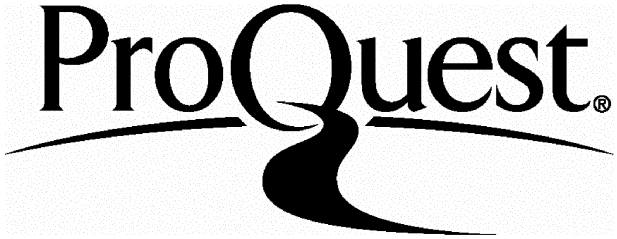
Characterisation of O-GlcNAc Modification Using Mass Spectrometry

ROBERT J. CHALKLEY

DEPARTMENT OF BIOCHEMISTRY AND
MOLECULAR BIOLOGY
UNIVERSITY COLLEGE LONDON
GOWER STREET
LONDON

This thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy from the University of London



Abstract

This thesis describes analytical biochemistry studies in the field of glycobiology, focusing on the use of advanced mass spectrometric techniques as a tool for studying the addition of single N-acetylglucosamine residues to intracellular proteins.

Transient glycosylation of nuclear and cytoplasmic proteins by O-GlcNAc has been detected in all eukaryotes studied. This modification is believed to control protein activity, stability and function, and performs a similar, but possibly antagonistic role in controlling the protein to that mediated by phosphorylation. However, a functional role of the modification is only known for a small number of proteins. Consequently, new methods for detecting and characterising the site specificity for this modification will enable more detailed structure/function studies to be carried out, and will provide the basic information required to elucidate the molecular details of biological activity of this post-translational modification. It is the development of these methods that is the subject of this thesis.

It is shown that through the application of quadrupole orthogonal-acceleration TOF technology, mass spectrometry can be used to identify sites of O-GlcNAc modification. Firstly, methods are developed for determining sites of modification of O-GlcNAcylated synthetic peptides. Next, these methods are employed to identify a previously known site of GlcNAc modification on α-crystallin, through analysing solution and in-gel digests of gel purified material. A detailed study of the post-translational state of serum response factor is carried out. In this work previously reported sites of phosphorylation and GlcNAc modification are confirmed and novel sites of O-GlcNAc and phosphate modification are identified.

This thesis demonstrates that a mass spectrometric approach is significantly more sensitive than previous techniques employed, and should become an important tool in the development of our understanding of this modification. The background and relevance of this advance in analytical methods, and its potential use in determining the role of O-GlcNAc modification is also discussed.

Abbreviations

The standard abbreviations for amino acids (1 letter and 3 letter) have been used. These are not listed below as their definitions can be found in any biochemistry textbook. For all abbreviations their definition is also given the first time they are introduced in the text.

BSA	Bovine serum albumin
CID	Collision-induced dissociation
CHCA	Alpha cyano 4-hydroxy cinnamic acid
Cys-Am	Acrylamide-modified cysteine
DIGE	Differential gel electrophoresis
DE	Delayed extraction
DHB	2,5-Dihydroxybenzoic acid
ECL	Electrochemical luminescence
ESI	Electrospray ionisation
FAB-MS	Fast atom bombardment - mass spectrometry
Gal	Galactose
GlcNAc	N-acetylglucosamine
GFAT	Glutamine:fructose-6-phosphate amidotransferase
HexNAc	N-acetylhexosamine
HPLC	High pressure liquid chromatography
HRP	Horseradish peroxidase
ICAT	Isotope coded affinity tag
IEF	Isoelectric focusing
IPG	Immobilised pH gradient
LINAC	Linear Acceleration
LC-MS	Liquid chromatography interfaced to mass spectrometry
Lys-C	Lysine specific endopeptidase
MALDI	Matrix-assisted laser desorption / ionisation

Met-Ox	Oxidised Methionine
MS-MS	Tandem mass spectrometry
Mw	Molecular weight
NEPHGE	Non-equilibrium pH gradient gel electrophoresis
O-GlcNAc	O-linked N-acetylglucosamine
OaTOF	Orthogonal-acceleration time-of-flight
PAGE	Polyacrylamide gel electrophoresis
PSD	Post-source decay
pI	Isoelectric point
Pro-C	Proline-specific endopeptidase
Q-TOF	Quadrupole oaTOF
RCA I	*Ricinus communis* agglutinin I lectin
SDS	Sodium dodecyl sulphate
sHSP	Small heat shock protein
SRE	Serum response element
SRF	Serum response factor
TIC	Total ion chromatogram
TOF	Time-of-flight
UV	Ultra-violet
VRC	Vanadyl ribonucleoside complex
WGA	Wheat germ agglutinin lectin

Acknowledgements

There are many people without whose help this thesis would not have been achieved. When I first started at the Ludwig Institute, I was greatly helped by the Dr. Elaine Stimson, Dr. Rainer Cramer, and the late Dr. Willy Richter; a more friendly man you could not hope to meet. This tradition has been continued by each new member of the lab, and I would like to thank all past and present members of the lab for the help they have given. A special mention must be given to Nick Totty, whose enthusiasm for research was infectious, and with whom many enjoyable discussions were held on the poor state of the English cricket team.

I am indebted to two people for the supply of samples. Prof. Gerry Hart donated GlcNAc-modified synthetic peptides, and these have been valuable standards for use in development of new experimental techniques. Secondly, I must thank Rob Nicholas at the ICRF. I had reached an impasse in my research where I had shown I could identify sites in synthetic peptides but I did not have a real sample to work upon. He was very generous in his gift of purified serum response factor, and this protein has contributed the largest part to the research in this thesis.

On the occasions when I have emerged from the dungeons of the mass spectrometry labs I have been greatly assisted by Dr. Rob Stein, Richard Foxon and the late Craig Brooks, who have been a great source of information and advice, and seemed to already have made up every buffer I ever needed.

I would also like to thank Prof. Mike Waterfield for his advice and supervision.

Of course, the biggest thank you must be extended to Prof. Al Burlingame. Having never heard of O-GlcNAcylation before this project started, I am grateful for his choice of an interesting topic of research, and I hope that through my work I have made many more people aware of this little known widespread modification.

Finally, I must extend my appreciation to my father, who as well as putting up with me around the house, has provided an efficient catering, laundry and secretarial service.

Table of Contents

List of Figures

List of Tables

List of Appendices

1. INTRODUCTION

1.1. Mass Spectrometry

1.1.1. What is a Mass Spectrometer?

A mass spectrometer is an instrument used to measure the mass to charge (m/z) ratio of gaseous ions, and is composed of three fundamental parts. The analyte is ionised in the 'source', converting it to gaseous ions. These are then separated on the basis of their m/z ratio in the 'analyser', before reaching a 'detector'.

This introduction will focus solely on technology utilised in the mass spectrometers used for this thesis.

1.1.2. Ionisation Methods

There are two major ionisation methods currently used for mass spectrometric analysis of biomolecules; matrix-assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI). Both of these ionisation methods produce ions by a proton transfer mechanism to form quasimolecular ions; protonated ions in positive ion mode and deprotonated ions in negative ion mode. The majority of the ions observed in MALDI positive ion spectra are $[M + H]^+$ ions. In the presence of salts, adduct ions such as $[M + Na]^+$ and $[M + K]^+$ ions are also formed. Ions formed by ESI are generally multiply-protonated, and peptides are most commonly observed as $[M + 2H]^{2+}$ or $[M + 3H]^{3+}$ ions, although higher degrees of protonation may be observed depending on the number of basic functions in any given peptide sequence.

1.1.2.1. MALDI

The formation of gaseous ions by firing a laser at a crystalline sample was first demonstrated in 1967[1]. However, it was the utilisation of a matrix chromophore capable of absorbing the laser energy [2, 3] that imparted practical analytical utility to the technique for biological samples. In matrix-assisted laser desorption/ionisation (MALDI), analyte is mixed with an excess of matrix molecules on the surface of a metal probe or plate. As the solution dries the analyte is incorporated into the matrix crystals.

When these are irradiated with a laser, the matrix absorbs the laser energy, causing matrix and analyte molecules to be desorbed from the surface of the metal plate to form a plume of gaseous ions and neutral molecules. During plume formation there is proton transfer from matrix to analyte molecules to create analyte ions. This ionisation method deposits less internal energy into the analyte ions than if desorbed directly from the laser shot, allowing fragile ions, such as biomolecules, to remain intact during the desorption/ionisation process. Hence, it is described as a 'soft' ionisation process [4].

The majority of MALDI is performed using ultraviolet (UV) lasers, with the most common lasers being nitrogen lasers, which emit at 337nm. Hence, the major requirements of the matrix are the ability to ionise sample molecules when irradiated at this wavelength, and to be stable under high vacuum and possibly catalyse decarboxylation that imparts the initial velocity into the plume [5]. The most commonly used matrices for peptide analysis are α-cyano-4-hydroxycinnamic acid (CHCA) [6] and 2,5-dihydroxybenzoic acid (DHB) [7].

Some matrices form homogenous spots of crystals, whilst others form crystals around the rim of the spot. Analyte molecules co-crystallise with the matrix, whilst salts and many other contaminants do not become incorporated into the crystals. This makes MALDI more tolerant of salts.

In order to reduce thermal transfer of energy during laser irradiation, desorption is performed using short duration laser shots of 1-200 ns. Hence, MALDI is a pulsed ionisation method and is ideal to interface with pulsed analysers such as time-of-flight (TOF) instruments (Section 1.1.3.1).

1.1.2.2. ESI

The other major ionisation method for biological samples is electrospray ionisation (ESI). Ions are formed by spraying the sample through a narrow capillary at a high voltage, causing the formation of a fine mist of highly charged droplets containing analyte and buffer molecules. As the buffer evaporates, often aided by a flow of drying gas, droplets shrink. There are two theories as to how the final charged analyte ions are formed. One theory states that each analyte molecule is in a separate droplet. The buffer evaporates from the droplet to leave the charged ions [8]. The other proposal is that

droplets contain many analyte ions. As the droplets shrink by buffer evaporation, the charge density of the droplets increases until a critical point at which it becomes unstable charged analyte ions are ejected from the droplet due to coulombic repulsion (Figure 1.1) [9].

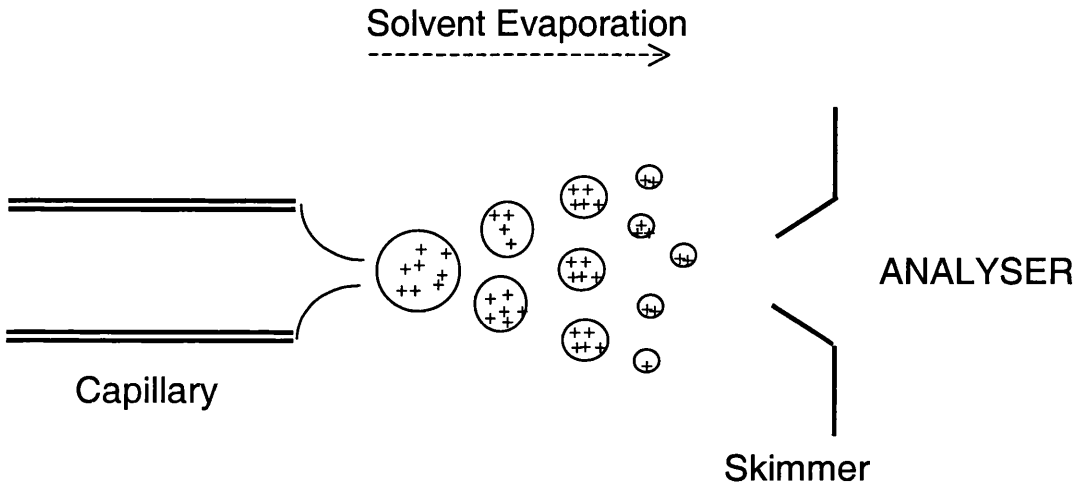


Figure 1.1: The mechanism of electrospray ionisation.
Sample is sprayed from a capillary at high voltage, forming a mist of highly charged droplets. As the solvent evaporates the charge density of the droplet increases until analyte ions are ejected.

ESI is another soft ionisation technique that is applicable to the analysis of proteins and oligonucleotides [10]. It is a concentration-dependent technique. Thus, if the sample is introduced in a smaller volume at a lower flow rate ESI-MS becomes more sensitive. Nanospray-MS (electrospray-MS at flow rates of nl/min) has become a pivotal tool in the identification of proteins and peptides [11].

Electrospray is a continuous ionisation method and as it works in the liquid phase it can be interfaced directly with liquid chromatographic separation strategies such as high performance liquid chromatography (HPLC) [12] and capillary electrophoresis (CE) [13].

1.1.3. Analysers

1.1.3.1. Time-of-Flight

Separating ions by TOF technology is fundamentally simple. Ions are accelerated to the same kinetic energy and are separated on the basis that the heavier the molecule the lower velocity it will have, so the longer it will take to travel a given distance to a detector:

$$zU = \frac{1}{2} m \frac{s^2}{t^2} \qquad \Rightarrow \qquad t \approx \sqrt{\frac{m}{z}}$$

where U is the acceleration voltage between the ion source and extraction grid, and z is the charge of the ion.

The first TOF mass spectrometers were not sensitive because they operated in a scanning mode where only a narrow timeframe of ions were detected in each scan [14]. By moving the time window, a complete spectrum could eventually be recorded.

The major breakthrough in sensitivity was realised upon the invention of time-to-digital recorders [15]. These allow all the ions from a TOF separation to be recorded in one scan, making TOF analysis more sensitive than alternative scanning analysers such as magnetic sectors and quadrupoles.

A thorough description of the principles of TOF analysis has been produced by Cotter [16].

