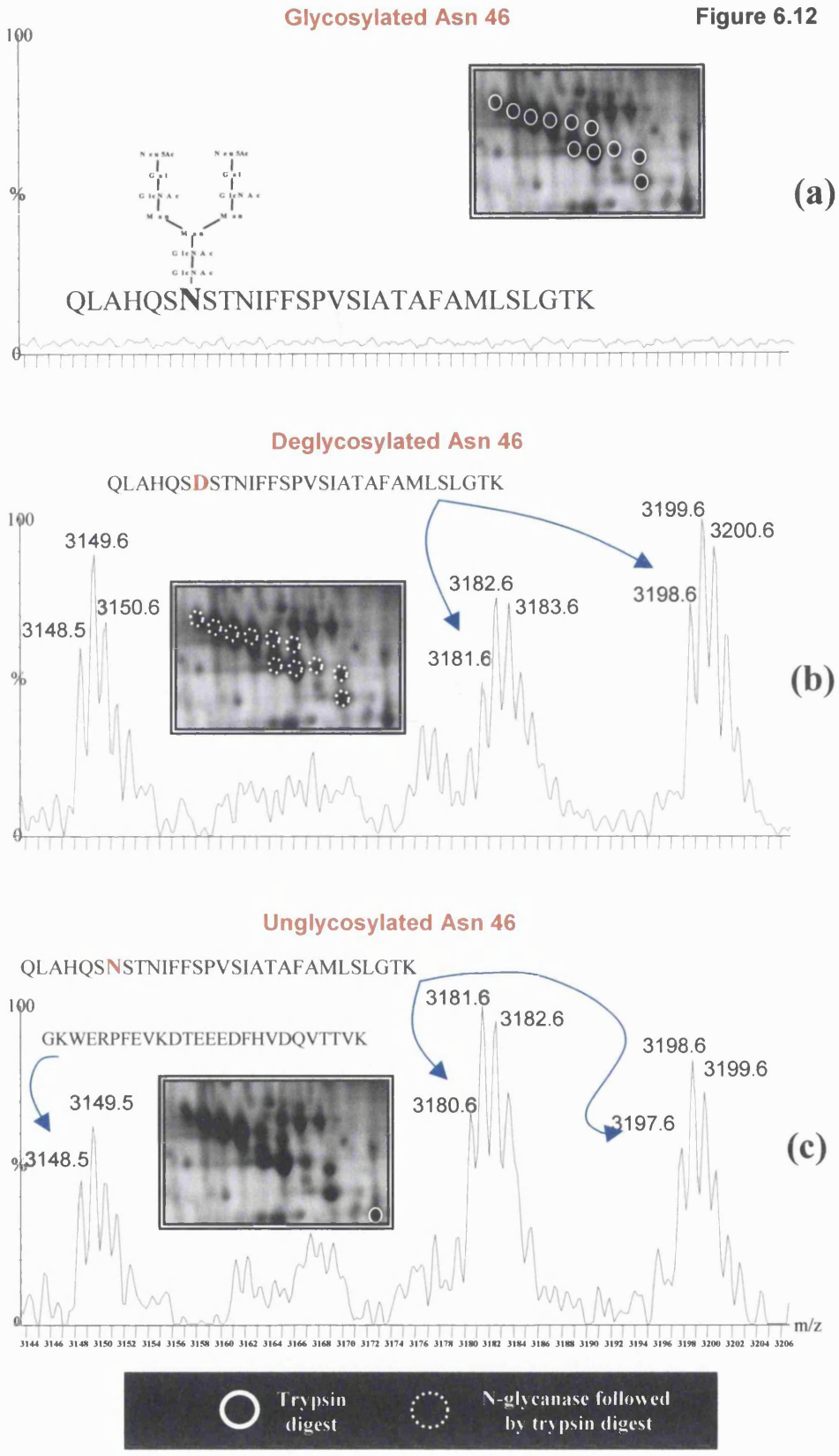


Figure 6.12 Mass spectral analysis of tryptic peptide containing Asn 46 of α_1-antitrypsin after;

(a) In-gel digestion of control α_1-antitrypsin.

(b) In-gel digestion of α_1-antitrypsin after prior enzymatic removal of the glycans. A new peptide of of masses 3181.6 and 3198.6 m/z corresponding to amino acids 40-69 is observed.

(c) In-gel digestion of the unglycosylated α_1-antitrypsin isoform showing peptides of masses 3180.6 and 3197.6 m/z corresponding to amino acids 40-69 with unglycosylated asparagine 46

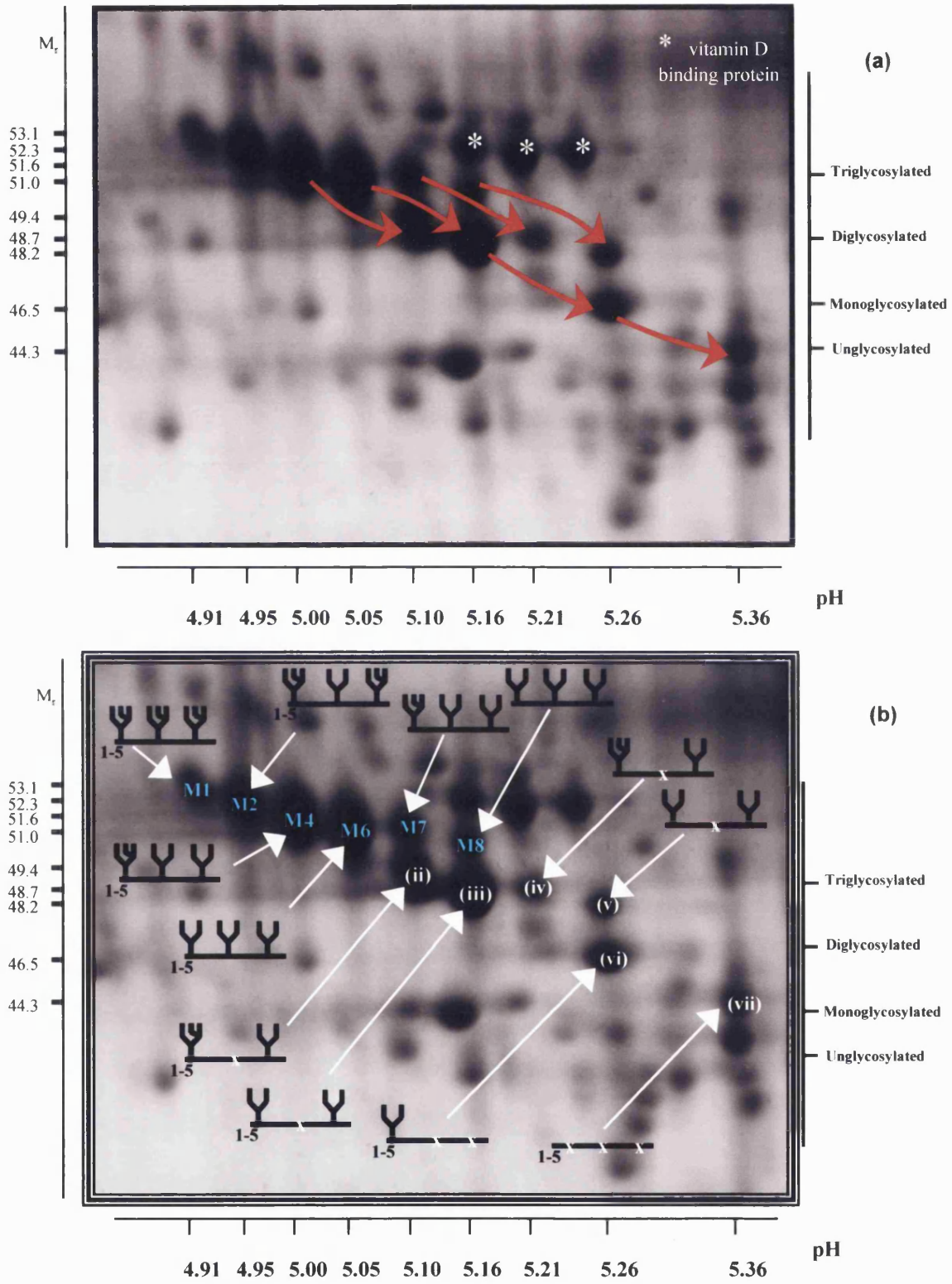


Figure 6.13 **Summary of the proposed micro- and macroheterogeneity of CDG-I plasma α_1-antitrypsin isoforms separated by 2D-PAGE.**

(a) 2D-PAGE showing the proposed movement of α_1-antitrypsin isoforms observed due to underglycosylation.

(b) Proposed macro- and microheterogeneity of α_1-antitrypsin isoforms observed in CDG-I plasma.

6.3.4 Conclusions.

Analysis of the site-specific glycosylation of each underglycosylated isoform of α_1–antitrypsin in the plasma samples of CDG-I patients, clearly demonstrated that the pattern of N-glycosylation site occupancy was not random. In the monoglycosylated isoform (vi), the first glycosylation site in the amino acid sequence, asparagine 46, was occupied in preference to the glycosylation sites at asparagines 83 and 247. Analysis of the diglycosylated isoforms showed that asparagines 46 and 247 were always glycosylated. Asparagine 83 was never glycosylated in any of the underglycosylated forms of α_1–antitrypsin. This pattern of preferential glycosylation of sites was identical in each of the three different CDG-I patients analysed. In all cases the fully glycosylated isoform was always the most abundant isoform present in plasma, with lesser amounts of the di-, mono- and non-glycosylated isoforms, in this order.

The pattern of occupancy of N-glycosylation sites in the α_1–antitrypsin isoforms present in plasma would depend on several factors. Firstly, the initial degree of glycosylation of the protein will depend upon the glycosylation capacity of the hepatocytes and the integrity of the nascent polypeptide chain. As the structural gene for α_1-antitrypsin is normal in CDG-I, the appearance of underglycosylated isoforms in the plasma must be due to the decreased capacity for glycosylation. Secondly, the flux of fully and partially glycosylated α_1–antitrypsin through the lumen of the rough endoplasmic reticulum to the sites of N-linked glycan processing will be determined by the editing /quality control mechanism, mediated by calnexin and other chaperones (Molinari and Helenius, 2000; Choudhury *et al.*, 1997). It has been shown that the correct folding of the nascent α_1–antitrypsin polypeptide in the ER during its synthesis involves a cyclic association with and release from calnexin until it is correctly folded (Choudhury *et al.*, 1997). This process is regulated by glucosylation and de-glucosylation of the N-linked oligosaccharide, which is mediated by the enzyme, UDP-glucose: glycoprotein glucosyltransferase. Incorrectly folded isoforms of α_1–antitrypsin are targeted to the proteasome for degradation by ubiquitination

of the calnexin in the calnexin-misfolded α_1–antitrypsin complex (Qu *et al.*, 1996). The overall decrease in the plasma concentration of α_1–antitrypsin in CDG-I suggests that a significant proportion of the α_1-antitrypsin is subjected to proteasomal degradation. The fully glycosylated isoforms constitute about 85% of the total plasma α_1–antitrypsin in CDG-I (equivalent to approximately 40% of the normal concentration) suggesting that underglycosylated isoforms are not exported efficiently. As non-glycosylated recombinant α_1–antitrypsin has been shown to be active *in vitro* (Rosenberg *et al.*, 1984; Travis *et al.*, 1985), it is probable that the small amount of non-glycosylated α_1–antitrypsin present in plasma from CDG-I patients is also functional. This suggests that either α_1–antitrypsin can fold up correctly independently of glycosylation by an alternative chaperone-mediated pathway (Molinari and Helenius, 2000) or some α_1–antitrypsin molecules bypass the calnexin/calreticulin editing mechanism (Cooper *et al.*, 1997).

The second reason for the heterogeneity is the processing of the N-linked glycans, which takes place predominantly in the Golgi apparatus. The glycans on the plasma isoforms of α_1–antitrypsin are fully processed in CDG-I, as shown in (Figure 6.10) (Mills *et al.*, 2001; Mills *et al.*, 2001a). In fact there is increased fucosylation and branching of the glycans on both the fully and partially glycosylated isoforms in CDG-I (Mills *et al.*, 2001), especially in CDG-Ic. This probably explains why isoform (i), which has two triantennary complex glycans, is only detected in CDG-Ic (Fig.1). The post-translational removal of the five N-terminal amino acids also takes place in the Golgi. This modification appears to be unaffected by underglycosylation, as judged by the appearance of isoforms (iv) and (v) in the plasma of CDG-I patients.

The final factor affecting the pattern of isoforms of α_1–antitrypsin in plasma is their clearance from the circulation. The absence or decrease in carbohydrate might be expected to lower the stability of an isoform but it may also diminish its removal from the circulation by carbohydrate-mediated endocytosis. Non-glycosylated recombinant human α_1–antitrypsin has been demonstrated previously as having a shorter half-life in rabbits than the normal glycosylated form (Travis *et al.*, 1985).

The pattern of N-glycosylation site occupancy observed in the mature forms of α_1–antitrypsin in plasma could represent the pattern of glycosylation after synthesis or the spectrum of glycoforms allowed through the editing/quality control mechanism. If it were the latter, glycosylation of asparagine 46 would be a minimal requirement for clearance of α_1–antitrypsin by the editing /quality control mechanism, except for the small amount of non-glycosylated α_1–antitrypsin. Underglycosylated α_1-antitrypsin is present in the plasma of patients with the (PIZ $_{Bristol}$) variant, in which genetic abrogation of the asparagine 83 glycosylation sequon occurs (Mills *et al.*, 2001a; Lovegrove *et al.*, 1997), confirming that glycosylation of asparagine 83 is not essential for secretion. Diglycosylated α_1–antitrypsin with glycans attached at asparagines 83 and 247 or 46 and 83, or monoglycosylated α_1–antitrypsin with a single glycan at asparagine 83 or 247 were not detected in plasma. These forms were therefore either removed by the editing/quality control mechanism or not synthesised in the first place.

If all isoforms were synthesised, then the pattern of glycosylation is very specific, with asparagine 46 being glycosylated most efficiently, followed by asparagine 247 and finally asparagine 83. The preferential occupation of the glycosylation site closest to the N-terminal is consistent with other data on the positional importance of glycosylation in the folding and maturation of glycoproteins (Gavel and von Heijne, 1990; Samandari and Brown, 1993). It is therefore interesting to speculate on the molecular basis for this pattern of glycosylation. Sequons containing threonine rather than serine have been reported to be glycosylated more efficiently *in vitro* (Shakin-Eshleman *et al.*, 1996; Mellquist *et al.*, 1998). However, all three glycosylation sites in α_1–antitrypsin have threonine in the third position of the glycosylation sequon. The middle amino acid of the Asn.X.Ser/Thr sequence has also been shown to affect the efficiency of glycosylation *in vitro* (Shakin-Eshleman *et al.*, 1996) and Asn 46 is followed by a serine, which was the most efficient residue in the viral glycoprotein experiments. Asn 247 precedes an alanine, which is also an efficient amino acid, whereas Asn 83 is followed by leucine, which was one of the least efficient amino acids. The amino acid following the sequon is also an important determinant of glycosylation efficiency in the viral glycoprotein

expression system, especially if the third amino acid is serine (Mellquist *et al.*, 1998). However the amino acids following the three sequons in α_1–antitrypsin, asparagine, glutamic acid and alanine, respectively, have comparable effects on the glycosylation of threonine-containing sequons. Therefore it is possible that asparagine 83 is glycosylated less efficiently under conditions of decreased glycosylation capacity because of the chemical nature of the X residue, leucine, in the sequon. The efficiency of glycosylation of a sequon in DNAse 1 has been shown to depend on the X amino acid and the tissue of origin (Nishikawa and Mizuno, 2001). It is probable that the major factor determining the pattern of underglycosylation of α_1–antitrypsin under conditions of inadequate glycosylation capacity, as in CDG-I, is the efficiency of glycosylation of sequons. Subsequently any underglycosylated α_1–antitrypsin isoforms that cannot fold properly will be removed by the quality control mechanism in the ER. The half-lives of the isoforms secreted into the blood circulation will also be affected by their state of glycosylation. Therefore the pattern of α_1–antitrypsin isoforms found in plasma will be a reflection of all these processes. It is probable from previously published data that the sequon closest to the N-terminal of a protein will generally be glycosylated. However, the occupancy of other sequons will depend upon the efficiency of glycosylation of each sequon in that particular protein. Therefore the pattern of occupancy of N-glycosylation sites of plasma glycoproteins in CDG-I will vary from protein to protein. The same factors probably determine the occupancy of glycosylation sites under conditions in which glycosylation capacity is not restricted, together with the physiological state and type of cell in which the protein is synthesised.

Chapter 7

General discussion and remarks

Chapter 7 - General discussion and remarks

7.0 General discussion and remarks

In the mid 20th century, the discovery of the double-helix by Watson and Crick (1953) created the field of molecular biology. The subsequent focus on DNA led to extensive research on the genetic basis of human disease and the origin and evolution of life. This pioneering work has culminated in the successful completion of the sequencing of the human genome as well as the complete genomic analysis of many other species (Lander *et al.*, 2001). A number of techniques have emerged that provide a powerful set of tools to study global and quantitative genome expression. These include differential display PCR (Liang and Pardee, 1992), cDNA microarrays (Lashkari *et al.*, 1997; DeRisi *et al.*, 1996) and serial analysis of gene expression (Velculescu *et al.*, 1995). The hypothesis of one gene encoding one protein has had to be reevaluated, with organisms being far more complex than is indicated by analysis of genetic material alone. Furthermore, the lack of correlation between transcriptional profiles and actual cellular protein levels has highlighted the need for information on proteins (Anderson and Seilhamer, 1997; Anderson and Anderson, 1998). Banks and colleagues (2000) have suggested that the reverse approach, examination of the functional output or the expressed protein complement of a genome, the 'proteome', can provide additional complementary information.

The idea of mapping the human proteome was first put forward almost 20 years ago (Anderson and Anderson, 1982) but it was not possible until 1993 when several laboratories published techniques for direct analysis of proteins separated by PAGE by mass spectrometry (Pappin *et al.*, 1993; Henzel *et al.*, 1993; Mann *et al.*, 1993). Several features of 2D-PAGE have made it difficult to replace in proteomic analysis and currently no other technique provides as much accurate and detailed information on a particular protein, including its relative quantity, co- and post-translational modifications, pI, and molecular weight (Yates *et al.*, 1993).

The major part of the work described in this thesis involved establishing a robust and reproducible 2D-PAGE method, optimisation of the in-solution and in-gel proteolytic and deglycosylation strategies and identification of proteins using MALDI TOF MS. The initial aims of the project were to analyse accurately (a) disease-causing and other non-disease causing amino acid mutations and, (b) N-glycosylation site-occupancy in plasma glycoproteins with multiple glycosylation sites.

7.1 Choice of biological material and the reasons for its selection

The α_1-antitrypsin deficiencies were chosen because they challenged the method in several ways. The glycoprotein, α_1-antitrypsin, exists normally as six major isoforms (M series), which differ in size and charge because of differences in the amino acid sequence and the branching and sialic acid content of their N-linked glycans. Therefore it was thus necessary to identify and define each of the M series of the α_1-antitrypsin isoforms in normal controls, using the techniques developed, before analysing the isoforms from patients with variant α_1-antitrypsin. Three different variants of α_1-antitrypsin were selected for analysis: a charged amino acid substitution (E342K), a neutral amino acid substitution (V213A) and the PI $Z_{Bristol}$ mutation, which has a neutral amino acid substitution (T85M) that abolishes the second glycosylation site in the molecule (asparagine 83, NLT→NLM).

The CDG-I disorder was chosen for the investigation of glycosylation site-occupancy in plasma glycoproteins because in this disorder there is a decreased capacity for glycosylation and results in the underglycosylation of proteins. Analysis of site occupancy under these conditions might reveal the factors that determine efficiency of glycosylation site occupancy. Three different sub-types of CDG type I syndrome were studied i.e. Ia, Ic and Ix. All of these produce a reduced supply of mature LLO and result in the hypoglycosylation of glycoproteins.