1.1.3.2. Quadrupoles

Quadrupoles generally consist of a set of four parallel rods. A DC voltage is applied to all four rods, with two opposing rods given a positive potential and the other two rods a negative potential. An alternating RF potential is applied to the rods, and at a given RF to DC ratio only ions of a certain m/z are able to pass through this mass filter, whereas all other m/z ions are unstable and collide with rods. By changing the RF to DC ratio a mass range can be scanned.

Quadrupole mass analysers are relatively inexpensive to make and can be very small in size. However, they suffer from a limited mass range of generally up to m/z 2000 –

4000. As they are scanning analysers they are significantly less sensitive than TOF analysers and produce spectra with significantly worse resolution, although by scanning a narrower mass range quadrupole mass resolution can be improved. They are, however, very useful in QTOF instruments (see Section 1.1.4.3), where they are used to select precursor ions for collisional activation.

1.1.4. Types of Mass Spectrometers

There are many different types of mass spectrometers. This review of different instruments only encompasses types of instrument used during this thesis.

1.1.4.1. MALDI-TOF

There are three main factors in ion formation that can impair the resolution and sensitivity of data in TOF separation: (1) temporal distribution; (2) spatial distribution; (3) initial velocity distribution. The development of lasers with short pulse durations and fast rise-times minimises the time variation in ion formation by MALDI. In MALDI, ions are usually formed from a sample plate that is perpendicular to the extraction grid. Thus, all ions are ideally produced at the same potential in the extraction field. However in reality, during ionisation a plume of ions is formed which have a distribution of different locations and velocities.

In MALDI ionisation initial spatial distribution is minimal, but ions are formed with a wide distribution of velocities. Partial correction for this velocity distribution can be achieved by introducing a short delay between the ionisation and acceleration of ions for TOF separation; a process referred to as time-lag-focusing or delayed extraction (DE) [17]. This converts the distribution of initial velocities into spatial distribution within the extraction field. The further the ion has travelled towards the extraction grid, the less it will experience the extraction field. The ions with less initial velocity will experience more of the extraction field and catch up the other ions at a point referred to as the 'space-focus plane'. By positioning the detector on this plane one compensates for the initial spatial distribution. Using a dual stage extraction allows the space-focus plane to be moved, so the detector can be placed at the end of a long TOF tube. This delay alters

the time zero for flight time measurement, such that the conversion of flight time to m/z is usually performed by calibration of a spectrum with known internal standards.

Kinetic energy distribution of ions can be compensated for after extraction using a reflectron [18]. This consists of an electric field mirror at the end of a first field-free drift region that retards ions then reflects them back into a second drift region. Ions of a certain m/z that have slightly more kinetic energy penetrate further into the reflectron so take longer to turn round into the second drift region. A detector placed at the end of the first drift region records a 'linear' spectrum, whilst the spectrum from a detector at the end of the second drift region is referred to as a 'reflectron' spectrum. Due to the compensation for kinetic energy distribution, reflectron spectra are better resolved than linear spectra. Figure 1.2 shows the configuration of a DE-MALDI-TOF instrument with a reflectron.

The introduction of a reflectron ion mirror also facilitates a form of fragmentation analysis called 'post source decay' (PSD) [19]. If an ion fragments in the source, the fragment ions will be separated according to their new masses. However, if fragmentation occurs during the first field-free region of the TOF, the fragment ions will continue to travel with the velocity of their parent ion mass until they reach the reflectron, but will travel with the velocity relating to their new mass in the second field free region. Hence, they are observed at an apparent mass in between their parent ion mass and fragment mass. Peaks formed by ions that have undergone fragmentation in the first drift region are known as 'metastable peaks'. These peaks are not properly focused onto the detector due to the change in kinetic energy, so produce poorly resolved peaks. In order to focus these ions one has to decrease sequentially the mirror voltage ratio of the reflectron to focus and record the whole mass range.

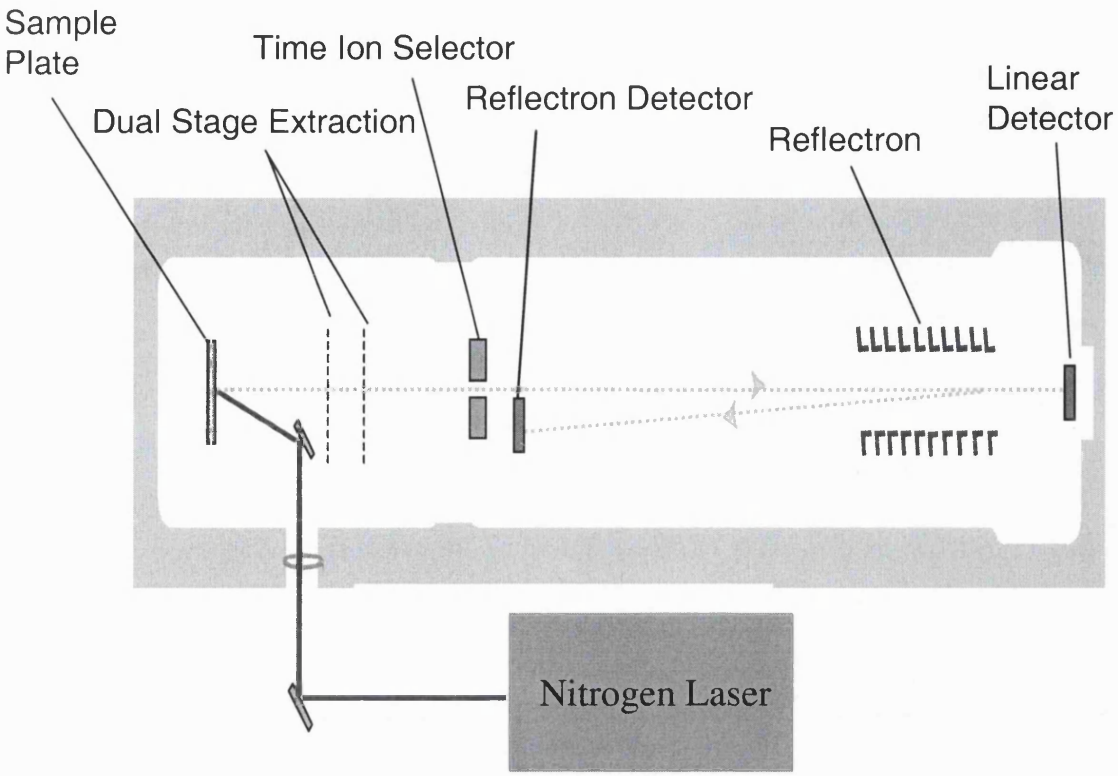


Figure 1.2: A schematic of a MALDI-TOF instrument with a reflectron.
Analyte on the sample plate is ionised by a nitrogen laser. After a short delay, ions are accelerated into the TOF analyser using a dual stage extraction. Ions can either be detected at the linear detector, or be directed by the reflectron ion mirror to the reflectron detector.

PSD spectra are acquired using an elevated laser fluence, as this promotes unimolecular dissociation. An ion is selected with a 'time-ion selector', which only allows ions into the TOF tube for a specified narrow window of time, *i.e.* m/z. Spectra are then acquired at different ion mirror voltage ratios to focus fragments of different masses. The spectra at different mirror voltage ratios are finally 'stitched' together to produce a full fragmentation spectrum.

1.1.4.2. ESI-TOF

Recently, TOF mass analysers have also been interfaced to ESI. The interfacing of TOF separation to ESI instruments is complicated by ESI being a continuous ionisation

method, whereas TOF mass analysis functions in a pulsed mode. Placing the analyser orthogonal to the ESI ion path has circumvented this problem. Ions are introduced into a push-pull region at the beginning of the TOF tube, where packets of ions are orthogonally accelerated in a pulsed manner through the mass analyser to the detector.

Placing the analyser orthogonal to the ion path has advantages. In this configuration, ionisation is completely decoupled from mass analysis. Hence, all the problems with initial time, space and kinetic energy distribution of ions are obviated, allowing an improvement in mass accuracy and resolution. There is also a dramatic reduction in chemical noise in the spectrum. On the other hand, only a fraction of the ions formed are resolved and detected, as a new packet of ions cannot be orthogonally accelerated for mass analysis until the previous TOF-MS separation has been completed. This feature is known as the 'duty cycle' of the instrument, and has an adverse impact on the machine's sensitivity.

1.1.4.3. Q-TOF

An adaptation of the ESI-TOF instrument is the quadrupole-TOF (QTOF) [20, 21]. A mass resolving quadrupole and a collision cell are placed before the TOF analyser (Figure 1.3). This allows the production of tandem mass spectra.

Prior to the commercial availability of these instruments, the majority of tandem mass spectrometry was performed on triple-quadrupole instruments. In these instruments the first quadrupole selects the precursor ion, the second acts as a collision cell and the third separates the fragments produced. As quadrupoles are scanning devices, these instruments only detect a small mass range of fragment ions formed at any given voltage ratio. By substituting the final quadrupole with a oa-TOF analyser, all fragment ions can be detected in a single scan, providing a significant increase in sensitivity.

However, there are types of tandem mass spectrometry experiments that are more sensitive when performed on triple quadrupole instruments. Precursor ion scanning experiments, where one identifies parent ions that fragment to form an ion of a certain mass, require the first mass spectrometer to scan the mass range whilst the second mass spectrometer only detects a certain mass. In TOF-MS analysis, a complete spectrum up to the mass of the desired product ion must be acquired, whereas a quadrupole does not

need to scan and can detect all ions at a given mass. Therefore, QTOF instruments have a lower duty cycle than triple quadrupole instruments for this type of experiment, making them less sensitive for precursor ion scanning. The sensitivity of a QTOF instrument can be improved using a linear acceleration (LINAC) collision cell[22]. In these instruments this collision cell is used as a decelerator and traps ions, releasing them in packets to coincide with acceleration into the TOF tube. By varying the pulse delay time for release of ions from the LINAC cell, the sensitivity for certain mass ranges can be improved. This technology has been incorporated into the QSTAR Pulsar (Applied Biosystems/MDS-Sciex), which has permitted a significant increase in sensitivity for precursor ion scanning [23].

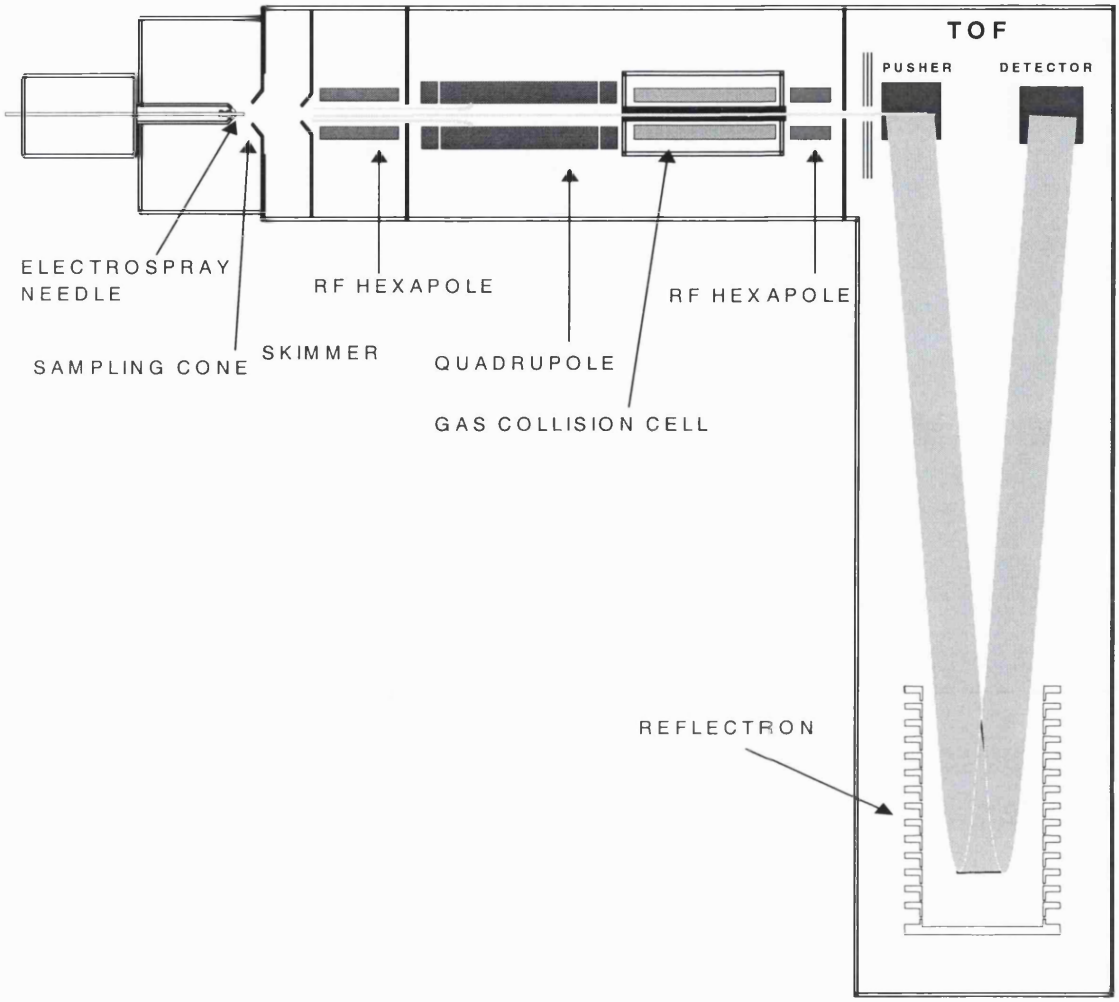


Figure 1.3: Typical design of a quadrupole-TOF instrument.

Ions are formed at the electrospray needle. In MS mode, ions are focused until they reach the pusher region of the TOF, where they are accelerated orthogonally at discrete timepoints. For MS-MS experiments a parent ion is selected by the quadrupole. This is then fragmented in the gas collision cell, before fragment ions are orthogonally accelerated and separated in the TOF analyser.

1.1.4.4. MALDI-Q-TOF

MALDI sources have been interfaced to Q-TOF instruments[24], producing mass spectrometers with better resolution and mass accuracy than axial configuration MALDI-TOF instruments, and significantly less background noise. MALDI-Q-TOF instruments

also contain a collision cell, which provides greater control over fragmentation than by PSD in a conventional axial MALDI-TOF mass spectrometer [25]. However, these instruments are not as sensitive as axial MALDI-TOF instruments, at the moment mainly due to inferior source design, but also due to their poorer duty cycle. Q-TOF machines are designed for continuous ion formation, so MALDI-Q-TOF instruments employ high repetition lasers to simulate continuous ion formation and thus deplete the sample rapidly. Conversely, their ability to acquire a complete resolved fragmentation spectrum in each scan, without the need to stitch together spectra acquired at different mirror voltage ratios as is necessary in MALDI-PSD, should increase their sensitivity for fragmentation analysis.

1.2. Proteomics

The proteome is the complete set of proteins expressed from the genome, including post-translational modifications of these proteins[26]. Hence, proteomics is essentially the cataloguing of all the proteins expressed in a system. Unfortunately, it is difficult to assign biochemical significance to the majority of results produced by this approach. Thus, the term 'functional proteomics' has been coined, and has been defined as 'the use of proteomics methods to monitor and analyse the spatial and temporal properties of the molecular networks and fluxes involved in living cells. Conversely, functional proteomics is also the use of proteomics methods to identify molecular species that participate in such networks via functional stimulation, perturbation, or isolation of these networks.'[27]

This is carried out at two different levels. At one level it is the identification of many proteins in a mixture, whilst at another level it is the characterisation of a given protein's post-translational modification state.

For mass spectrometric identification of proteins it is desirable to analyse a single protein at a time, or to have as simple a mixture as possible. Thus, for a complex starting material such as a cell lysate, powerful separation methods are employed prior to mass spectrometric analysis.

1.2.1. 2D-PAGE

One of the most powerful separation methods is two-dimensional polyacrylamide gel electrophoresis (2D-PAGE)[28, 29]. Proteins migrate in the first dimension on the basis of their isoelectric point (pI) using isoelectric focusing (IEF), then in a second dimension by their molecular weight using SDS-PAGE. This technique is not only able to separate different proteins but also resolves between versions of the same protein that are differentially post-translationally modified. For example, phosphorylation makes a protein more acidic, so will cause the protein to migrate to a different pI in the first dimension of separation. 2D-PAGE has been reported to successfully separate nearly 9000 different protein features[30]. However, despite this impressive resolution, mass spectrometry often identifies multiple proteins in each spot on the gel.

2D-PAGE is very good at visualising differences between samples. Many proteomic experiments involve comparing treated cells (such as growth factor stimulated) with untreated cells. 2D-PAGE images of the two samples are compared to look for upregulation/downregulation or the appearance/disappearance of spots, which are then excised from the gel and identified by mass spectrometry. The production of reproducible 2D gels is a skilled art, and images are never completely superimposable. The analysis of these gels can be a slow process, but there is software developed specifically for comparing gels, such as Melanie (Geneva Bioinformatics, Switzerland) and PDQuest (BioRad, Hercules, CA, USA).

The development of differential gel electrophoresis (DIGE) should make sample comparison by 2D-PAGE simpler [31]. In this technique, the samples to be compared are labelled with different fluorescent tags. The samples are then combined and separated by 2D-PAGE on the same gel. By scanning this gel at the different wavelength of excitation of each fluorescent dye one creates an image of each sample. As the samples are resolved on the same gel the images can be superimposed, and using different colours for each, one can easily see the appearance and disappearance of features.

There are several protein staining methods. The most common is Coomassie Brilliant Blue staining (Coomassie), which is rapid and simple. The Coomassie binds to the SDS molecules that coat proteins separated by SDS-PAGE. However, it is not particularly sensitive, with a lower limit of detection of 50 – 100 ng of a protein on the gel. Silver

staining is more sensitive with a limit of detection of one – two ng of protein. This stain modifies the protein itself. Many silver staining protocols use glutaraldehyde, making them incompatible with mass spectrometry, but mass spectrometry friendly silver staining protocols have been developed [32]. Spots stained with these protocols are often analysed without destaining, but the quality of data from a spot that is silver stained is generally significantly worse than if the same spot was Coomassie stained. Destaining silver stained spots improves the quality of data obtained [33-35].

Recently there have been significant developments using fluorescent dyes. The latest of these are as sensitive as silver staining but have wider dynamic ranges and are more quantitative than silver [36]. The problem with these stains is that they are not visible to the naked eye. Thus, in order to excise proteins of interest a spot-cutting robot is required; otherwise a second stain must be utilised to observe the spot before manual excision.

1.2.1.1. Drawbacks of 2D-PAGE

Although 2D gels have very good resolving ability, there are limitations to the technique. There are a select number of proteins that appear to be regularly identified by 2D-PAGE, and low abundance proteins such as signalling proteins are rarely detected. Differences observed using this technique when looking at complex mixtures such as cell lysates, are in relatively abundant proteins, which are often not the proteins of interest. Therefore, 2D-PAGE is ideal for identifying global changes such as markers for disease states, but is usually less successful at discerning signalling cascades, as signalling proteins are generally expressed at low levels in the cell.

Separation of proteins by IEF is efficient for proteins with isoelectric points of pI 5 – pI 9, but resolution is poor for proteins of very acidic or basic pI. For basic proteins an alternative first dimensional separation is non-equilibrium pH gradient gel electrophoresis (NEPHGE)[37]. In this technique all proteins are loaded at the positive end of the gel and during electrophoresis the basic proteins migrate to the negative end of the gel. However, unlike IEF where the proteins migrate to their isoelectric point then stop moving, in NEPHGE proteins will migrate off the gel if left too long. Thus,

electrophoresis is stopped at a user-defined time. This means NEPHGE gels vary significantly from one to another.

There are also problems with losing proteins during 2D-PAGE. Acidic and basic proteins sometimes precipitate during the first dimension, many large proteins (greater than 150 kDa) do not enter the second dimension gel and small proteins (less than 10 kDa) generally migrate off the bottom of the gel. Hydrophobic membrane proteins are also poorly represented due to their poor solubility in electrophoresis buffers. 2D-PAGE protocols designed for membrane proteins have been developed, but have had limited success [38].

1.2.2. Alternatives to 2D-PAGE

Different liquid chromatographies can be combined to form two dimensional separation strategies [39]. Strong cation exchange has been combined with reverse phase chromatography to characterise a protein complex [40], and this technology was applied to large scale cataloguing of a yeast proteome, where 1484 proteins were identified in a single experiment [41]. Proteins identified included proteins of very high and low pH, low abundance proteins such as transcription factors, high molecular weight proteins and membrane spanning proteins. Hence, this approach appears to be less biased than 2D-PAGE towards types of proteins identified, and identifies a wider variety of components in a sample than 2D-PAGE. The disadvantage is that it is not quantitative, so it cannot be used for comparing samples.

An interesting alternative to a 1D gel is the use of non-porous reverse phase high performance liquid chromatography (HPLC) to separate proteins, followed by mass spectrometric detection of eluting proteins [42]. The total ion chromatogram (TIC) can be used to plot a virtual 1D gel, which can be compared using software in the same way one would compare lanes on a 1D gel. UV chromatograms can also be acquired which can give quantitative data. By post-column splitting the sample and collecting fractions, when a change in pattern is observed in the virtual 1D gel images, the fraction of interest can be digested and the proteins identified. This approach gives higher resolution than a 1D gel, and is able to detect co-eluting proteins. It will detect many low mass proteins

lost by 1D SDS-PAGE, but the largest proteins identified in this study were only 90 kDa [42].

Approaches have been proposed using isotopic labelling of proteins and mass spectrometry to quantify changes in protein levels. Isotope-coded affinity tags (ICAT) [43] are tags that contain two functional parts. An iodoacetamide group at one end of the tag molecule reacts with free cysteine residues. At the other end of the tag is a biotin moiety, which is used to affinity purify the tag and bound peptides from mixtures using avidin beads. Between these two groups is a linker. In one form of the tag the linker contains eight deuteriums, whilst the other tag contains hydrogens. This means the two tags have the same chemical properties, but one tag is 8 Da greater in mass than the other.

The protocol for ICAT comparison of two samples is summarised in Figure 1.4. Proteins from the two samples to be compared are reduced to create free thiol groups on cysteines. These are labelled using the ICAT reagent. Samples are then mixed, and the proteins are digested. Cysteine-containing peptides are isolated using avidin affinity chromatography, then samples are analysed by LC-MS and LC-MS-MS. Identical peptides from the two samples elute at similar retention times and by comparing the intensity of the two peaks in the LC-MS spectrum, relative changes in protein concentration can be extrapolated. The individual peptides are then identified from their fragmentation spectra to determine the protein they were derived from.

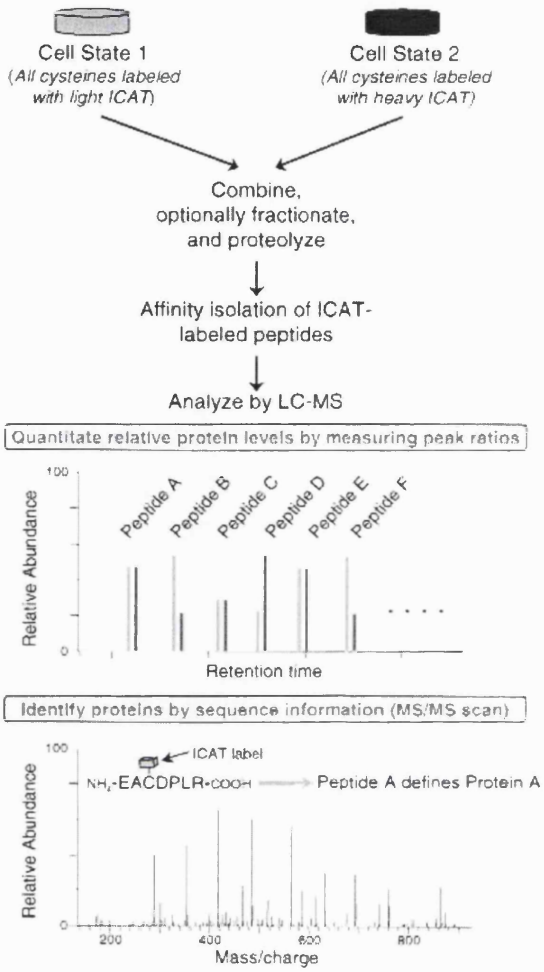


Figure 1.4: Overview of the process of ICAT labelling for subsequent mass spectrometric quantification of proteins.

Samples from cell states 1 and 2 are labelled using light or heavy ICAT reagent. Samples are combined and digested. Cysteine-containing peptides are affinity isolated using the biotin on the tag. These peptides are then analysed by mass spectrometry, and from the ratio of peaks containing the different tags, relative quantification can be made between the two samples. This figure is reproduced from Gygi et al.[43].

1.3. Characterising Proteins by Mass Spectrometry

Although mass spectrometry can measure an accurate molecular mass of an intact protein, this mass is not sufficient to identify it, as many proteins are cleaved to smaller products and most proteins carry some form of post-translational modification.

Hence, the general strategy for protein identification involves enzymatic cleavage of proteins into peptides, which are then analysed to identify the protein. The favoured enzyme is trypsin, which cleaves C-terminal to lysine and arginine residues, producing peptides mostly in the mass range of 1000 - 3000 Da. Having produced a protein digest there are two strategies for identifying the protein.

1.3.1. Peptide Mass Fingerprinting

A spectrum of the protein digest is acquired, generally on a MALDI-TOF instrument. The masses of all the peptides observed are entered into a search engine and can be used to identify the protein, as the specificity of cleavage of the enzyme is known. For every entry in a protein database one can create a list of theoretical masses of peptides that would be produced if the protein was digested using a certain enzyme. Hence, each protein in the database will create a characteristic 'mass fingerprint' of peptides. The list of peak masses observed in the mass spectrum is compared with the theoretical 'mass fingerprints' of all proteins in the database to determine the identity of the protein.

1.3.2. Peptide Sequence Tag

The alternative method for identifying the protein is to select individual peptides and fragment them to determine their sequence. This is best carried out by ESI-CID-MS-MS. Electrospray low energy CID spectra of tryptic peptides are relatively simple to interpret. Tryptic peptides are generally observed as doubly and triply charged ions by ESI-MS. When fragmented the charge is preferentially retained on the C-terminal basic residue (tryptic peptides end in arginines or lysines). Hence, a ladder of sequence ions extending from the C-terminus is observed. There are several approaches to interpreting the fragmentation data. The simplest approach is to enter the masses of all the fragment ions and search against theoretical fragment ions of all peptides in the database that have the correct parent ion mass and are predicted to be formed after cleavage using a specified

enzyme. This search is performed using a program such as MS-TAG [44]. Alternatively, a short sequence can be interpreted from the fragmentation spectrum, and by combining this short sequence (minimum three residues) with the fragment masses either side this sequence and the parent ion mass, the peptide can be identified using the program PeptideSearch [45]. If a long stretch of sequence can be interpreted from the fragmentation spectrum, then the search can be carried out with sequence alone, using, for example, MS-Pattern [44]. Each of these different search methods have their advantages and disadvantages, so searching data by more than one approach will increase the chance of identifying the peptide.