7.2 Accurate detection of amino acid substitutions and polymorphic alterations in α_1-antitrypsin.

Using 2D-PAGE, proteolysis, deglycosylation and mass spectral methods developed in the course of this investigation, it was possible to confirm the identity of each α_1-antitrypsin isoform and mutations and polymorphisms in the α_1-antitrypsin allele. The use of MALDI TOF MS to analyse the polypeptide backbone and the glycans attached to each isoform allowed the unequivocal assignment of each α_1-antitrypsin isoform to the M series and was used as the basis for the diagnosis of α_1-antitrypsin deficiencies and polymorphisms.

The neutral amino acid substitution V213A, although not resulting in any detectable changes in pI on the 2D-PAGE, could be identified using MALDI TOF MS to detect a tryptic peptide of a different mass. In contrast, the charged amino acid mutation E342K (PI Z) did result in changes in pI that were detected by 2D-PAGE and confirmed by subsequent MALDI TOF MS. Similarly, the PI $Z_{Bristol}$ mutation was detected by 2D-PAGE because of the changes in both the pI and molecular weight of the underglycosylated α_1-antitrypsin isoforms, which were confirmed by MALDI TOF MS.

These findings show clearly that structural variants of α_1-antitrypsin can be analysed by the proteomic techniques developed during the course of this work and that their application to the analysis of samples from patients with α_1-antitrypsin deficiencies is feasible. To summarise, the use of the proteomics can be used as a powerful tool in the identification and confirmation of mutations, polymorphisms and changes in the post-translational modification of plasma α_1-antitrypsin.

7.3 Detection of the site-specific glycosylation in hypoglycosylated proteins purified from CDG-I plasma.

Using the proteomic techniques developed in the course of this work it was possible to identify several glycoproteins in the plasma of patients with CDG-I that were

underglycosylated. Two of these proteins, transferrin and α_1-antitrypsin, were purified further and analysed to determine the site-specific glycosylation of each isoform.

The use of high-resolution 2D-PAGE and MALDI TOF MS allowed the identification of the isoforms of α_1-antitrypsin that contained fully (tri-), bi-, mono- and non-glycosylated isoforms. The state of glycosylation of each of the three glycosylation sites in each isoform of α_1-antitrypsin, was established by monitoring whether a 1 Da increase in the mass of the peptide containing each sequon was observed following treatment with N-glycanase. The removal of a glycan by N-glycanase results in the conversion of an asparagine to an aspartic acid residue, with an increase in mass of 1 Da which can readily be detected using MALDI TOF MS. Using this method it was not only possible to confirm which sites where occupied in each α_1-antitrypsin isoform but using the pI / Mr data from the 2D-PAGE, it was also possible to deduce the identity of the glycans present on each isoform. The results indicated that for the correct maturation of the protein, the first glycosylation site in α_1-antitrypsin (asparagine 46) is the most important, followed by the glycosylation site at asparagine 247, and the least important sequon being asparagine 83.

Analysis of plasma transferrin using the same techniques suggested that the glycosylation site nearest the N-terminal (Asn 413) was occupied preferentially to the second glycosylation site (Asn 611). Whilst the techniques used were successful for monitoring all three glycosylation sites in α_1-antitrypsin, it was not possible to confirm unequivocally that in transferrin the asparagine 413 was preferentially occupied over asparagine 611. However, other data implied strongly that the sequon nearer the N-terminus was also the most favorable for the correct maturation of the transferrin molecule.

7.4 Limitations of the capability of the techniques.

The role of mass spectrometry in proteomic analyses is often restricted to that of the general identification of proteins. However, this study attempted to use MALDI

TOF MS to identify or focus on specific areas of the protein amino acid sequence that contained known mutations of polymorphisms. It becomes apparent that the detection of peptides covering less than 20% of the total amino acid sequence although sufficient to identify the protein by mass mapping, was inadequate for the type of analyses proposed in this work and thus new strategies would have to be introduced to overcome these problems.

The reasons for the low amino acid sequence coverage are unclear but are probably multifactorial and include;

(i) Some of the resulting fragments produced after in-gel proteolytic digestion were too small or large to be analysed by MALDI TOF MS operating in reflectron mode i.e. smaller peptides fall into the matrix region of the mass spectra (< 650 m/z) or are too big to be analysed in the reflectron (> 6000 m/z).

(ii) Some of the peptides are lost during the extraction or C-18 clean-up.

(iii) Non-specific proteolytic cleavage of proteins.

(iv) The ionisation of the digestion products resulted in particular peptides or glycopeptides, being suppressed or they were not amenable to MALDI TOF MS analysis.

The latter reason is probably the largest factor for the low amino acid sequence coverage. However, successful attempts to increase the amino acid sequence coverage were made, using several proteases to digest each protein and was shown to increase the sequence coverage in some cases to 100% (Chapter 4).

Another problem encountered with MALDI TOF MS analysis of peptide mixtures was that of quantitation. Although MALDI TOF MS has been shown to be quantitative when used with appropriate internal standards (Section 3.7.2.4, Gusev *et al.*, 1996), caution is required in using peak heights to quantify the relative proportion of the site-occupancy between two glycosylation sites.

Perhaps the major limitation to this study was the lack of access to a mass spectrometer with good sequencing capabilities. Although the MALDI TOF mass spectrometer used in this study has the ability to sequence peptides by PSD

(TofSpec E, Micromass, UK), it was found that larger amounts of material were necessary to obtain sequence information than could be excised from polyacrylamide gels. However, even when larger amounts of material were obtained, it was often difficult to interpret the mass spectra obtained or no significant post-source decay products were observed to provide information on the amino acid sequence.

7.5 Future work

A mass spectrometer with a capacity for sequencing should be used to confirm the identity of those peptides and glycopeptides containing amino acid substitutions and missing glycan structures. It is also conceivable that the softer ionisation method of electrospray may have facilitated the analysis of intact N-glycopeptides, a phenomenon that could not be achieved using the *uv*-MALDI TOF MS. Although the macroheterogeneity of N-glycoproteins was demonstrated in this work, the ability to analyse intact N-glycopeptides would also have allowed the determination of the microheterogeneity of each of the glycosylation sites. The use of a mass spectrometer fitted with an electrospray ionisation source would also have allowed the biochemical techniques developed in this study to be extended to the analysis of other post-translational modifications such as phosphorylation, sulphation, O-linked glycosylation or acetylation and therefore applied to the study of other inborn errors of metabolism.

Although the techniques developed in this work demonstrated the potential of proteomics, the techniques are essentially destructive i.e. the proteins are degraded prior to analysis. Therefore, this does not allow the protein to be analysed in its native conformation or as a complex with other proteins, often an important factor for the correct functioning of a protein or enzyme. For this reason, future research in proteomics will require a preliminary separation that isolates proteins in their native state permitting the study of protein-protein interactions.

References

Abbas MA and Latham J (1967) The Instability of Evaporating Charged Drops. *J Fluid Mechanics* **30**: 663-670.

Aebi M and Hennet T (2001) Congenital Disorders of Glycosylation: Genetic Model Systems Lead the Way. *Trends in Cell Biology* **11**: 136-141.

Amado M, Almeida R, Schwientek T and Clausen H (1999) Identification and Characterization of Large Galactosyltransferase Gene Families: Galactosyltransferases For All Functions. *Biochim Biophys Acta* **1473**: 35-53.

Anderson L and Seilhamer J (1997) A Comparison of Selected mRNA and Protein Abundances in Human Liver. *Electrophoresis* **18**: 533-537.

Anderson NG and Anderson L (1982) The Human Protein Index. *Clin Chem* **28**: 739-748.

Anderson NL and Anderson NG (1998) Proteome and Proteomics: New Technologies, New Concepts, and New Words. *Electrophoresis* **19**: 1853-1861.

Apweiler R, Hermjakob H and Sharon N (1999) On the Frequency of Protein Glycosylation, as Deduced From Analysis of the SWISS-PROT Database. *Biochim Biophys Acta* **1473**: 4-8.

Austen BM and Westwood OMR (1991) Protein Targeting and Secretion (edited by Rickwood D). IRL Press, Oxford.

Bagshaw RD, Callahan JW and Mahuran DJ (2000) Desalting of In-Gel-Digested Protein Sample With Mini-C18 Columns for Matrix-Assisted Laser Desorption Ionization Time of Flight Peptide Mass Fingerprinting. *Anal Biochem* **284**: 432-43

Baker EN and Lindley PF (1992) New Perspectives on the Structure and Function of Transferrins. *J Inorg Biochem* **47**: 147-160.

Banks RE, Dunn MJ, Hochstrasser DF, Sanchez J-C, Blackstock W, Pappin DJ and Selby PJ (2000) Proteomics: New Perspectives, New Biomedical Opportunities. *Lancet* **356**: 1749-1756.

Barber M, Bordoli RS, Sedgwick RD and Taylor AN (1981) Fast Atom Bombardment of Solids as an Ion-Source in Mass-Spectrometry. *Nature* **293**: 270-275.

Barton N W, B rady R O, D ambrosia J M, D i B isceglie A M, D oppelt S H, H ill S C, Mankin HJ, Murray GJ, Parker RI, Argoff CE, Grewal RP and Yu K-T (1991) Replacement Therapy for Inherited Enzyme Deficiency - Macrophage-Targeted Glucocerebrosidase for Gaucher's Disease. *N Engl J Med* **324**: 1464 - 1470.

Barton NW, Furbish FS, Murray GJ, Garfield M and Brady RO (1990) Therapeutic Response to Intravenous Infusions of Glucocerebrosidase in a Patient with Gaucher Disease. *Proc Natl Acad Sci USA* **87**: 1913-1916.

Bathurst IC, Travis J, George PM and Carrell RW (1984) Structural and Functional Characterization of the Abnormal Z α_1-Antitrypsin Isolated From Human Liver. *FEBS Lett* **177**: 179-183.