Many laboratories analyse their digests by LC-ESI-MS-MS. From one LC-MS run a large number of MS-MS spectra can be acquired. Software that can automatically interpret MS-MS spectra and combine the matches from all the spectra is a vital time-saving tool, and two search engines are able to do this: SEQUEST [46] and MASCOT [47].

In many cases the protein analysed does not contain a database entry for the species of interest. When a highly homologous database entry from another species exists, protein identification is usually straightforward. However, if the homology is low, new software has been developed that can combine several stretches of sequence data to give confident assignments [48].

1.3.3. Analysing Post-translational Modifications using Mass Spectrometry

The majority of proteins are post-translationally modified. The best-studied modifications are phosphorylation and glycosylation, but other modifications such as acetylation, sulphation or attachment of a fatty acid also take place [49]. These modifications can affect the structure, stability and function of the protein.

1.3.3.1. Mass Spectrometry and Phosphorylation

Phosphate groups are added to the amino acids serine, threonine, tyrosine and occasionally histidine [50] and aspartic acid [51]. There are many enzymes that

phosphorylate proteins (kinases), and each recognises a different consensus sequence for addition of the modification.

The classical method for identifying sites of phosphorylation is through the use of radioactivity. ^{32}P is added to proteins either by adding radioactive phosphate to the cell culture or by incubating proteins/peptides with ^{32}P and a specific kinase for *in vitro* labelling. Edman sequencing of phosphorylated peptides and monitoring for release of radioactivity is employed to identify sites of modification.

Phosphorylation can also be studied using mass spectrometry. The addition of a phosphate group increases the mass of a peptide by 80 Da. Thus, search engines analysing peptide mass fingerprints can allow for possible phosphorylated peptides in their searches.

However, phosphatidic links are relatively labile, and can be cleaved in the mass spectrometer. When analysing serine or threonine phosphorylated peptides by MALDI-TOF-MS, two new peaks are often seen in the reflectron spectrum; a peak 98 Da smaller than the phosphopeptide formed by loss of the H_3PO_4 during ionisation, which can be used for identifying phosphorylation sites [25], and a metastable peak slightly higher in mass than this dephosphorylated peak, formed by loss of the phosphoric acid in the first drift region of the TOF. The lability of the phosphatidic bond can be exploited for detecting phosphopeptides. In negative ion ESI mass spectrometry, the loss of the phosphate group forms a unique ion at m/z 79. Precursor ion scanning for peptides which when fragmented form a negatively charged ion at m/z 79 can locate phosphopeptides in a mixture [52].

Phosphorylated peptides are ionised more readily in negative ion mode than positive ion mode [53]. Hence, another technique for locating phosphopeptides is to acquire peptide mass fingerprints of protein digests in positive and negative ion modes, and look for peaks that have increased significantly in intensity or appeared in the negative ion spectrum. Another commonly used technique for phosphorylation analysis is to acquire a mass fingerprint, then treat the sample with a phosphatase before acquiring a second mass fingerprint. The two spectra are compared, looking for the disappearance of one peak and the appearance/increase in intensity of a new peak that is 80 Da smaller [54].

Affinity purification of phosphorylated peptides from phosphoprotein digests as a precursor to mass spectrometric analysis simplifies samples and makes identification of phosphorylation sites simpler. Immobilised metal ion affinity chromatography (IMAC) can be used to selectively enrich phosphopeptides [55]. Alternative affinity purification strategies either by adding a biotin tag to the phosphate group [56] or replacing the phosphate with a biotin-tagged moiety [57] have also been demonstrated.

1.3.3.2. Mass Spectrometry and Glycosylation

Extra-cellular and cell surface proteins are generally modified with a complex diversity of glycan side chains. There are three major types of glycans: (1) N-linked glycans attached to asparagine residues in a N X S/T motif where X can be any residue except a proline; (2) O-linked glycans linked through serines and threonines at sites containing no consensus sequence for linkage, and (3) glycosylphosphatidylinositol(GPI) anchored proteins, modified at their C-termini with a combination of lipid and carbohydrate groups. N-linked glycans are generally the most complicated, containing a core of two N-acetylglucosamine (GlcNAc) residues and three mannose residues, to which branched sugar chains may be attached. For a given site on a protein there can be a complex mixture of different glycans (glycoforms) attached [58]. O-linked glycosylation is much more varied and can be anything from a monosaccharide to a complex polysaccharide chain.

Mass spectrometry is a major tool for the identification and characterisation of protein glycosylation [59, 60]. When a glycopeptide is fragmented by CID, the majority of the fragmentation is of the glycan rather than the peptide. Hence, sugar moieties are often removed enzymatically using PNGaseF for N-linked glycans and a β-elimination reaction employing a strong base for O-linked sugars. The released glycan is then analysed by mass spectrometry, often in combination with enzymatic digestion of the glycan to determine residue linkage. Unfortunately, the strong basic conditions generally used for β-elimination of O-linked sugars, cause degradation of the attached peptide, preventing determination of the site of modification. Classically, sites of glycosylation are determined using Edman degradation analysis by the presence of a gap in the sequence at the modified residue.

Sites of N-linked glycosylation can be predicted due to the existence of a consensus sequence for modification. Sites can also be determined by mass spectrometry, as the N-glycosidic linkage is as stable as the peptide backbone, allowing facile observation of glycosylated fragment ions [61]. However, sites of O-linked glycosylation are much more difficult to determine, as the O-glycosidic bond is significantly more labile than the peptide backbone. Thus, fragmentation spectra are dominated by 'deglycosylated' fragment ions, and low intensity glycosylated fragment ions have proven difficult to observe. When a single hexose residue was attached, sites of modification and peptide sequence could be determined using high energy CID, but when sites contain more than one sugar residue fragmentation was dominated by carbohydrate cleavage [62]. Sites of modification of O-linked glycopeptides attached through N-acetylgalactosamine (GalNAc) residues have been determined using high energy CID [61], MALDI-PSD and low energy CID on a triple quadrupole instrument [63], though in each of these cases large amounts of starting material was required to observe the low intensity glycosylated fragment ions.

However, with the advent of quadrupole oa-TOF instruments, the Peter-Katalinic group have demonstrated that O-GalNAc-linked sites can be determined at high sensitivity[64], and using this technology a novel type of O-glycosylation has been identified [65].

The lability of the glycosidic bond has been exploited for the detection of glycosylated peptides within a mixture. If an ESI-MS analysis of a digest is carried out using an elevated orifice potential, which causes in-source fragmentation, glycan-specific oxonium ions are formed at m/z 204 for HexNAc residues, m/z 163 for hexose, m/z 292 for sialic residues and m/z 366 for the disaccharide Hex-HexNAc. Using parent ion scanning for these ions, glycosylated peptides that form these sugar-specific fragment ions can be identified from an LC-MS analysis [66].

As well as glycosylation of extra-cellular proteins, nuclear and cytoplasmic proteins can also be glycosylated [67, 68]. By far the most studied nuclear glycosylation is GlcNAcylation [69].

1.4. What is GlcNAcylation?

GlcNAcylation is the addition of a single GlcNAc sugar residue O-linked to serine and threonine residues to form the structures in Figure 1.5.

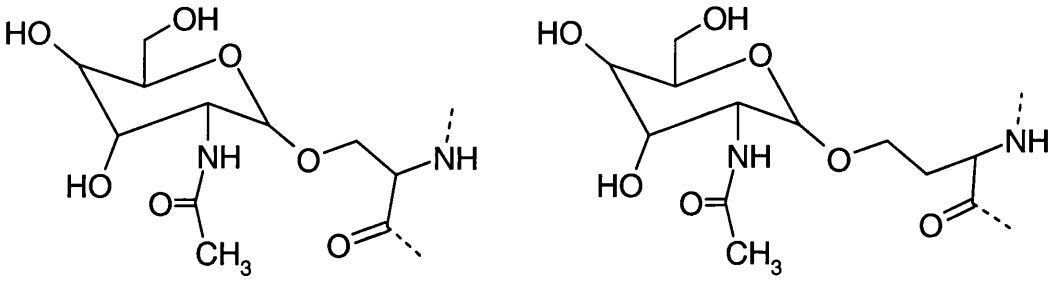


Figure 1.5: Structures of GlcNAc-modified serine and threonine residues.

The first evidence of this modification was from the work of Torres and Hart [67]. Whilst labelling terminal GlcNAc residues of the extra-cellular matrix of murine lymphocytes with [^{3}H]galactose using a galactosyltransferase, they discovered that the majority of the labelling was on proteins internalised in the cell. Subsequent work identified that GlcNAc-modified proteins are nuclear and cytoplasmically localised [70]. The modification has been detected in all eukaryotes studied, including plants, animals and fungi as well as proteins from viruses that infect eukaryotes.

Modified proteins are involved in many different aspects of cellular physiology from structural proteins to proteins involved in signal transduction. Table 1.1 lists proteins identified to be O-GlcNAc-modified. Due to the difficulty in identifying sites of modification, only a subset of the known GlcNAc-modified proteins have had sites of modification identified (Table 1.2). Although there is no consensus sequence for these sites, they have general features in common. Many sites have a proline residue in the two or three residues N-terminal to the site of modification, and most sites are surrounded by other serines and threonines. Using this information, if a protein is known to be modified, software has been developed to predict likely sites [71].

Table 1.1: Identified O-GlcNAc-modified proteins.

Nuclear Proteins	**Cytosolic Proteins**	**Other Proteins**
Estrogen receptors[72, 73]	α-Crystallin[74, 75]	β-Amyloid precursor protein[76]
c-myc[77]	Cytokeratins 8, 13, 18[78, 79]	Adenovirus fiber[80]
Nuclear pore proteins[81]	Neurofilaments H, M, L[82, 83]	Malarial proteins[84]
p53 tumour suppressor[85]	O-GlcNAc transferase[86]	P67 translational regulatory protein[87]
RNA Polymerase II[88, 89]	Synapsin[90]	Schistosome proteins[91]
Serum response factor[92]	MAP proteins[93]	Trypanosome proteins[94]
SV40 T antigen[95]	Tau[96]	Cytomegalovirus basic phosphoprotein (BPP)[97]
v-erb-a oncoprotein[98]	Talin[99]	
Sp1 and other transcription factors[100]	Clathrin assembly protein AP3[101]	
Casein kinase II	Ankyrin G[102]	
Chromatin binding proteins[103]	Glycogen synthase kinase-3[104]	
HnRNP G[105]		

Table 1.2: Determined sites of O-GlcNAc modification of proteins.

Protein	Site of Modification	Reference
Keratin 18	ARPVSSAASV	[106]
	SRISVSRSTS	[106]
αA-crystallin	AIPVSREEK	[74, 107]
αB-crystallin	EKPAVTAAPK	[75]
Serum Response	LAPVSASV	[92]
Factor	NLPGTTSTIQT	[92]
	SPSAVSSAD	[92]
	TQTSSSGT	[92]
Cytomegalovirus	SVPVSGSA	[97]
BPP	YPPSSTAK	[97]
Neurofilament-L	YVETPRVHISSV	[82]
	SGYSTAR	[83]
	SAPVSSSLSV	[83]
Neurofilament-M	GSPSSGFR	[83]
	GSPSTVSSS	[82]
	QPSVTISSK	[82]
	VPTETRSS	[83]
Neurofilament-H	PKSPATVK	[83]
	ARTSVSSVSAS	[83]
RNA Polymerase II	SPSSP	[88]
	TPTSP	[88]
	SPTSPT	[88]
Talin	VDPACTQ	[99]
	ANQLTND	[99]
c-myc	LLPTPPLS	[77]
Nuclear pore protein	PADTSDP	[108]
p62		
Erythrocyte band 4.1	TITSETPSS	[109]
Estrogen receptor α	TKASGMA	[110]
	SKPTVFN	[110]
	LATTSSTS	[73]
Estrogen receptor β	SVPSSTGNL	[111]
SV40 large T antigen	SEEMPSSDDEA	[95]
Sp1	QGVSLGQTSS	[112]

1.4.1. The Enzymes Involved in GlcNAcylation

GlcNAcylation is a transient modification, which has a turnover rate significantly higher than the protein itself [70]. The enzyme responsible for the addition of GlcNAc residues to proteins, a uridine diphospho-N-acetlyglucosamine:polypeptide β-N-acetlyglucosaminyltransferase (O-GlcNAc transferase) has been identified, purified [113], and subsequently cloned [86, 114]. The protein purified from rat liver cytosol is a heterotrimer of two 110kDa subunits and a 78kDa subunit [113], although in other tissues the 78kDa subunit appears not to be expressed, and the enzyme is a homotrimer [115]. The 78kDa subunit is closely related to the 110kDa subunit and is probably a cleavage product of the larger subunit. Monomeric 110kDa subunit is enzymatically active, but the homotrimer has a different binding affinity for the donor UDP-GlcNAc, suggesting the subunit composition may regulate the enzyme's activity [115]. Unlike the addition of phosphate, where there are many different kinases each recognising a different consensus sequence, there appears to be only one O-GlcNAc transferase. Thus, the regulation of this enzyme must be complex.

The C-terminal half of the protein, which is the catalytic part, shares no homology to any other glycosyltransferase [114]. The N-terminal half contains multiple tetratricopeptide repeats (TPRs), with 11 repeats in the human form of the protein. TPRs are degenerate 34 amino acid sequences present in proteins which are involved in a wide variety of processes [116]. Repeats of this region form super-helical structures, the groove of which is thought to bind the alpha helices of other proteins, thus facilitating protein-protein interactions [117]. Different proteins binding to this region are believed to influence the substrate specificity of the enzyme [86]. The transferase protein is tyrosine phosphorylated upon one of these repeat regions [115] and also GlcNAcylates itself [114], so is likely to be regulated in a number of different ways (Figure 1.6).

The O-GlcNAc transferase gene has been mapped to the X chromosome, and knockout studies have determined this protein to be essential for viability[118], demonstrating O-GlcNAcylation to be a fundamental process.

A cytosolic and nuclear localised β-N-acetylglucosaminidase (O-GlcNAcase) has been identified, purified[119] and cloned[120]. It is a ubiquitously expressed protein, with the highest levels of expression in the pancreas, placenta and brain. It is inhibited by

GlcNAc but not GalNAc, unlike other hexosaminidases. Several inhibitors of this protein have been identified including O-(2-acetamido-2-deoxy-o-glucopyranosylidene)-amino-N-phenylcarabamate (PUGNAc)[121] and streptozotocin[122]. Overnight incubation of cells with PUGNAc increases O-GlcNAc levels on selected proteins[121].

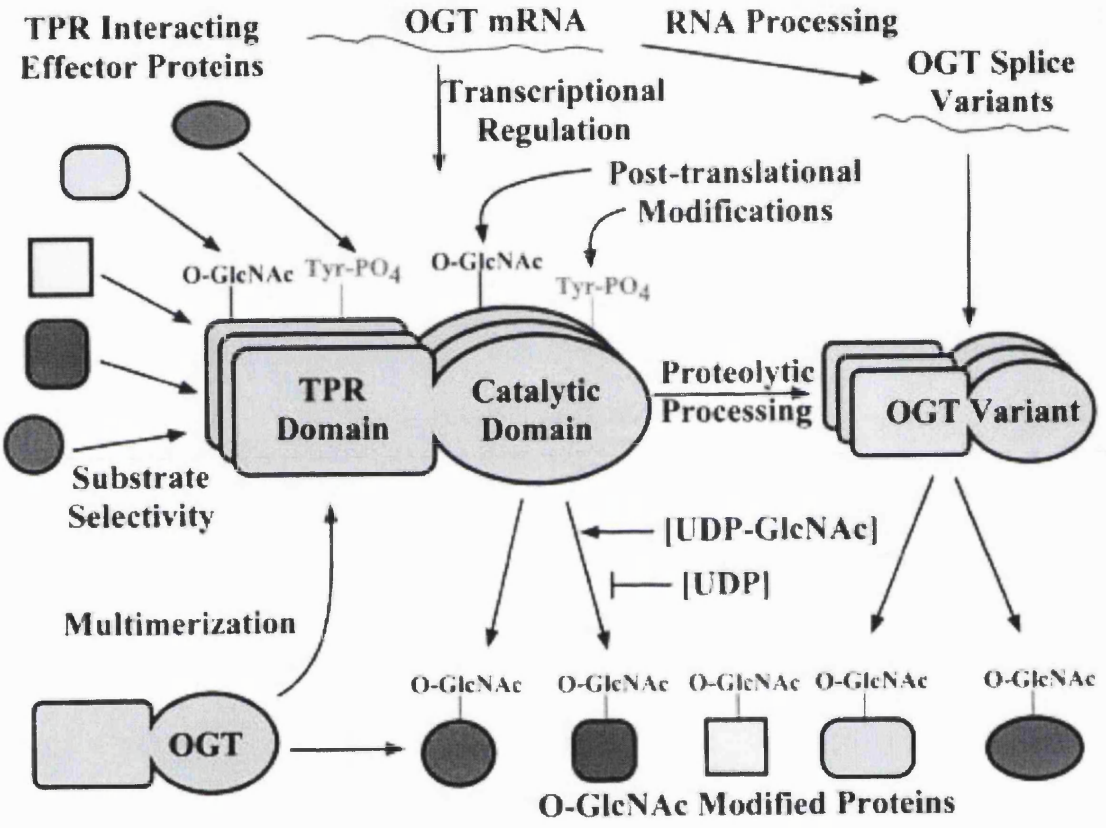


Figure 1.6: Possible modes of regulation of the O-GlcNAc transferase protein.
The enzyme is usually a homotrimer of three 110 kDa subunits. However, in some tissues such as the liver it contains a 78 kDa subunit, which is believed to be a cleavage product of the larger subunit. Individual subunits also show activity on their own. The protein is tyrosine phosphorylated and O-GlcNAc-modified. The presence of a TPR domain suggests regulation by protein-protein interactions. Its shows increased activity when levels of UDP-GlcNAc are high. This figure is reproduced from Comer et al. *[123]*

1.4.2. What is the Function of the Modification?

A plant protein SPINDLY (SPY) is highly homologous to O-GlcNAc transferase in both its N-terminal tetratricopeptide repeat domain and C-terminal catalytic domain [124]. SPY is a negative regulator of many of the effects of gibberellins [125], which are hormones in plants that trigger signal transduction pathways involved in growth and development. As SPY is known to regulate signal transduction in plants, and not just through gibberellin [126], it is probable that O-GlcNAcylation performs a similar function in animals.

All identified GlcNAc-modified proteins are potential phosphoproteins, and for many proteins the two modifications appear to be mutually exclusive [79, 88, 89]. Indeed, the site of O-GlcNAcylation in some cases is the same as a site of phosphorylation [72, 77, 95]. Hence, the two modifications compete for a site and the protein function will differ depending whether this residue exists in an unmodified, phosphorylated or GlcNAc-modified state. This immediately suggests that a function of GlcNAcylation may be to regulate phosphorylation of proteins. In support of this, altering cellular levels of phosphorylation causes a change in levels of O-GlcNAc modification [127, 128].

A number of diseases are associated with protein hyperphosphorylation. In the neurons of patients suffering from Alzheimer's disease the microtubule-associated protein tau is hyperphosphorylated. This protein is involved in the formation of abnormal filaments that cause neuronal cell death. Tau is also a target for O-GlcNAc modification, yet GlcNAc modification cannot be detected in hyperphosphorylated tau filaments [129]. Another protein involved in Alzheimer's is the β-amyloid protein, which aggregates to form insoluble plaques. The β-amyloid precursor protein is O-GlcNAc-modified, and to an altered extent in Alzheimer's patients [76]. Hence, a loss of GlcNAcylation has been suggested as a contributory factor in the disease's progression.

The largest subunit of RNA polymerase II contains an extended C-terminal domain consisting of over 50 repeats of a consensus sequence YSPTSPS. This region is a target for phosphorylation by a number of different kinases. Phosphorylation by the transcription factor TF IIH kinase is important for the transition from transcription initiation to elongation [130]. Phosphorylation is believed to enhance promoter clearance by disrupting interactions between proteins involved in the pre-initiation complex,

allowing the RNA polymerase II to travel along the RNA, although exactly how it does this has not been determined. Other phosphorylation events by the elongation factor pTEFb are believed to stabilise the elongation complex[131]. Phosphorylation also affects interactions between RNA polymerase II and mRNA splicing factors, mRNA capping enzymes and polyadenylation factors [132]. Conversely, phosphorylation of RNA polymerase II prevents formation of the pre-initiation complex [133].

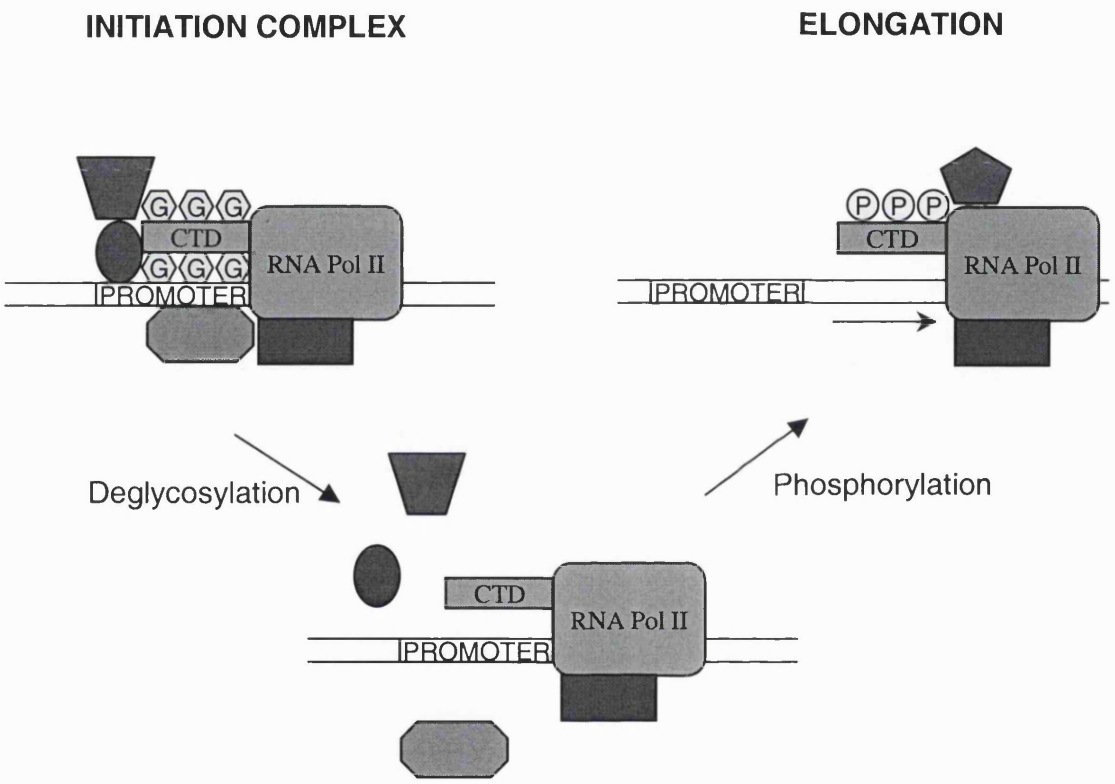


Figure 1.7: Proposed model of the different post-translational states of the C-terminal domain of the largest subunit of RNA polymerase II.

O-GlcNAc modification is predicted to be necessary for formation of the pre-initiation complex, probably through the binding of transcription factors through its TPR domain. For the transition from initiation to elongation, RNA polymerase II must be first deglycosylated, which frees the binding to the promoter region by releasing transcription factors. The protein is then phosphorylated, which allows the binding of elongation factors that stabilise the elongation process.

The C-terminal domain of the largest subunit of RNA polymerase II is O-GlcNAc-modified in the pre-initiation complex [88]. Although the individual residues modified by phosphorylation and GlcNAcylation may not be the same, the two modifications are mutually exclusive at the protein[88], enzyme and peptide level [89]. Hence, it is postulated that the GlcNAc-modified form of the protein is a storage form and contributes to protein-protein interactions involved in the formation of the pre-initiation complex. Once this is formed, the protein must be deglycosylated and then phosphorylated before elongation can take place[88] (Figure 1.7). This extra level of modulation prevents premature gene transcription before the pre-initiation complex is properly attached to the DNA.

c-Myc is a nuclear protein that dimerises with Max and regulates gene transcription of proteins involved in cell differentiation [134] and apoptosis [135]. It can be phosphorylated on threonine 58 and serine 62 [136]. Threonine 58 is also a site of O-GlcNAc modification [77]. This residue is a known mutational hot spot in Burkitt lymphomas and is associated with enhanced transforming activity and tumour formation [137, 138]. Thus, this site is key to the regulation of c-myc's activity. It remains to be determined which of the effects of mutation of this site are related to the lack of phosphorylation or lack of glycosylation, but as the levels of c-myc remain relatively constant throughout the cell cycle [139], regulation of this protein must be largely through its post-translational state.

The transcription factor Sp1 controls the expression of TATA-less genes, which encode housekeeping proteins. Phosphorylation of Sp1 alters its DNA binding activity [140]. It is also heavily O-GlcNAc-modified, with an average of eight modifications per molecule [100]. GlcNAc modification has multiple effects on the protein. Over-expression of O-GlcNAc transferase caused Sp1 to become O-GlcNAc-modified and inhibited Sp1 driven transcription [141]. This inhibition could be related to the demonstrated inhibition of protein-protein interactions of the glycosylated protein with itself or the TATA binding protein associated factor TAF110 [112]. An absence of GlcNAc modification targets the protein for proteasome-mediated degradation [142]. The signal for degradation is carried in the N-terminal 54 amino acids of the protein, and this region, when fused onto other proteins, targets them for destruction by the

proteasome [143]. The effects of O-GlcNAcylation on Sp1 are consistent with the theory that O-GlcNAc modification is a signal of the cell's nutritional state. In a starved state Sp1 becomes deglycosylated and destroyed, preventing protein transcription and thus conserving nutrients for energy utilisation.

SV40 Large T antigen is a 90kDa protein that is required for the infection of cells by the SV40 virus. It allows viral protein expression and DNA replication, and can initiate and maintain the transformation of a host cell through binding to tumour suppressor proteins p53 and the retinoblastoma protein [144]. A number of phosphorylation sites have been identified on this protein, and two of these sites, serines 111 and 112, also constitute a site of O-GlcNAc modification [95]. Hence, the two modifications compete for the same site, implying one modification regulates the other.