Bause E (1983) Structural Requirements of N-Glycosylation of Proteins. Studies With Proline Peptides as Conformational Probes. *Biochem J* **209**: 331-336.

Beavis RC and Chait BT (1990) High-Accuracy Molecular Mass Determination of Proteins Using Matrix-Assisted Laser Desorption Mass Spectrometry. *Anal Chem* **62**: 1836-1840.

Beuhler RJ, Flanigan E, Greene LJ and Friedman L (1974) Proton Transfer Mass Spectrometry of Peptides. A Rapid Heating Technique for Underivatized Peptides Containing Arginine. *J Am Chem Soc* **96**: 3990-3999.

Beutler E and Grabowski GA (2001) Gaucher Disease. In: The Metabolic and Molecular Bases of Inherited Diseases III 8th Edition, 2001, p. 3635-3668, Scriver CR, Beaudet AL, Sly WS and Valle D (eds.) McGraw-Hill, New York.

Beutler E, Gelbart T, Lee P, Trevino R, Fernandez MA and Fairbanks VF (2000) Molecular characterization of a Case of Atransferrinemia. *Blood* **96**: 4071-4074.

Billeci T.M and Stults J.T (1993) Tryptic Mapping of Recombinant Proteins by Matrix-Assisted Laser-Desorption Ionisation Mass Spectrometry. *Anal Chem* **65**: 1709-1716.

Bironaite D, Lindgren S and Janciauskiene S (2001) Fibrillogenic C-Terminal Fragment of Alpha-1-Antitrypsin Activates Human Monocytes Via Oxidative Mechanisms. *Cell Tissue Res* **305**: 87-98.

Bjellqvist B, Ek K, Righetti PG, Gianazza E, Gorg A, Wesrermeier R, Postel W (1982) Isoelectric-Focusing in Immobilized pH Gradients - Principle, Methodology and Some Applications. *J Biochem Biophys Meth* **6**: 317-339.

Brady RO, Murray GJ and Barton NW (1994) Modifying Exogenous Glucocerebrosidase for Effective Replacement Therapy in Gaucher Disease. *J Inherit Metab Dis* **17**: 510-519.

Brantly M, Nukiwa T and Crystal RG (1988) Molecular Basis of Alpha-1-Antitrypsin Deficiency. *Am J Med* **84**: 13-31.

Brantly M L, W ittes J T, Vogelmeier C F, H ubbard R C, F ells G A a nd C rystal R G (1991) Use of a Highly Purified α_1-Antitrypsin Standard to Establish ranges for the Common Normal and Deficient α_1-Antitrypsin Phenotypes. *Chest* **100**:703-708.

Brockhausen I, Carver J and Schacter H (1988) Control of Glycoprotein Synthesis. XIV. The Use of Oligosaccharide Substrates and HPLC to Study the Sequential Pathway for N-Acetylglucosaminyltransferases I, II, III, IV, V and VI in the Biosynthesis of Highly Branched N-Glycans by Hen Oviduct Membranes. *Biochem Cell Biol* **66**: 1134-1151.

Brown RS and Gilfrich NL (1991) Design and Performance of a Matrix-Assisted Laser Desorption Mass Spectrometer Utilising a Pulsed Nitrogen Laser. *Analytica Chimica Acta* **248**: 541-552.

Carrell RW and Lomas DA (1997) Conformational Disease. *Lancet* **350**:134-138.

Carrell RW and Lomas DA (2002) Alpha$_1$-Antitrypsin Deficiency - a Model for Conformational Diseases. *N Engl J Med* **346**: 45-52.

Chai W, Piskarev V and Lawson AM (2001) Negative-Ion Electrospray Mass Spectrometry of Neutral Underivatised Oligosaccharides. *Anal Chem* **73**: 651:657.

Chantret I, Dupre T, Delenda C, Bucher S, Dancourt J, Barnier A, Charollais A, Heron D, Bader-Meunier B, Danos O, Seta N, Durand G, Oriol R, Codogno P and Moore SE (2002) Congenital Disorders of Glycosylation Type Ig is Defined b y a D eficiency i n D olichyl-P-Mannose : Man(7)GlcNAc(2)-PP-dolichyl Mannosyltransferase. *J Biol Chem* **277**: 25815-25822.

Charlwood J, Clayton P, Keir G, Mian N and Winchester B (1998) Defective Galactosylation of Serum Transferrin in Galactosemia. *Glycobiology* **8**: 351-357.

Chernushevich IV, Loboda AV and Thomson BA (2001) An Introduction to Quadrupole-Time of Flight Mass Spectrometry *J Mass Spectrom* **36**: 849-865.

Choudhury P, Liu Y, Bick RJ and Sifers RN (1997) Intracellular Association Between UDP-Glucose:Glycoprotein Glucosyltransferase and an Incompletely Folded Variant of Alpha1-Antitrypsin. *J Biol Chem* **272**: 13446-13451.

Choy FYM, Woo M and Potier M (1986) In Situ Radiation-Inactivation Size of Fibroblast Membrane-Bound Acid Beta-Glucosidase in Gaucher Type I, Type 2 and Type 3 Disease. *Biochim Biophys Acta* **870**: 76-81.

Chui D, Oh-Eda M, Liao Y-F, Panneerselvam K, Lal A, Marek KW, Freeze HH, Moremen KW, Fukuda MN and Marth JD (1997) Alpha-Mannosidase-II Deficiency Results in Dyserythropoiesis and Unveils an Alternate Pathway in Oligosaccharide Biosynthesis. *Cell* **90**: 157-167.

Clauser KR, Baker PR and Burlingame AL (1999) Role of Accurate Mass Measurement (+/- 10ppm) in Protein Identification Strategies Employing MS or MS/MS and Database Searching. *Anal Chem* **71**: 2871-2882.

Clayton PT (2001) Applications of Mass Spectrometry in the Study of Inborn Errors of Metabolism. *J Inherit Metab Dis* **24**: 139-150.

Cooper GR, Brostrom CO and Brostrom MA (1997) Analysis of the Endoplasmic Reticular Ca^{2+} Requirement for α_1-Antitrypsin Processing and Transport Competence. *Biochem J* **325**: 601-608.

Covey TR, Bonner RF, Shushan BI and Henion J (1988) The Determination of Protein, Oligonucleotide and Peptide Molecular Weights by Ion-Spray Mass Spectrometry. *Rapid Commun Mass Spectrom* **2**: 249-256.

Cox DW (2001) α_1–Antitrypsin Deficiency. In: The Metabolic and Molecular Bases of Inherited Diseases IV 8th edition, 2001, p.5559-5584, Scriver CR, Beaudet AL, Sly WS and Valle D (eds.) McGraw-Hill, New York.

Cox DW and Billingsley GD (1989) Rare Deficiency Types of α_1–Antitrypsin: Electrophoretic Variation and DNA Haplotypes. *Am J Hum Genet* **44**: 844-854.

Cox DW, Billingsley GD and Callahan JW (1986) Aggregation of Plasma Z Type α_1–Antitrypsin Suggests Basic Defect for the Deficiency. *FEBS Lett* **205**: 255-260.

Cox DW, Johnson AM and Fagerhol MK (1980) Report of Nomenclature Meeting for α_1-Antitrypsin. INSERM. Rouen/Bois-Guillaume-1978. *Hum Genet* **53**: 429-433.

Creaser CS, Reynolds JC and Harvey DJ (2002) Structural Analysis of Oligosaccharides by Atmospheric Pressure Matrix-Assisted Laser Desorption/Ionisation Quadrupole Ion Trap Mass Spectrometry. *Rapid Commun Mass Spectrom* **16**: 176-184.

Crimaudo C, Hortsch M, Gausepohl H and Meyer DI (1987) Human Ribophorins I and II: the Primary Structure and Membrane Topology of Two Highly Conserved Rough Endoplasmic Reticulum-Specific Glycoproteins. *EMBO J* **6**: 75-82.
Davis BJ (1964) Disc Electrophoresis-II: Method and Application to Human Serum Proteins. *Ann N.Y. Acad Sci* **121**: 404-427.

Dawson G and Ellory JC (1985) Functional Lysosomal Hydrolase Size as Determined by Radiation Inactivation Analysis. *Biochem J* **226**: 283-288.

De Virgilio M, Weninger H and Ivessa NE (1998) Ubiquitination Is Required for the Retro-Translocation of a Short-Lived Luminal Endoplasmic Reticulum Glycoprotein to the Cytosol for Degradation by the Proteasome. *J Biol Chem* **273**: 9734-9743.

Dell A (1987) F.A.B.-Mass spectrometry of Carbohydrates. *Adv Carbohydr Chem Biochem* **45**: 19-72.

Dell A and Morris HR (2001) Glycoprotein Structure Determination by Mass Spectrometry. *Science* **291**: 2351-2356.

DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen YD, Su YA and Trent JM (1996) Use of a cDNA Microarray to Analyse Gene Expression Patterns in Human Cancer. *Nature Genetics* **14**: 457-460.

Diettrich O, Mills K, Johnson AW, Hasilik A and Winchester BG (1998). Application of Magnetic Chromatography to the Isolation of Lysosomes From Fibroblasts of Patients With Lysosomal Storage Disorders. *FEBS Letts* **441**: 369-72.

Dole M, Mack LL, Hines RL, Mobley RC, Ferguson LD and Alice MB (1968) Molecular Beams of Macroions. *J Chem Phys* **42**: 22-40.

Dove A (1999) Proteomics: Translating Genomics into Products? *Nature Biotech* **17**: 233-236.

Dunn MJ (1995a) Electrophoretic Analysis Methods. In: Protein Purification Methods, 1995, p. 18-39, Harris ELV and Angal S (eds.) IRL Press, Oxford.

Dunn MJ (1995b) Determination of Total Protein Concentration. In: Protein Purification Methods, 1995, p. 10-18, Harris ELV and Angal S (eds.) IRL Press, Oxford.

Edmunds T and Zhang K (1999) Personal Communication Regarding the Analysis of Ceredase and Cerezyme Glycosylation by Genzyme Inc.

Ellgaard L and Helenius A (2001) ER Quality Control: Towards an Understanding at the Molecular Level. *Curr Op Cell Biol* **13**: 431-437.

Errington DM, Bathurst IC, Janus ED and Carell RW (1982) In Vitro Synthesis of M and Z Forms of Human α_1-Antitrypsin. *FEBS Lett* **148**: 83-86.

Fagerhol MK (1967) Serum Pi Types in Norwegians. *Acta Path Microbiol Scand* **70**: 421 - 428.

Fagerhol MK and Laurell CB (1967) The Polymorphism of 'Prealbumins', α_1–Antitrypsin in Human Sera. *Clin Chim Acta* **16**:199-203.

Fan JQ and Lee YC (1997) Detailed Studies on Substrate Structure Requirements of Glycoamidases A and F. *J Biol Chem* **272**: 27058-27064.
FEBS Lett **296**: 300-304.

Fu J, Ren M and Kreibich G (1997) Interactions Among Subunits of the Oligosaccharyltransferase Complex. *J Biol Chem* **272**: 29687-29692.

Gavel Y and von Heijne G (1990) Sequence Differences Between Glycosylated and Non-Glycosylated Asn-X-Thr/Ser Acceptor Sites: Implications for Protein Engineering. *Protein Eng* **3**: 433-442.

Gennaro LA, Delaney J, Vouros P, Harvey DJ and Domon B (2002) Capillary Electrophoresis / Electrospray Ion Trap Mass Spectrometry for the Analysis of Negatively Charged Derivatized and Underivatised Glycans. *Rapid Commun Mass Spectrom* **16**: 192-200.

Gharahdaghi F, Weinberg CR, Meagher DA, Imai BS and Mische SM (1999) Mass Spectrometric Identification of Proteins From Silver-Stained Polyacrylamide Gel: a Method for the Removal of Silver Ions to Enhance Sensitivity. *Electrophoresis* **20**: 601-605.

Gorelick FS and Shugrue C (2001) Exiting the Endoplasmic Reticulum. *Mol Cell Endocrin* **177**: 13-18.

Gorg A, Obermaier C, Boguth G, Harder A, Scheibe B, Wildgruber R, Weiss W (2000) The Current State of Two-Dimensional Electrophoresis With Immobilized pH Gradients. *Electrophoresis. 6:1037-1053.*

Goya N, Miyazaki S, Kodate S and Ushio B (1972) A Family of Atransferrinemia. *Blood* **40**: 239-245.

Grabowski GA, Barton MW, Pastores G, Dambrosia JM, Banerjee TK, McKee MA, Parker C, Schiffmann R, Hill SC and Brady RO (1995) Enzyme Therapy in Type-1 Gaucher Disease - Comparative Efficacy of Mannose-Terminated Glucocerebrosidase From Natural and Recombinant Sources. *Ann Intern Med* **122**: 33-39.

Grabowski GA, Pastores G, Brady RO and Barton NW (1993) Safety and Efficacy of Macrophage Targeted Recombinant Glucocerebrosidase Therapy. *Pediatr Res* **33**: 139A

Grace ME and Grabowski GA (1990) Human Acid Beta-Glucosidase: Glycosylation is Required for Catalytic Activity. *Biochem Biophys Res Commun* **168**: 771-777.

Grace ME and Grabowski GA (1993) Molecular Enzymology of Acid β-Glucosidase. In: Glucosidases: Biochemistry and Molecular Biology, 1993, p. 40-56, Esen R (ed.) ACS Press, Washington DC.

Grubenmann CE, Frank CG, Kjaergaard S, Berger EG, Aebi M and Hennet T (2002) ALG12 Mannosyltransferase Defect in Congenital Disorder of Glycosylation Type Ig. *Hum Mol Gen* **11**: 2331-2339.

Grunewald S, Matthijs G, and Jaeken J (2002) Congenital Disorders of Glycosylation: a Review. *Pediatr Res* **52**: 618-624.

Guile GR, Rudd PM, Wing DR, Prime SB and Dwek RA (1996) A Rapid High-Resolution High-Performance Liquid Chromatographic Method for Separating Glycan Mixtures and Analysing Oligosaccharide Profiles. *Anal Biochem* **240**: 210-226.

Gusev AI and Wilkinson WR, Proctor A and Hercules DM (1995) Improvement in Signal Reproducibility and Matrix / Co-Matrix Effects in MALDI Analysis. *Anal Chem* **67**: 1034-1041.

Gusev AI, Wilkinson WR, Proctor A and Hercules DM (1996) Direct Quantitative Analysis of Peptides Using Matrix Assisted Laser Desorption Ionisation. *Fresenius J Anal Chem* **354**: 455-463

Hale JE, Butler JP, Knierman MD and Becker GW (2000) Increased Sensitivity of Tryptic Peptide Detection by MALDI TOF MS-TOF Mass Spectrometry is Achieved by Conversion of Lysine to Homoarginine. *Anal Biochem* **287**: 110-117.

Hanrahan S, Charlwood J, Tyldesley R, Langridge J, Bordoli R, Bateman R and Camilleri P (2001). Facile Sequencing of Oligosaccharides by Matrix-Assisted Laser Desorption / Ionisation on a Hybrid Quadrupole Orthogonal Acceleration Time-of-Flight Mass Spectrometer. *Rapid Commum Mass Spectrom* **15**: 1141-115.

Harduin-Lepers A, Vallejo-Ruiz V, Krzewinski-Recchi MA, Samyn-Petit B, Julien S and Delannoy P (2001) The Human Sialyltransferase Family. *Biochimie* **83**: 727-737.

Harnik-Ort V, Prakash K, Marcantonio E, Colman DR, Rosenfeld MG, Adensik M, Sabatini DD and Kribich G (1987) Isolation and Characterisation of cDNA Clones for Rat Ribophorin I: Complete Coding Sequence and in Vitro Synthesis and Insertion of the Encoded Product into Endoplasmic Reticulum Membranes. *J Cell Biol* **104**: 855-863.

Hart G (1997) Dynamic O-Linked glycosylation of Nuclear and Cytoskeletal Proteins. *Ann Rev Biochem* **66**: 315-335.

Harvey DJ (1999) Matrix-Assisted Laser Desorption / Ionization Mass Spectrometry of Carbohydrates. *Mass Spectrometry Reviews* **18**: 349-451.

Harvey DJ (2000) Electrospray Mass Spectrometry and Fragmentation of N-Linked Carbohydrates Derivatised at the Reducing Terminus. *J Am Soc Mass Spectrom* **11**: 900-915.

Harvey DJ (2001) Identification of Protein-Bound Carbohydrates by Mass Spectrometry. *Proteomics* **1**: 311-328.

Helenius A and Aebi M (2001) Intracellular Functions of N-Linked Glycans. *Science*: **291** 2364-2369.

Helenius J, Ng DTW, Marolda CL, Walter P, Valvano MA and Aebi M (2002) Translocation of Lipid-Linked Oligosaccharides Across the ER Membrane Requires Rft1 Protein. *Nature* **415**: 447-450.

Hennet T (2002) The Galactosyltransferase Family. *Cell Mol Life Sci* **59**: 1081-1095.

Henry H, Froehlich F, Perret R, Tissot J-D, Eilers-Messerli B, Lavanchy D, Dionisi-Vici C, Gonvers J-J and Bachmann C (1999) Microheterogeneity of Serum Glycoproteins in Patients with Chronic Alcohol Abuse Compared with Carbohydrate-Deficient Glycoprotein Syndrome Type I. *Clin Chem* **45**: 1408-1413.

Henzel WJ, Billeci TM, Stults JT, Wong SC, Grimley C and Watanabe C (1993) Identifying Proteins From 2-Dimensional Gels by Molecular Mass Searching of Peptide-Fragments in Protein-Sequence Databases. *P Natl Acad Sci USA* **90**: 5011-5015.

Hercz A, Katona E, Cutz E, Wilson JR and Barton M (1978) Alpha1-Antitrypsin: the Presence of Excess Mannose in the Z Variant Isolated From Liver. *Science* **201**:1229-1232.

Herscovics A (2002) Importance of Glycosidases in Mammalian Glycoprotein Biosynthesis. *Biochim Biophys Acta* **1473**: 96-107.

Hirose M (2000) The Structural Mechanism for Iron Uptake and Release by Transferrins. *Biosci Biotechnol Biochem* **64**: 1328-1336.

Hochstrasser DF and Merril CR. (1988) 'Catalysts' for Polyacrylamide Gel Polymerization and Detection of Proteins by Silver Staining. *Appl Theor Electrophor* **1**:35-40.

Hochstrasser DF, Patchornik A and Merril CR (1988) Development of Polyacrylamide Gels That Improve the Separation of Proteins and Their Detection by Silver Staining. *Anal Biochem* **173**:412- 423

Ikebe N, Akaike T, Miyamoto Y, Hayashida K, Yoshitake J, Ogawa M and Maeda H (2000) Protective Effect of S-Nitrosylated α_1-Protease Inhibitor on Hepatic Ischemia-Reperfusion Injury. *J Pharmacol Exp Ther* **295**: 904-911.

Imperiali B and Hendrickson TL (1995) Asparagine-Linked Glycosylation: Specificity and Function of Oligosaccaryl Transferase. *Bioorg Med Chem* **3**: 1565-1578.

Jachymek W, Niedziela T, Petersson C, Lugowski C, Czaja J and Kenne L (1999) Structures of the O-Specific Polysaccharides from *Yokenella Regensburgei* (*Koserella Trabulsii*) Strains PCM 2476, 2477, 2478, and 2494: High-Resolution Magic-Angle Spinning NMR Investigation of the O-Specific Polysaccharides in Native Lipopolysaccharides and Directly on the Surface of Living Bacteria. *Biochem* **38**: 11788-11795.

Jaeken J, van Eyk HG, van der Heul C, Corbeel L, Eeckels R and Eggermont E (1984) Sialic Acid Deficient Serum and Cerbrospinal Fluid Transferrin in a Newly Regonized Genetic Syndrome. *Clin Chim Acta* **144**: 245-247.