The tumor suppressor protein p53 is involved in DNA maintenance and can promote and repress the transcription of different proteins. A constitutively active form of p53 is O-GlcNAc-modified in the basic C-terminal region of the protein [85]. The GlcNAc modification allows the protein to bind with high affinity to DNA, whereas the deglyocosylated version cannot bind to DNA with the same affinity. Hence, the GlcNAc masks a region that suppresses DNA binding and subsequent transcriptional activation.

Protein translation is also regulated by O-GlcNAc modification. eIF-2 is a key regulator of protein synthesis. When phosphorylated by eIF-2 kinase it inhibits protein translation. An O-GlcNAc-modified 67 kDa protein (p67) binds to eIF-2 [87]. This prevents eIF-2 from being phosphorylated by its kinase, thus precluding eIF-2 from inhibiting protein translation. However, under starvation conditions p67 becomes deglycosylated, leading to it being degraded [145]. This frees eIF-2 kinase to phosphorylate eIF-2, and protein translation is inhibited.

Levels of O-GlcNAc modification are closely linked to levels of free glucose in the cell. In the starved state, glucose is utilised as an energy source. However, when glucose levels are high, it is converted into glucosamine through the hexosamine pathway. The pivotal enzyme in this switch is the enzyme glutamine:fructose-6-phosphate amidotransferase (GFAT). This enzyme converts fructose-6-phosphate into glucosamine for glycoprotein production. Under starvation conditions, adenylate cyclase phosphorylates and inactivates the GFAT enzyme, preventing glucose being converted

into glucosamine [146]. Levels of O-GlcNAc-modified proteins are reduced under starvation conditions [142] and upon protein kinase A or C activation [128]. Exposure of cells to elevated glucosamine increases O-GlcNAc modification[147]. In cells containing GFAT mutated at the phosphorylation site serine 205, levels of O-GlcNAc modification were increased by the addition of glucosamine, but not by the addition of glucose[146]. Thus, the regulation of this protein could be a key controlling factor in the levels of O-GlcNAc modification.

Hyperglycaemia, where blood glucose levels are too high, is a symptom of the disease diabetes mellitus. There is growing evidence to suggest that the utilisation of glucose into the hexosamine pathway through GFAT is a contributory factor to the disease's pathogenesis. Over-expression of GFAT in mice caused the development of insulin resistance [148]. This coincided with a reduction in protein levels of the GLUT4 glucose transporter. The N-acetylglucosamine homologue streptozotocin is toxic to pancreatic β-cells[147]. Streptozotocin is known to inhibit the GlcNAcase and also increases cellular O-GlcNAc levels by increasing production of the O-GlcNAc transferase [122]. This is also correlated with increased flux of glucose into the hexosamine pathway. Under hyperglycaemic conditions O-GlcNAc transferase shows increased protein expression and activity [149]. Abnormal O-GlcNAcylation levels are hence thought to be a contributory factor in many diseases associated with diabetes caused by abnormal proliferation of smooth muscle cells in arterial walls leading to cataracts, renal failure and atherosclerosis [150].

The nuclear pore complex contains the highest concentration of O-GlcNAc-modified proteins in the cell [81]. An antibody recognising GlcNAc as part of its epitope was able to identify eight nucleoporins as being O-GlcNAc-modified [81]. Wheat germ agglutinin (WGA), a lectin that specifically binds to terminal GlcNAc residues, is able to block nuclear transport of proteins by binding to the nuclear pores, suggesting a role for O-GlcNAc in nuclear transport [151]. However, capping of GlcNAc moieties with galactose did not inhibit nuclear transport suggesting O-GlcNAc is not critical for this process [152].

O-GlcNAc-modified peptides are highly immunogenic, and are naturally presented by MHC class I molecules [153]. Antibodies raised against the nuclear pore protein p62

target a GlcNAc-modified epitope as their substrate, and these antibodies show cross-reactivity against other GlcNAc-modified proteins [81, 154]. Recently, a mouse monoclonal antibody has been purified that is specific to O-GlcNAc and shows no cross-reactivity to peptides or other sugar moieties [155]. It appears to be specific to O-GlcNAc-modified serine or threonine and has more widespread reactivity than the p62 monoclonal antibody. Importantly, it is also suitable for immunoprecipitation of GlcNAc-modified proteins, so should be a valuable tool in glycoproteomic research.

1.5. Determination of Sites of O-GlcNAc Modification

The traditional method for determination of sites of GlcNAc modification is through the use of radioactivity [156]. GlcNAc-modified proteins are identified by radiolabelling GlcNAc residues with [^{3}H]galactose *in vitro* using a galactosyltransferase. Modified proteins are enzymatically digested and peptides are separated by multiple rounds of HPLC. Radiolabelled peptides are located by screening fractions for radioactivity. When a fraction containing only the radiolabelled peptide is isolated the peptide can be identified by Edman sequencing. Alternatively, the peptide can be identified using mass spectrometry to measure its molecular mass, from which the peptide can be identified from a theoretical enzymatic digest of the protein. The site of modification is determined by monitoring cycles of Edman degradation for the release of the radioactively labelled amino acid. Sites can also be identified or confirmed using site-directed mutagenesis of predicted modified residues in the protein, then analysis to determine whether the mutation prevents radiolabelling of the protein with [^{3}H]galactose.

Mass spectrometry has also been used to indirectly identify sites of GlcNAc modification. Strong base catalysed β-elimination of GlcNAc residues O-linked to synthetic peptides followed by fragmentation analysis was able to identify modified residues [157, 158]. When a sugar residue is β-eliminated from a glycopeptide a hydroxyl moiety is also cleaved from the side chain of the modified amino acid, converting a glycoserine into a dehydroalanine, and a glycothreonine into a dehydrobutyric acid. This causes the formation of a residue 18 Da smaller in mass than an unmodified residue, which can be exploited in the determination of GlcNAcylation sites. To this effect, Greis *et al.* [157] investigated the ability to perform β-elimination of

GlcNAc, GalNAc and LacNAc residues from synthetic peptides using sodium hydroxide. All three sugar moieties could be eliminated, with O-GlcNAc being the most labile, and thus needing the shortest incubation time. They were able to identify the site of modification from the β-eliminated product using ESI-CID-MS-MS on a triple quadrupole mass spectrometer.

The presence of large amounts of sodium in the sample meant that a reverse phase clean-up step was necessary after the β-elimination reaction. The need for this extra step decreased the sensitivity of the protocol. Thus, a more sensitive protocol was demonstrated using ammonium hydroxide as the base [158]. This is volatile buffer, so can be easily removed by vacuum centrifugation. Using this protocol a site of modification could be determined starting with only one picomole of pure synthetic peptide. However, it was found that the rate of β-elimination was dependent on the peptide sequence and also the residue of attachment. Glycothreonine residues required much longer incubations to eliminate the sugar moiety than glycoserines. Hence, using this protocol required individual method optimisation for each new peptide that was analysed. Thus, although a site of modification could be determined from one pmole of peptide, more sample would be required for initial method optimisation.

β-Elimination reactions for identifying sites of post-translational modification is currently undergoing extensive research in a number of laboratories, primarily for the purpose of identifying phosphorylation sites [56, 57, 159]. An O-phosphate group is similarly liberated in strong basic conditions to O-glycans. However, the phosphate moiety is much less labile meaning stronger basic conditions are required to release the phosphate group. Hence, a two tier approach of releasing O-glycans in mild basic conditions followed by stronger conditions to free phosphate groups could be proposed. It has also been suggested that prior phosphatase treatment of samples intended for glycosylation analysis would be beneficial, to prevent misleading data [159].

1.6. Aims of this Thesis

The major stumbling block in the understanding of O-GlcNAcylation is the difficulty of detection of the modification and subsequent determination of sites of modification. Mass spectrometry is potentially a much more sensitive method of detection than the

radioactivity/Edman sequencing approach currently used, and should be able to identify sites of modification at lower levels than have previously been achieved. Initial work characterising this modification by mass spectrometry has determined that the O-GlcNAc moiety is highly labile, and people have therefore sought to derivatise the modification site using a base catalysed β-elimination reaction before fragmentation analysis. However, this technique has only been applied to synthetic peptides, and has not determined any novel sites of modification.

The aim of this thesis is to demonstrate that mass spectrometry can be utilised to determine sites of O-GlcNAc modification in proteins. It will be demonstrated that sites of modification can be determined without the need to perform any prior derivatisation of the peptide, and the sensitivity for determining sites of modification on synthetic peptides is significantly higher using mass spectrometry than by any previous protocol (Chapter 3). In Chapter 4, a previously determined site of modification of αA-crystallin is confirmed by mass spectrometry, and the sensitivity of the technique for determining sites in proteins is investigated. Chapter 5 is an in-depth analysis of the post-translational modification state of serum response factor (SRF). As well as confirming previous sites of modification, previously undetected sites of GlcNAcylation and phosphorylation are determined. Finally, initial investigations into identifying novel GlcNAc-modified proteins by a proteomic approach are made (Chapter 6).

2. MATERIALS AND METHODS

All chemicals were purchased from Sigma-Aldrich (Poole, Dorset, UK), unless otherwise stated.

2.1. Experiments using Synthetic Peptides

2.1.1. Synthetic Peptides

Three sets of synthetic peptides were supplied by Professor Gerry Hart (John Hopkins School of Medicine, Baltimore, USA). The sequence of each set was based on previously identified O-GlcNAcylated regions of proteins, and each set contained unmodified, phosphorylated and O-GlcNAcylated peptides.

The first set of samples had the sequence KKFELLPTPPLSPSRR, which is a sequence found in the N-terminal transcriptional activation domain of c-myc. Threonine 58 has been shown to be a site of O-GlcNAcylation [77], and threonine 58 and serine 62 are both phosphorylation sites [160]. In these samples, the threonine was unmodified, phosphorylated or glycosylated. These three peptides are referred to in the text as N-Myc (unmodified), P-Myc (phosphorylated) and G-Myc (glycosylated).

The second set of samples had the sequence YSPTSPSK, a consensus sequence that is repeated over fifty times at the C-terminus of the largest subunit of RNA polymerase II. This region is multiply phosphorylated [133] and O-GlcNAcylated [88]. Modification of these peptides was on the middle serine. These peptides are referred to as N-CTD, P-CTD and G-CTD.

The final set had the sequence PGGSTPVSSANMM, which is a sequence found near the C-terminus of casein kinase II. The threonine residue in this sequence is phosphorylated in cells arrested at the G2 to M transition of the cell cycle [161]. The phosphorylated peptide was modified on the threonine, whilst the O-GlcNAcylated peptide was modified on serine 8. These peptides are referred to as N-CKII, P-CKII and G-CKII.

2.1.2. MALDI

MALDI mass spectra were acquired using either α-cyano-4-hydroxycinnamic acid (CHCA) in methanol (Hewlett-Packard, Böblingen, Germany), or using a saturated solution of 2,5-dihydroxybenzoic acid (DHB) in water. MALDI-MS and MALDI-PSD spectra were acquired on the Voyager Elite XL (Applied Biosystems, Foster City, CA, USA). MALDI-CID-MS-MS and MALDI precursor ion scanning experiments were acquired on a QSTAR, fitted with a prototype MALDI source (a collaboration between Applied Biosystems and MDS Sciex, Ontario, Canada).

2.1.3. ESI

Samples for positive ion ESI-MS, CID-MS-MS and nanospray-MS-MS were dissolved in 50% methanol/0.1% formic acid. Samples for negative ion ESI-MS and CID-MS-MS were dissolved in 12.5% ammonium acetate/40% methanol.

Precursor ion scanning experiments were performed on a QSTAR. All other ESI-MS spectra were acquired on a Q-TOF (Micromass, Manchester, UK).

2.1.4. On-line LC-ESI-MS

HPLC was carried out using the Ultimate/Famos/Switchos suite of instruments (LC Packings, Amsterdam, NL). Samples were loaded onto a guard column (300 μm ID x 5 cm C18 PepMap) in the injection loop, and washed using 0.1% formic acid at 40 μl/min for 2 minutes using the switchos pump. Peptides were then separated on an analytical column (75 μm ID x 15 cm C18 PepMap) at a flow rate of 200 nl/min employing a gradient from 5% to 40% buffer B (80% acetonitrile/0.1% formic acid) over a period of 32 minutes.

Nano-ESI-MS was performed on the Q-TOF. Automatic function switching between survey MS and MS-MS modes was used. When the doubly charged peak corresponding to the G-CKII peptide was detected above 4 counts/sec in the mass spectrum it was automatically selected for tandem MS analysis, and was fragmented until the peak fell below this threshold again. This peptide was fragmented using a collision energy of 20 V.

2.1.5. Galactosyltransferase labelling of Synthetic Peptides

Synthetic peptides were galactosylated using a protocol adapted from a published procedure [162]. The labelling buffer contained 3.75 µl 25 mM ammonium bicarbonate, 0.25 µl galactose buffer (0.1 mM galactose, 0.1 mM Hepes-NaOH, 0.15 mM NaCl, 50 µM $MgCl_2$), 1µl autogalactosylated galactosyltransferase (1 mU) and 0.5 µl 160 µM UDP-galactose in 25 mM ammonium bicarbonate containing 0.5 mM 5'-AMP. Labelling was carried out at 37°C for 6 h to maximise levels of labelling.

2.1.6. β-Elimination of O-GlcNAc

β-Elimination of O-GlcNAc was performed using a previously published protocol [158]. G-CTD peptide was incubated in 25% NH_4OH for 16 h at 45°C. The sample was then vacuum centrifuged to dryness to remove the volatile base, before re-suspension in water.