Jaeken J, Vanderschueren-Lodeweyckx M, Casaer P, Snoeck L, Corbeel L, Eggermont E and Eeckels R (1980) Familial Psychomotor Retardation With Markedly Fluctuating Serum Prolactin, FSH and GH Levels, Partial TBG Deficiency, Increased Serum Arylsulphatase A and Increased CSF Protein: a New Syndrome? *Pediatr Res* **14**: 179.

Janciauskiene S and Eriksson S (1994) The Interaction of Hydrophobic Bile Acids With the α_1-Proteinase Inhibitor. *FEBS Lett* **343**: 141-145.

Janus ED, Phillips NT and Carrell RW (1985) Smoking, Lung Function, and α_1-Antitrypsin Deficiency. *Lancet* **19**: 152-154.

Jeppsson J-O (1976) Amino Acid Substitution Gly-Lys in α_1-Antitrypsin Pi Z. *FEBS Lett* **65**: 195-197.

Jeppsson JO, Lilja H and Johansson M (1985) Isolation and Characterisation of Two Minor Fractions of α_1–Antitrypsin by High Performance Liquid Chromatography Chromatofocusing. *J Chromatogr* **327**: 173-177.

Kaplan HA, Welply JK and Lennarz WJ (1987) Oligosaccharyl Transferase: the Central Enzyme in the Pathway of Glycoprotein Assembly. *Biochim Biophys Acta* **906**: 161-173.

Karas M and Hillenkamp F (1988) Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Daltons. *Anal Chem* **60**: 2299-2301.

Kaufmann R (1995) Matrix-Assisted Laser Desorption Ionization (MALDI TOF MS) Mass Spectrometry: a Novel Analytical Tool in Molecular Biology and Biotechnology. *J Biotechnol* **41**: 155-175.

Keir G, Winchester BG and Clayton P (1999) Carbohydrate-Deficient Glycoprotein Syndromes: Inborn Errors of Protein Glycosylation. *Ann Clin Biochem* **36**: 20-36.

Kelleher DJ and Gilmore R (1997) DAD1, the Defender Against Apoptotic Cell Death, Is a Subunit of the Mammalian Oligosaccharyltransfease. *Proc Natl Acad Sci USA* **94**: 4994-4999.

Kelleher DJ, Kreibich G and Gilmore R (1992) Oligosaccharyltransferase Activity is Associated With a Protein Complex Composed of Ribophorins I and II and a 48kD Protein. *Cell* **69**: 55-65.

Keough T (1985) Cationization of Organic-Molecules Using Fast Atom Bombardment Mass-Spectrometry. *Anal Chem* **58**: 2027-2034.

Kim D and Yu M-H (1996) Folding Pathway of Human α_1-Antitrypsin: Characterisation of an Intermediate That is Active but Prone to Aggregation. *Biochem Biophys Res Commun* **226**: 378-384.

Knauer R and Lehle L (1999) The Oligosaccharyltransferase Complex From Yeast. *Biochim Biophys Acta* **1426**: 259-273.

Knoell DL, Ralston DR, Coulter KR and Wewers MD (1998) Alpha-1-Antitrypsin and Protease Complexation is Induced by Lipopolysaccharide, Interleukin-1-Beta, and Tumor Necrosis Factor-Alpha in Monocytes. *Am J Respir Crit Care Med* **157**: 246-255.

Kornfeld R and Kornfeld S (1985) Assembly of Asparagine-Linked Oligosaccharides. *Ann Rev Biochem* **54**: 631-664.

Krause E, Wenschuh H and Jungblut PR (1999) The Dominance of Arginine-Containing Peptides in MALDI TOF MS-Derived Tryptic Mass Fingerprints of Proteins. *Anal Chem* **71**: 4160-4165.

Kussmann M and Roepstorff P (2000) Sample Preparation Techniques for Peptides and Proteins Analyzed by MALDI-MS. *Methods Mol Biol* **146**: 405-24.

Küster B, Wheeler SF, Hunter AP, Dwek RA, Harvey DJ (1997) Sequencing of N-Linked Oligosaccharides Directly From Protein Gels: In-Gel Deglycosylation Followed by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry and Normal-Phase High-Performance Liquid Chromatography. *Anal Biochem* **250**: 82-101.

Laemmli UK (1970) Cleavage of Structural Proteins During the Assembly of the Head of Bacteriophage T4. *Nature* **227**: 680-685.

Lander ES, Linton LM, Birren B *et al.* (2001) Initial Sequencing and Analysis of the Human Genome. *Nature* **409**: 860-921.

Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, Brown PO and Davis RW (1997) Yeast Microarrays for Genome Wide Parallel Genetic and Gene Expression Analysis. *Proc Natl Acad Sci USA* **94**: 13057-13062.

Laurell C-B and Eriksson S (1963) The Electrophoretic α_1-Antitrypsin Deficiency. *Scand J Clin Lab Invest* **15**: 132-133.

Lee C, Levin A and Branton D (1987) Copper Staining: A Five-Minute Protein Stain for SDS-Polyacrylamide Gels. *Anal Biochem* **166**: 308-312.

Lejeune PJ, Mallet B, Farnarier C and Kaplanski S (1989) Changes in Serum Level and Affinity for Concanavalin-A of Human Alpha-1-Proteinase Inhibitor in Severe Burn Patients-Relationship to Natural-Killer Cell-Activity. *Biochim Biophys Acta* **990**: 122-127.

Liang P and Pardee AB (1992) Differential Display of Eukaryotic Messenger-RNA by Means of the Polymerase Chain-Reaction. *Science* **257**: 967-971.

Lord JM, Davey J, Frigerio L and Roberts LM (2000) Endoplasmic Reticulum-Associated Protein Degradation. *Semin Cell Dev Biol* **11**: 159-164.

Lovegrove J, Jeremiah D, Gillet GT, Temple IK, Povey S and Whitehouse DB (1997) A New α_1–Antitrypsin Mutation, THR-MET85 ($PIZ_{Bristol}$) Associated with Novel Electrophoretic Properties. *Ann Hum Genet* **61**: 385-391.

Mann M, Hojrup P and Roepstorff P (1993) Use of Mass-Spectrometric Molecular-Weight Information to Identify Proteins in Sequence Databases. *Biol Mass Spectrom* **22**: 338-345.

Mann M, Meng CK and Fenn JB (1989) Interpreting Mass-Spectra of Multiply Charged Ions. *Anal Chem* **61**: 1702-1708.

Maret A, Potier M, Salvayre R and Douste-Blazy L (1983) Modification of Subunit Interaction in Membrane-Bound Acid Beta-Glucosidase from Gaucher Disease. *FEBS Lett* **160**: 93-97.
Maret A, Salvayre R, Negre A and Douste-Blazy L (1981) Properties of the Molecular Forms of Beta-Glucosidase and Beta-Glucocerebrosidase from Normal Human and Gaucher Disease Spleen. *Eur J Biochem* **115**: 455-461.

Mechref Y and Novotny MV (2002) Structural Investigations of Glycoconjugates at High Sensitivity. *Chem Rev* **102**: 321-369.

Mellquist JL, Kasturi L, Spitalnik SL and Shakin-Eshleman SH (1998) The Amino Acid Following an Asn-X-Ser/Thr Sequon is an Important Determinant of N-Linked Core Glycosylation Efficiency. *Biochemistry-US* **37**: 6833-6837.

Meyer TS and Lamberts BL (1965) Use of Coomassie Brilliant Blue R250 for the Electrophoresis of Microgram Quantities of Parotid Saliva Proteins on Acrylamide-Gel Strips. *Biochim Biophys Acta* **107**:144-145.

Miele RG, Castellino FJ and Bretthauer RK (1997) Characterization of the Acidic Oligosaccharides Assembled on the Pichia Pastoris-Expressed Recombinant Kringle 2 Domain of Human Tissue-Type Plasminogen Activator. *Biotechnol Appl Biochem* **26**: 79-83.

Mills K, Johnson AW, Clayton PT, Diettrich OGP and Winchester BG (2000) A Strategy for the Identification of Site Specific Glycosylation in Glycoproteins Using MALDI TOF MS. *Tetrahedron-Assymetry* **11**: 75-93.

Mills K, Mills PB, Clayton PT, Johnson AW, Whitehouse D and Winchester BG (2001a) Identification of α_1-Antitrypsin Variants in Plasma With the Use of Proteomic Technology. *Clin Chem* **47**: 2012-2022.

Mills PB, Mills K, Clayton PT, Johnson AW, Whitehouse D and Winchester BG (2001) Congenital Disorders of Glycosylation Type I Leads to Altered Processing of N-Linked Glycans, as Well as Underglycosylation. *Biochem J* **359**: 249–254.

Miyamoto Y, Akaike T and Maeda H (2000) *S*-Nitrosylated Human α1-Protease Inhibitor. *Biochim Biophys Acta* **1477**: 90-97.

Molinari M and Helenius A (2000) Chaperone Selection During Glycoprotein Translocation into the Endoplasmic Reticulum. *Science* **288**: 331-333.

Montreuil J, Genevieve S and Mazurier J (1997) Transferrin Superfamily. In: Glycoproteins II, 1997, p. 203-241 Montreuil J, Vliegenthart JFG and Schachter H (eds.) Elsevier Science BV, Amsterdam.

Moremen KW (2002) Golgi Alpha-Mannosidase II Deficiency in Vertebrate Systems: Implications for Asparagine-Linked Oligosaccharide Processing in Mammals. *Biochim Biophys Acta* **1573**: 225-235.
Moremen KW, Trimble RB and Herscovics A (1994) Glycosidases of the Asparagine-Linked Oligosaccharide Processing Pathway. *Glycobiology* **4**: 113-125.

Munro S (2001) What Can Yeast Tell us About N-Linked Glycosylation in the Golgi Apparatus? *FEBS Lett* **498**: 223-227.

Naitoh A, Aoyagi Y and Asakura H (1999). Highly Enhanced Fucosylation of Serum Glycoproteins in Patients With Hepatocellular Carcinoma. *J Gastroenterol Hepatol* **14**: 436-445.