2.1.7. Poros Purification of Samples for Nanospray-MS

A BSA digest spiked with G-CKII was cleaned-up using Poros R2 20µm OD media (Applied Biosystems) using a published protocol [163]. Briefly, a slurry of R2 media in methanol/water is loaded into a nanoES purification capillary (Protana, Odense, Denmark). The end is broken and the media is equilibrated using 0.1% formic acid. Sample is loaded in 0.1% formic acid and further washed in this buffer, before eluting using 50% methanol/0.1% formic acid directly into a nanospray needle (Micromass).

2.1.8. Precursor Ion Scanning

Precursor ion scanning was performed for the production of fragments at m/z 204.1±0.05 Th. The mass range of m/z 500 to m/z 1000 was scanned over 5 seconds, with a step size of 0.2 Th, employing a collision offset of 20 V with nitrogen as the collision gas.

2.2. Alpha Crystallin

Bovine α-Crystallin was purchased from Sigma (Poole, UK). HPLC grade water and acetonitrile were purchased from Rathburn (Walkerburn, Scotland), formic acid and

trifluoroacetic acid (TFA) from Romil (Cambridge, UK). The HPLC columns were obtained from LC Packings (Amsterdam, Netherlands). Sequencing grade modified trypsin was from Promega (Southampton, UK)

2.2.1. 1D SDS-PAGE

400 ng of bovine α-crystallin was loaded onto a 12% polyacrylamide mini-gel (8 cm x 10 cm x 0.5 mm) and run on a Hoeffer mini-gel system (Amersham-Pharmacia Biotech (APBiotech), Amersham, UK). The gel was then stained using Coomassie for 1 h and destained for 3 h, before bands were excised.

2.2.2. Solution Digestion

200 ng of α-crystallin was digested using six nanograms of modified trypsin in 25 mM NH_4HCO_3 and incubated at 37°C for two hours.

2.2.3. In-gel Digestion

Excised gel bands were further destained in 25 mM NH_4HCO_3/50% acetonitrile, vacuum centrifuged to dryness, then allowed to swell in 5 µl 25 mM NH_4HCO_3 containing 4 ng/µl modified trypsin for 5 minutes. Gel pieces were then overlayed with a further 25 µl 25 mM NH_4HCO_3 and digested overnight at 37°C.

Peptides were extracted using two changes of 20 µl 50% acetonitrile/5% TFA, vacuum centrifuged to dryness and re-suspended in 10 µl water prior to analysis.

2.2.4. On-Line LC-ESI-MS

HPLC was carried out as Section 2.1.4.

Nano-ESI-MS was performed on the Q-TOF. Automatic function switching between survey MS and MS-MS modes was used. Using this method, when a multiply-charged peak above 4 counts/sec was detected in the mass spectrum it was automatically selected for tandem MS analysis. Up to two precursor ions could be selected for MS-MS at any time. In the initial analysis of the sample, four MS-MS spectra at two different collision offsets were acquired on each selected parent ion, then it switched back to MS mode to look for other ions. For ions of m/z 400 - 900, collision offsets of 28 V and 32 V were

used, whereas for peptides of m/z 900 - 2000, collision offsets of 30 V and 35 V were used. Once GlcNAc-modified peptides were located, a second run was carried out, performing MS-MS only on modified peptides. This time MS-MS was performed for as long as the peak stayed above 4 counts/sec (typically 40 - 50 seconds). The triply charged GlcNAc-modified peptide from αA-crystallin was fragmented using 22 V and 25 V collision offsets.

2.2.5. Precursor Ion Scanning

Precursor ion scanning was performed essentially as Section 2.1.8, except precursor ion scans were either for the production of m/z 204.086±0.05 Da or for the production of m/z 204.086±0.01 Da.

2.3. Serum Response Factor

Human serum response factor (SRF) was a kind gift from Dr Rob Nicholas (Imperial Cancer Research Fund, UK). The protein was over-expressed in baculovirus, then purified [164]. This was the same sample that was used to identify O-GlcNAc modification sites using a combination of mass spectrometry, radioactivity and Edman sequencing [92]. The sample was supplied in 20 mM hepes pH 7.9, 300 mM KCl, 0.2 mM EDTA, 0.2 mM EGTA, 0.1% NP-40, 10% glycerol, 1mM DTT. The presence of detergents and salts made the sample unsuitable for direct analysis by mass spectrometry. Hence, protein was purified by 1D SDS-PAGE on a 10% polyacrylamide gel, before in-gel digestion using the relevant enzyme. For both tryptic and chymotryptic digests 20 ng enzyme was used, and digests were performed using the protocol in Section 2.4.3.

Sequencing-grade modified trypsin (Promega, Southampton, UK), sequencing-grade chymotrypsin (Roche, Lewes, UK) and proline-specific endopeptidase (pro-C)(AMS Biotechnology, Abingdon, UK) were used.

2.3.1. MALDI Mass Fingerprinting

0.5 µl of sample was mixed on target with 0.5 µl of DHB (saturated solution in water). Sample was allowed to dry at room temperature before introduction into the mass spectrometer. Spectra were acquired on the Reflex III (Bruker Daltonics, Bremen,

Germany), surveying a range of m/z 600 – 7000. Spectra were internally calibrated using tryptic autolysis peaks at m/z 842.510 and 2212.105 if observed, or externally calibrated from a nearby spot using a peptide calibration mixture (Calibration mixture 2 from Sequazyme kit, Applied Biosystems).

Database searching was carried out using an in-house version of MS-FIT[44]. Searches allowed for 100 ppm mass error, unmodified or acrylamide modified cysteines, oxidised methionines and conversion of N-terminal glutamine to pyro-glutamate. For tryptic digests up to three missed cleavages were allowed. For chymotryptic digestions and combined tryptic/chymotryptic digests up to six missed cleavages were permitted.

2.3.2. On-line LC-ESI-MS

HPLC was carried out as in Section 2.1.4. Nano-ESI-MS was performed as in Section 2.4.4. For subsequent CID-MS-MS analysis of observed GlcNAc-modified peaks, using the automatically acquired CID-MS-MS spectra as a reference, collision offsets were adjusted to optimise for the observation of GlcNAc-modified fragment ions.

2.3.3. HPLC Fractionation of Chymotryptic Digest

HPLC was carried out using an ABI 140D pump and 785A detector (Applied Biosystems). Digest was loaded onto a guard column (300 μm ID x 5cm C18 PepMap) in the injection loop and washed using 20 μl 0.1% formic acid. Peptides were then separated on an analytical column (300 μm ID x 15cm C18) at a flow rate of 5 μl/min with a gradient of 5 to 40% Buffer B (80% acetonitrile) over a period of 32 minutes. Fractions were collected when peaks were observed on the UV chromatogram. Fractions were vacuum-centrifuged to dryness and re-suspended in 4 μl water for MALDI screening.

2.3.4. Pro-C Sub-digestion

5 μl (2 mU) of pro-C in 25 mM NH_4HCO_3 was added to the HPLC fraction containing the GlcNAc-modified peptide. Digestion was carried out at 37°C for 1 h.

2.3.5. Database Searching

Searches performed using MS-FIT and MS-TAG [44] were performed using an on-site version of Prospector and were searched against a database containing only the myc-tagged SRF protein. MASCOT searches were performed against the NCBI.nr database, a non-redundant database [165].

2.3.6. Extracted Ion Chromatograms

All extracted ion chromatograms were performed against the average mass of the peptide rather than the monoisotopic mass. This is to ensure that the ions from all isotopes of the peptide contribute to the extracted ion chromatogram.

2.4. Proteomics to Find Modified Proteins

2.4.1. Cell Line

The cell line utilised for all experiments was the EGFR over-expressing head and neck tumour carcinoma cell line HN5 [166].

2.4.2. Nuclear Isolation

All the nuclear isolation procedures were carried out on ice, and all centrifugation performed at 4°C.

2.4.2.1. Preparation of Crude Nuclei

The nuclear isolation protocol used was a previously published protocol [167]. Flasks of cells were scraped into phosphate buffered saline (PBS) and pelleted at 1200rpm (300xg) for 5 minutes in a Sorvall TechnoSpin R centrifuge (Hertfordshire, UK). The supernatant was discarded and cells were re-suspended in 2 ml buffer A (10 mM Hepes-NaOH pH 7.4, 10 mM NaCl, 1 mM EGTA, 1 mM $MgCl_2$) and left to swell for 10 minutes. 100 µl of this cell lysate solution was removed and used for assessment of nuclear clean-up process. Cells were homogenised by five passes in a dounce homogeniser (teflon head with 1 – 1.5 mm clearance). 2 ml of 0.7% Triton X-100, 3 mM $MgCl_2$, 250 mM sucrose was added to the cells. The sample was split into two, and each

half (2 ml) was layered onto a solution of buffer B (300 mM sucrose, 50 mM Hepes-NaOH pH 7.4, 50 mM NaCl, 5 mM $MgCl_2$, 1mM EGTA). The samples were then centrifuged at 1500 rpm (400xg) for five minutes to give crude nuclear pellets.

2.4.2.2. *Purification of Nuclei – Mellman*

This clean up step is used in the laboratory of Dr Ira Mellman (Ludwig Institute for Cancer Research, Yale University, CT, USA).

One of the crude nuclear pellets was re-suspended in 2 ml RSB (10 mM Tris-HCl pH 7.2, 10 mM NaCl, 3 mM $MgCl_2$), using a transfer pipette to efficiently break up the pellet. Nuclei were passed four times through a dounce homogeniser before being pelleted at 3,200 rpm in a Multifuge 3 S-R Centrifuge (Merck). This process was repeated before overlaying sample onto 0.88 M sucrose and pelleting at 3,200 rpm for five minutes to give a purified nuclear pellet.

2.4.2.3. *Purification of Nuclei – Gerner*

This clean up step is based on a published protocol for tissue culture cells [167]. Crude nuclei were re-suspended in 2 ml 0.2% sodium deoxycholate, 0.4% Tween-40, 2mM vanadyl ribonucleoside complex, 3 mM $MgCl_2$ in buffer A. Nuclei were passed four times through a dounce homogeniser and incubated for 10 minutes at 4°C. Nuclei were then pelleted through a cushion of 2 ml buffer B at 3,200 rpm for five minutes to give a purified nuclear pellet.

2.4.3. *Galactosyltransferase Labelling*

Radiolabelling of GlcNAc-modified proteins using [^{3}H]-galactose was carried out using a previously published protocol [162].

2.4.3.1. *Enzyme Autogalactosylation*

Firstly, the galactosyltransferase must be autoglactosylated to prevent subsequent radiolabelling of the enzyme. 10 U galactosyltransferase (Oxford Glycosciences, Abingdon, UK) was dissolved in 100 µl water. Half of this was then added to 59 µl of the autogalactosylation buffer (50 µl 100 mM Tris-HCl pH 7.2, 10 mM $MnCl_2$, pH 7.4, 5

µl 8 mM UDP-galactose, 2 µl aprotinin (10 mg/ml in water) and 2 µl 50 mM β-mercaptoethanol). Autogoalactosylation was carried out by incubation for 1 h at 37°C, before being chilled on ice. Enzyme was precipitated by piecemeal addition of 61mg ammonium sulphate followed by incubation overnight at 4°C. Precipitate was pelleted at 16,000xg for five minutes, supernatant was removed and the pellet was re-suspended in 50 µl 25 mM Hepes-NaOH pH 7.4, 5 mM $MnCl_2$, and then dialysed against this buffer for 4 h. The autogalactosylated enzyme buffer swelled during dialysis and when mixed with 50 µl glycerol gave approximately 150 µl enzyme at 35 U/ml.

2.4.3.2. In vitro *Galactosylation of Nuclear Proteins*

The nuclear pellet was re-suspended in 100 µl labelling buffer without label (75 µl 10mM Hepes-NaOH pH 7.3), 0.15 M NaCl, 0.3% NP-40, 20 µl autogalactosylated galactosyltransferase, 5 µl galactose buffer (0.1 M D-galactose, O.1M Hepes-NaOH pH 7.3, 0.15 M NaCl, 50 mM $MgCl_2$, 5% NP-40). Nuclei were then split into two tubes. To one tube 5 µl UDP-[^{3}H]-galactose (1µCi/µl), 25mM 5'-AMP was added ('Hot'). To the other tube 5 µl 18µM UDP-galactose was added ('Cold').

Samples were incubated at 37°C for 90 minutes before proteins were precipitated in 10% tricarboxylic acid (TCA) in acetone at –20°C for 3 h. Precipitate was pelleted and washed using acetone.

2.4.4. 2D PAGE

Hot and cold labelled nuclear pellets were re-suspended in 342 µl 'lysis buffer' (8 M urea (Merck, Leics, UK) , 2 M thiourea (Merck) , 4% CHAPS, 65 mM DTT (Merck)), 4 µl ampholine 3 – 10 (Amersham-Pharmacia Biotech (APBiotech) Amersham, UK), 4 µl pharmalyte 3 – 10 (APBiotech) and 2 µl bromophenol blue, and loaded onto pH 3 – 10 IPG strips (APBiotech). These were overlayed with mineral oil and left to rehydrate for 20 h.

Strips were run on a IPGphor (APBiotech) at 200 V for the first hour, stepped to 300 V for 1 h, then a gradient from 300 – 3500 V over 1.5 h. Finally they were run at 3500 V for 19 h.

Strips were then re-equilibrated in 6M urea, 2% SDS, 1% DTT, 50mM Tris-HCl pH 6.8 for 20 minutes.