Nakashima T, Sekiguchi T, Kuraoka A, Fukushima K, Shibata Y, Komiyama S and Nguyen DN, Becker GW and Riggin RM (1995) Protein Mass Spectrometry: Applications to Analytical Biotechnology. *J Chromatogr A* **705**: 21-45.

Nicola AJ, Gusev AI, Proctor A, Jackson EK and Hercules DM (1995) Application of the Fast-Evaporation Sample Preparation Method for Improving Quantification of Angiotensin II by Matrix-Assisted Laser Desorption/Ionization. *Rapid Commun Mass Spectrom* **9**:1164-1171.

Nishikawa A and Mizuno S (2001) The Efficiency of N-linked Glycosylation of Bovine DNAase I Depends on the Asn-Xaa-Ser/Thr Sequence and the Tissue of Origin. *Biochem J* **355**: 245-248.

Nishimoto T (1993) Molecular Cloning of a Human cDNA Encoding a Novel Protein, DAD1, Whose Defect Causes Apoptotic Cell Death in Hamster BHK21 Cells. *Mol Cell Biol* **13**: 6367-6374.

Nishimoto T (1993) Molecular-Cloning of a Human cDNA-Encoding a Novel Protein, DAD1, Whose Defect Causes Apoptotic Cell-Death in Hamster BHK-21-Cells. *Mol Cell Biol* **13**: 6367-6374.

Nukiwa T, Brantly M, Ogushi F, Fells G, Satoh K, Stier L, Courtney M and Crystal RG (1987) Characterization of the M1(Ala213) Type of α1-Antitrypsin, a Newly Recognized Common 'Normal' α1-Antitrypsin Haplotype. *Biochemistry* **26**: 5259-5267.

O'Farrell PH (1975) High Resolution 2D-PAGE Electrophoresis of Proteins. *J Biol Chem* **250**: 4007-4021.

O'Neill RR, Tokoro T, Kozak CA and Brady RO (1989) Comparison of the Chromosomal Localization of Murine and Human Glucocerebrosidase Genes and of the Deduced Amino Acid Sequences. *Proc Natl Acad Sci USA* **86**: 5049-5053.

Ornstein L (1964) Disc Electrophoresis-I: Background and Theory. *Ann N.Y. Acad Sci* **121**: 321-349.
Ortiz ML, Calero M, Patron CF, Castellanos L and Mendez E (1992) Imidazole-SDS-Zn Reverse Staining of Proteins in Gels Containing or Not SDS and Microsequence of Individual Unmodified Electroblotted Proteins.

Osiecki-Newman KM, Fabbro D, Dinur T, Boas S, Gatt S, Legler G, Desnik RJ and Grabowski GA (1986) Human Acid Beta-Glucosidase: Affinity Purification of the Normal Placental and Gaucher Disease Splenic Enzymes on *N*-Alkyl-Deoxynojirmycin-Sepharose. *Enzyme* **35**: 147-153.

Paakko P, Kirby M, du Bois RM, Gillissen A, Ferrans VJ and Crystal RG (1996) Activated Neutrophils Secrete Stored Alpha-1-Antitrypsin. *Am J Respir Crit Care Med* **154**: 1829-1833.

Packer NH, Lawson MA, Jardine DR, Sanchez JC and Gooley AA (1998) Analyzing Glycoproteins Separated by Two-Dimensional Gel Electrophoresis. *Electrophoresis* **19**: 981-988.

Papac DI, Wong A and Jones AJ (1996) Analysis of Acidic Oligosaccharides and Glycopeptides by Matrix-Assisted Laser Desorption/Ionisation Time-of-Flight Mass Spectrometry. *Anal Chem* **15**: 3215-3223.

Pappin DJC, Hojrup P and Bleasby AJ (1993) Rapid Identification of Proteins by Peptide-Mass Fingerprinting. *Curr Biol* **3**: 327-332.

Parodi AJ (2000) Protein Glycosylation and its Role in Protein Folding. *Annu Rev Biochem* **69**: 69-93.

Patton WF (2001) Detecting Proteins in Polyacrylamide Gels and on Electroblot Membranes. In: Proteomics from Protein Sequence to Function, 2001, p. 65-86, Pennington SR and Dunn MJ (eds.) Bios Scientific Publishers Ltd, Oxford.

Pipkorn R, Boenke C, Gehrke M, Hoffmann R (2002) High-Throughput Peptide Synthesis and Peptide Purification Strategy at the Low Micromol-Scale Using the 96-Well Format. *J Pept Res* **59**:105-114.

Qu D, Teckman JH, Omura S and Perlmutter DH (1996) Degradation of a Mutant Secretory Protein, α_1-Antitrypsin Z, in the Endoplasmic Reticulum Requires Proteasome Activity. *J Biol Chem* **271**: 22791-22795.

Rahbek-Nielsen H, Roepstorff P, Reischl H, Wozny M, Koll H and Haselbeck A (1997) Glycopeptide Profiling of Human Urinary Erythropoietin by Matrix-Assisited Laser Desorption/Ionization Mass Spectrometry. *J Mass Spectrom* **32**: 948-958.

Reason AJ, Dell A, Romero PA and Herscovics A (1991) Specificity of the Mannosyltransferase Which Initiates Outer Chain Formation in Saccharomyces Cerevisiae. *Glycobiology* **1**: 387-391.

Reynolds JA and Tanford C (1970) The Gross Conformation of Protein-Sodium Dodecyl Sulfate Complexes. *J Biol Chem* **243**: 5161-5165.

Roepstorff P and Fohlman J (1984) Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed Mass Spectrom* **11**: 601.

Rosenberg S, Barr PJ, Najarian RC and Hallewell RA (1984) Synthesis in Yeast of a Functional Oxidation-Resistant Mutant of Human Alpha-Antitrypsin. *Nature* **312**: 77-80.

Rosenfeld J, Capdevielle J, Guillemot JC and Ferrara P (1992) In-Gel Digestion of Proteins for Internal Sequence Analysis After One- or Two-Dimensional Gel Electrophoresis. *Anal Biochem* **203**: 173-179.

Rossen R, Schade AL, Butler WT and Kasel JA (1966) The Proteins in Nasal Secretion: a Longitudinal Study of the GammaA-Globulin, GammaG-Globulin, Albumin, Siderophilin, and Total Protein Concentrations in Nasal Washings From Adult Male Volunteers. *J Clin Invest* **45**: 768-776.

Roth J (2002) Protein N-Glycosylation Along the Secretory Pathway: Relationship to Organelle Topography and Function, Protein Quality Control, and Cell Interactions. *Chem Rev* **102**: 285-303.

Samandari T and Brown JL (1993) A Study of the Effects of Altering the Sites for N-Glycosylation in α-1-Proteinase Inhibitor Variants M and S. *Prot Sci* **2**: 1400-1410.

Sanchez JC, Rouge V, Pisteur M, Ravier F, Tonella L, Moosmayer M, Wilkins MR and Hochstrasser DF (1997) Improved and Simplified In-Gel Sample Application Using Reswelling of Dry Immobilized pH Gradients. *Electrophoresis* **18**: 324-327.

Sato Y and Beutler E (1993) Binding, Internalization, and Degradation of Mannose-Terminated Glucocerebrosidase by Macrophages. *J Clin Invest* **91**: 1909-1917.

Schachter H (1986) Biosynthetic Controls that Determine the Branching and Microheterogeneity of Protein-Bound Oligosaccharides. *Biochem Cell Biol* **64**: 163-181.

Schachter H (1991) The 'Yellow Brick Road' to Branched Complex N-Glycans. *Glycobiology* **1**: 453-461.

Schachter H (2000) The Joys of HexNAc. The Synthesis and Function of N- and O-Glycan Branches. *Glycoconjugate J* **17**: 465-483.

Schachter H (2001a) Biosynthesis of N- and O-Glycans. In: Glyko catalogue - Tools For Glycobiology 2001-2002 pgs 7.9-7.11.

Schachter H (2001b) Congenital Disorders Involving Defective N-Glycosylation of Proteins. *Cell Mol Life Sci* **58**: 1085-1104.

Schmitt S, Glebe D, Alving K, Tolle TK, Linder M, Geyer H, Linder D, Peter-Katalinic J, Gerlich W.H and Geyer R (1999) Analysis of the Pre-S2 N- and O-Linked Glycans of the M Surface Protein from Human Hepatitis B Virus. *J Biol Chem* **274**: 11945-11957.

Schultze HE, Gollner I, Heide K, Schonenberger M and Schwick G (1955) Zur Kenntnis der Alpha Globuline des Menschlichen Normalserums. *Z Naturforsch* **10**: 463-466.

Schwick H (1977) Therapy of Atransferrinemia with Transferrin. *J Clin Chem Biochem* **16**: 73.

Shakin-Eshleman SH, Spitalnik SL and Kasturi L (1996) The Amino Acid at the X Position of an Asn-X-Ser Sequon is an Important Determinant of N-Linked Core-Glycosylation Efficiency. *J Biol Chem* **271**: 6363-6366.

Sharp HL (1971) α_1-Antitrypsin Deficiency. *Hosp Pract* **5**: 83.

Sharp HL, Bridges RA and Krivit W (1969) Cirrhosis Associated With α_1-Antitrypsin Deficiency: A Previously Unrecognized Inherited Disorder. *J Lab Clin Med* **73**: 934-939.

Shevchenko A, Wilm M, Vorm O and Mann M (1996) Mass Spectrometric Sequencing of Proteins From Silver-Stained Polyacrylamide Gels. *Anal Chem* **68**: 850-858.

Silberstein S, Kelleher DJ and Gilmore R (1992) The 48-KDa Subunit of the Mammalian Oligosaccharyltransferase Complex is Homologous to the Essential Yeast Protein WBP1. *J Biol Chem* **267**: 23658-23663.

Siuzdak G (1994) The Emergence of Mass Spectrometry in Biomedical Research. *Proc Natl Acad Sci USA* **91**: 11290-11297.

Spengler B, Kirsch D, Kaufmann R and Jaeger E (1992) Peptide Sequencing by Matrix-Assisted Laser-Desorption Mass Spectrometry. *Rapid Commun Mass Spectrom* **6**: 105-108.

Spiro RG (2002) Protein Glycosylation: Nature, Distribution, Enzymatic Formation, and Disease Implications of Glycopeptide Bonds. *Glycobiology* **12**: 43R-56R.

Stahl B, Klabunde T, Witzel H, Krebs B, Steup M, Karas M and Hillenkamp F (1994) The Oligosaccharides of the Fe(III)-Zn(II) Purple Acid Phosphatase of the Red Kidney Bean. Determination of the Structure by a Combination of Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry and Selective Enzymic Degradation. *Eur J Biochem* **220**: 321-330.

Staudacher E, Altmann F, Wilson IBH and März L (1999) Fucose in N-Glycans: From Plant to Man. *Biochim Biophys Acta* **1473**: 216-236.

Stibler H and Jaeken J (1990) Carbohydrate Deficient Serum Transferrin in a New Systemic Hereditary Syndrome. *Arch Dis Child* **63**: 107-111.

Sunner JA, Kulatunga R and Kebarle P (1986) Fast-Atom-Bombardment and Gas-Phase Basicities. *Anal Chem* **58**: 1312-1316.

Sutton-Smith M, Morris HR, Dell A (2000) A Rapid Mass Spectrometric Strategy Suitable for the Investigation of Glycan Alterations in Knockout Mice. *Tet: Asymm* **11**: 363-369.

Svenger T (1985) Plasma Protease Inhibitors in α_1-Antitrypsin-Deficient Children. *Pediatr Res* **19**: 834-838.

Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y and Yoshida T (1988) Protein and Polymer Analysis up to m/z 100,000 by Laser Ionisation Time-of-Flight Mass Spectrometry. *Rapid Commun Mass Spectrom* **2**: 151-153.

Tarentino AL and Plummer TH (1994) Substrate Specificity of Flavobacterium Meningosepticum Endo F2 and Endo F3: Purity is the Name of the Game. *Glycobiology* **4**: 771-773.

Thiede B, Lamer S, Mattow J, Siejak F, Dimmler C, Rudel T and Jungblut PR (2000) Analysis of Missed Cleavage Sites, Tryptophan Oxidation and N-Terminal Pyroglutamylation After In-Gel Tryptic Digestion. *Rapid Commun Mass Spectrom* **14**: 496-502.

Thompson JJ (1899) On the Masses of the Ions in Gases at Low Pressures. *Philosophical Magazine* **48**: 547-567.

Thorstensen K and Romslo I (1990) The Role of Transferrin in the Mechanism of Cellular Iron Uptake. *Biochem J* **271**: 1-10.

TofSpec E & SE Operator Manual (1997) Overview of TofSpec E & SE Instruments pgs 9-18.

Torgerson DF, Skowronski RP and Macfarlane R D (1974) New Approach to the Mass Spectroscopy of Non-Volatile Compounds. *Biophys Res Commun* **60**: 616-621.

Travis, J, Owen M, George P, Carrell R, Rosenberg S, Hallewell, RA and Barr PJ (1985) Isolation and Properties of Recombinant DNA Produced Variants of Human α_1-Proteinase Inhibitor. *J Biol Chem* **260**: 4384-91985.

Tremblay LO and Herscovics A (2000) Characterization of a cDNA Encoding a Novel Human Golgi Alpha 1,2-Mannosidase (IC) Involved in N-Glycan Biosynthesis. *J Biol Chem* **275**: 31655-31660.

Tsarbopoulos A, Becker GW, Occolowitz JL and Jardine I (1988) Peptide and Protein Mapping by CF-252-Plasma Desorption Mass-Spectrometry. *Anal Chem* **171**: 113-123.

Tsuji S, Choudary PV, Martin BM, Winfield S, Barranger JA and Ginns EI (1986) Nucleotide Sequence of cDNA Containing the Complete Coding Sequence for Human Lysosomal Glucocerebrosidase. *J Biol Chem* **261**: 50-53.

Ueda M., Mashiba S and Uchida K (2002) Evaluation of Oxidized Alpha-1-Antitrypsin in Blood as an Oxidative Stress Marker Using Anti-Oxidative α1-AT Monoclonal Antibody. *Clinica Chimica Acta* **317**: 125-131.

Van Eijk HG, Van Noort WL, De Jong G and Koster JF (1987) Human-Serum Sialo Transferrins in Diseases. *Clin Chim Acta* **165**: 141-145.

Van Schaftingen E and Jaeken J (1995) Phosphomannomutase Deficiency is a Cause of Carbohydrate-Deficient Glycoprotein Syndrome Type I. *FEBS Lett* **377**: 318-320.

Varki A (1993) Biological Roles of Oligosaccharides all of the Theories are Correct? *Glycobiology* **3**: 97-130.

Varki A, Cummings R, Esko J, Freeze H, Hart G and Marth J (1999a) N-Glycans. In: E ssentials o f G lycobiology, 1 999, p . 8 5-100, V arki A , C ummings R , E sko J, Freeze H, Hart G and Marth J (eds.) Cold Spring Harbor L aboratory P ress, New York.

Varki A, Cummings R, Esko J, Freeze H, Hart G and Marth J (1999b) Evolution of Glycan Diversity. In: Essentials of Glycobiology, 1999, p. 31-40, Varki A, Cummings R, Esko J, Freeze H, Hart G and Marth J (eds.) Cold Spring Harbor Laboratory Press, New York.

Vashist S, Kim W, Belden WJ, Spear ED, Barlowe C and Ng DTW (2001) Distinct Retrieval and Retention Mechanisms are Required for the Quality Control of Endoplasmic Reticulum Protein Folding. *J Cell Biol* **155**: 355-368.

Vaughan L, Lorier M and Carrell RW (1982) α_1–Antitrypsin Microheterogeneity: Isolation and the Physiological Significance of Isoforms. *Biochem Biophys Acta* **701**: 339-345.

Velculescu VE, Zhang L, Vogelstein B and Kinzler KW (1995) Serial Analysis of Gene Expression. *Science* **270**: 484-487.

Vestal M and Juhasz P (1998) Resolution and Mass Accuracy in Matrix-Assisted Desorption Ionization-Time-of-Flight. *J Am Soc Mass Spectrom* **9**: 892-911.

Vorm O, Roepstorff P and Mann M (1994) Improved Resolution and Very High Sensitivity in MALDI TOF of Matrix Surfaces Made by Fast Evaporation. *Anal Chem* **66**: 3281-3287.

Wada Y, Nishikawa A, Okamoto N, Inui K, Tsukamoto H, Okada S and Taniguchi N (1992) Structure of Serum Transferrin in Carbohydrate-Deficient Glycoprotein Syndrome. *Biochem Biophys Res Commun* **189**: 832-836.

Wasinger VC, Cordwell SJ, CerpaPoljak A Yan JX, Gooley AA, Wilkins MR, Duncan MW, Harris R and Williams KL (1995) Progress With Gene-Product Mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* **16**: 1090-1094.

Watson JD and Crick FHC (1953) A Structure for Deoxyribose Nucleic Acid. *Nature* **171**: 737-738.

Welch S and Longmead L (1990) A Comparison of the Structure and Properties of Normal Transferrin and a Genetic Variant of Human Transferrin. *Int J Biochem* **22**: 275-282.

Westbrook JA, Yan JX, Wait R, Welson SY and Dunn MJ (2001) Zooming-in on the Proteome: Very Narrow-Range Immobilised pH Gradients Reveal More Protein Species and Isoforms. *Electrophoresis* **22**: 2865-2871.

Wiley WC and McLaren IH (1955) Time-of-Flight Mass Spectrometer with Improved Resolution. *Rev Sci Instrum* **26**: 1150-1157.

Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF and Williams KL (1996) Progress with Proteome Projects: Why All Proteins

Expressed by a Genome Should be Identified and How to do it. *Biotechnol Genet Eng* **13**: 19-50.

Williams J (1982) The Evolution of Transferrin. *Trends Biochem Sci* **8**: 272-275.

Wong SF, Meng CK and Fenn JB (1988) Multiple Charging in Electrospray Ionization of Poly(Ethylene Glycols). *J Phys Chem* **92**: 546-550.

Wu Y, Whitman I, Molmenti E, Moore K, Hippenmeyer P and Perlmutter DH (1994) A Lag in Intracellular Degradation of Mutant α_1-Antitrypsin Correlates With the Liver Disease Phenotype in Homozygous PiZZ α_1-Antitrypsin Deficiency. *Proc Natl Acad Sci USA* **91**: 9014-9018.

Yamashita K, Ideo H, Ohkura T, Ideo H, Ohno K and Kanai M (1993) Electrospray Ionization-Mass Spectrometric Analysis of Serum Transferrin Isoforms in Patients with Carbohydrate-Deficient Glycoprotein Syndrome. *J Biochem (Tokyo)* **114**: 766-769.

Yates JR, Speicher S, Griffin PR and Hunkapiller T (1993) Peptide Mass Maps – a Highly informative Approach to Protein Identification. *Anal Biochem* **214**: 397-408.

Yoshida L, Liberman J, Gaidulis L and Ewing C (1976) Molecular Abnormality of Human α1-Antitrypsin Variant (Pi Z) Associated with Plasma Activity Deficiency. *Proc Natl Acad Sci U S A* **73**: 1324 - 1327.

Zamfir A and Peter-Katalinic J (2001) Glycoscreening by On-Line Sheathless Capillary Electrophoresis / Electrospray Ionization-Quadrupole Time of Flight-Tandem Mass Spectrometry. *Electrophoresis* **22**: 2448-2457.

Zenobi R and Knochenmuss R (1998) Ion Formation in MALDI Mass Spectrometry. *Mass Spectrom Rev* **17**: 337-366.

Zhang W, Czernik AJ, Youngwirth T, Aebersold R and Chait B (1995) Matrix-Assisted Laser Desorption Mass Spectrometric Peptide Mapping of Proteins Separated by Two-Dimensional Gel Electrophoresis: Determination of Phosphorylation in Synapsin I. *Protein Sci* **3**: 677-686.