Second dimensions were run on 9 – 16% gradient polyacrylamide gels. The light solution contained 120 ml water, 60 ml 1.5 M Tris-HCl, 60 ml acrylamide/piperazine diacrylamide (PDA) 650 µl dimethyl piperazine and 850 µl 10% ammonium persulphate. The heavy solution contained 60 ml water, 60 ml 1.5 M Tris-HCl, 120 ml acrylamide/PDA, 650 µl piperizine diacrylamide and 850 µl 10% ammonium persulphate. Gradient gels were cast using a Hoeffer Gradient Maker (APBiotech).

Second dimension gels were overlayed with warm agarose solution containing bromophenol blue and IPG strips were layered onto the second dimension gel and the agarose was allowed to set.

Second dimensions were run at 25 mA per gel for 0.3 h, then 40 mA per gel for 4.5 h.

2.4.5. Coomassie Staining

Gels were stained using Coomassie (1 mg/ml Coomassie Brilliant Blue in 45% methanol, 10% acetic acid) for 1 h, then destained in 40% ethanol, 10% acetic acid.

2.4.6. Fluorescent Signal Amplification of Radioactivity

The 2D gel of the radiolabelled nuclear fraction was soaked in EN^{3}HANCE fluorescent scintillant (PE Life Sciences, Cambridge, UK) for 1 h, with constant gentle shaking. Scintillant was precipitated in cold water for 30 minutes, and then the gel was dried onto 3M paper. The dried gel was exposed to film for one month.

2.4.7. Western Blotting

Nuclear fractions for immunoblotting were run on 7.5% acrylamide gels. Following electrophoresis, proteins were blotted onto PVDF using a wet blotter (Biorad). The transfer buffer was 192 mM glycine, 25 mM Tris-HCl, 3 mM SDS, 20% methanol. Proteins were transferred at 200 mA constant current for five hours.

Membrane was blocked overnight in 5% milk powder in TBST (50 mM Tris-NaOH pH 8.0, 150 mM NaCl, 0.05% Tween 20, 0.001% NaN_3).

Membranes were probed for 3 h using dilutions (in TBST) of the following antibodies: 1 in 3,000 mouse anti-E-cadherin (BD Transduction Labs, NJ, USA), 1 in 20,000 mouse anti-MEK2 (BD Transduction Labs), 1 in 5,000 goat anti-calnexin (Autogen Bioclear, Wiltshire, UK) or 1 in 5,000 goat anti-Lamin B (Autogen Bioclear). Membranes were washed three times using TBST, and then probed using the relevant secondary antibody, either anti-mouse-HRP (Biorad) or anti-goat-HRP (Dako, Cambridgeshire, UK) for 1 h. After a further 3 washes using TBST, bands were visualised by electrochemical luminescence (ECL) using SuperSignal West Pico (Pierce, Cheshire, UK).

2.4.8. Wheat Germ Agglutinin – Horseradish Peroxidase Probing

Probing for GlcNAc-modified proteins using wheat germ agglutinin conjugated to horseradish peroxidase (WGA-HRP) is a similar protocol to that used for western blotting (Section 2.4.7). Following electrophoresis, proteins were transferred to PVDF. Membranes were blocked using 4% BSA instead of milk proteins. Membranes were then washed 3 times in TBST before probing for 3 h using WGA-HRP diluted 1 in 10,000 in TBST. After a further three washes the image was visualised by ECL using Supersignal West Pico (Pierce).

On-blot glycosidic digestion was carried out using 5 U PNGaseF and 1 U Sialidase in 25 mM ammonium bicarbonate. After the membrane was blocked as above, it was incubated in glycosidic digestion buffer for 24 h. The membrane was then washed and probed using WGA-HRP as above in either TBST or TBST in the presence of 0.5 M GlcNAc.

2.4.9. WGA-Sepharose Affinity Chromatography

WGA affinity chromatography was carried out using a previously published protocol [168]. A WGA-sepharose affinity column was constructed using 1 ml of WGA-sepharose beads. This column was then equilibrated by three washes using 1 ml loading buffer (50 mM Tris-HCl pH 7.2, 0.1 mM $CaCl_2$, 0.5 M NaCl, 0.1% Triton X-100, 0.5 mM DTT).

A purified nuclear pellet (Section 2.4.2) was [^{3}H]-galactose labelled *in vitro* (Section 2.4.3.2), re-suspended in loading buffer, sheared using a 25 gauge needle (BD, Oxford, UK) and pelleted. The supernatant was loaded onto the WGA-sepharose column and the eluent was collected. The flowthrough was re-loaded and collected a further two times to ensure maximal protein binding. The eluent from the third loading was kept as 'flowthrough'.

The column was washed by two 1ml loadings of wash buffer (50 mM Tris-HCl pH 7.2, 0.1 mM $CaCl_2$, 1 M urea, 1% Triton X-100) to remove non-specifically bound proteins, and eluents were collected as 'wash 1' and 'wash 2'. Finally, specifically bound proteins were eluted using two 1 ml loadings of 0.5 M GlcNAc in loading buffer, and collected as 'elute 1' and 'elute 2'.

Proteins in each fraction were precipitated by overnight incubation in 10% TCA in acetone at –20°C. Precipitated proteins were pelleted, washed using acetone then loaded onto a 1D gel. Proteins were separated by SDS-PAGE, visualised using Coomassie and radiolabelled proteins detected as in Section 2.4.6.

The gel was Coomassie stained to visualize proteins, then soaked in EN^{3}HANCE, dried and exposed to film for a month to detect radiolabelled proteins (see Section 2.5.6 for more detailed procedure).

3. EXPERIMENTS USING SYNTHETIC PEPTIDES

A set of synthetic peptides was used as standards for the development of mass spectrometric techniques for analysing GlcNAc-modified proteins. Once methods were developed, they could then be applied to the analysis of peptides from protein digests.

3.1. Stability of GlcNAc Modification in the Mass Spectrometer

3.1.1. Positive Ion ESI-MS

The O-Glycosidic link between a GlcNAc residue and serines and threonines is highly labile. Indeed, this has previously been exploited for precursor ion scanning to locate glycopeptides [66, 169].

The fragility of this linkage is illustrated in Figure 3.1, which shows ESI mass spectra of the synthetic GlcNAc-modified peptide G-CTD at two different cone voltages. At a low voltage of 25 eV, the doubly charged glycosylated molecular ion at m/z 535.27 is the only peptide-related peak in the spectrum. However, at a more normal cone voltage of 35 eV, there is significant loss of the sugar residue to form the singly charged deglycosylated ion at m/z 866.42.

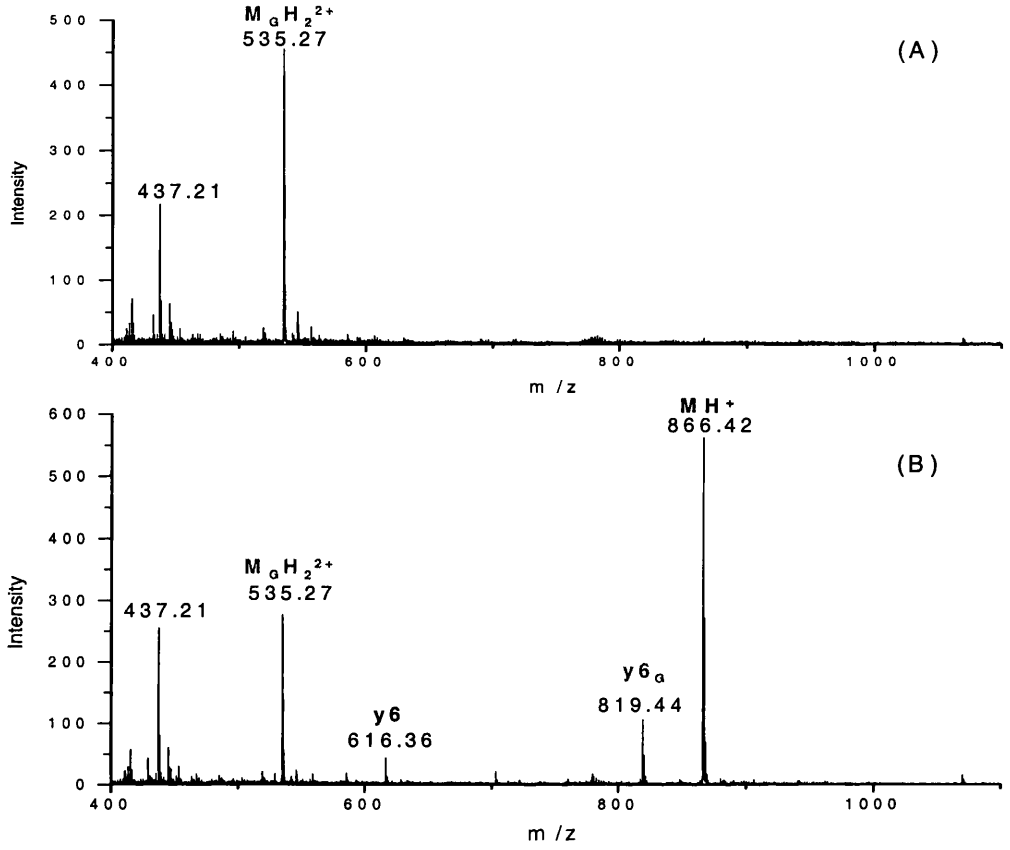


Figure 3.1: ESI-MS spectra of the G-CTD peptide.

These spectra were acquired at cone voltages of (A) 25 eV and (B) 35 eV. The peak at m/z 535.27 corresponds to the doubly charged glycosylated parent ion, whereas the peak at m/z 866.42 is the singly charged 'deglycosylated' parent ion. GlcNAc-modified ions are indicated with a $_G$. The peak at m/z 437.21 is a contaminant.

3.1.2. Positive Ion MALDI-MS

A similar effect is observed in MALDI-MS. In MALDI-MS the amount of energy imparted during ionisation can be varied either by varying the laser energy, or by using different matrices. The most widely used matrix is CHCA, which is a relatively 'hot' matrix, *i.e.* it transfers a lot of energy into the sample molecules during the ionisation process. A MALDI-MS spectrum of the G-CTD peptide in CHCA is shown in Figure 3.2. The glycosylated parent ion is observed as $[M + H]^+$ m/z 1069.56, $[M + Na]^+$ m/z 1091.57 and $[M + K]^+$ m/z 1107.54. There are also two ions formed by the loss of the GlcNAc residue. The ion at m/z 866.47 is caused by the prompt loss of the sugar residue in the source of the mass spectrometer. A second unresolved ion is present at m/z 880.61. This is a metastable ion, formed by loss of the GlcNAc during the first drift region of the reflectron.

The spectrum of the same peptide in DHB is shown in Figure 3.3. DHB is a 'cooler' matrix, and thus imparts less energy into the sample molecules during the ionisation process. This results in less removal of the sugar moiety during ionisation, and the deglycosylated prompt and metastable ions in Figure 3.3 are significantly less intense than when using the 'hotter' matrix (Figure 3.2).

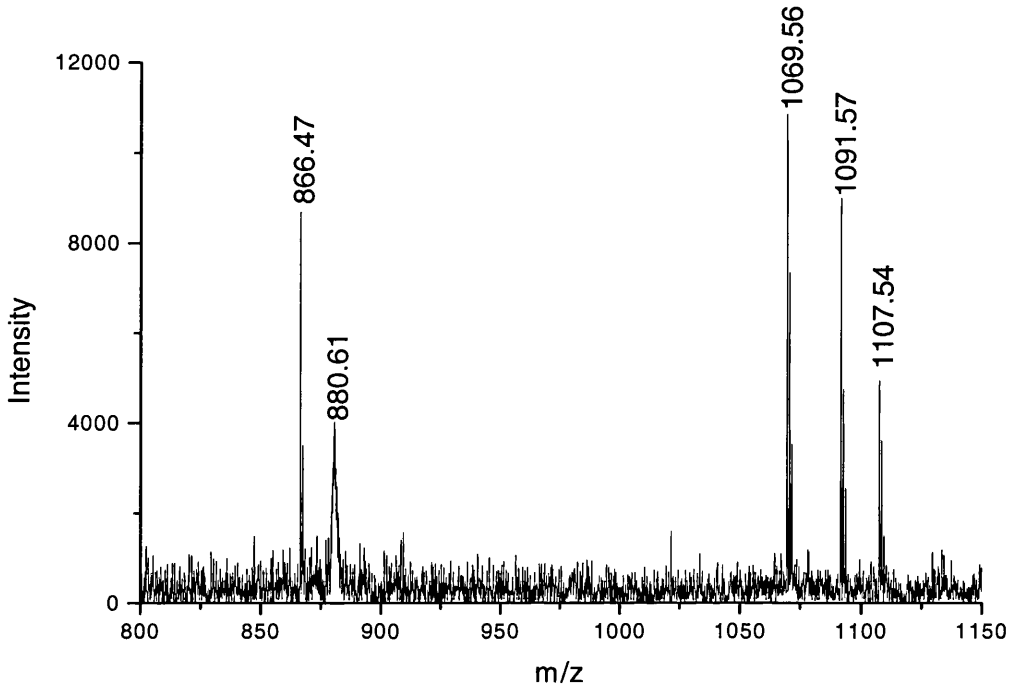


Figure 3.2: MALDI-MS spectrum of the GlcNAc-modified peptide G-CTD, acquired using CHCA as the matrix.

Three peaks correspond to the parent ion at $[M + H]^+$ m/z 1069.56, $[M + Na]^+$ m/z 1091.57, $[M + K]^+$ m/z 1107.54. The other two peaks are formed by the loss of the GlcNAc residue. The peak at m/z 866.47 is a product of cleavage of the sugar residue in the source region. The unresolved peak at m/z 880.61 is a metastable peak corresponding to the loss of the sugar moiety during the first stage of the reflectron TOF.

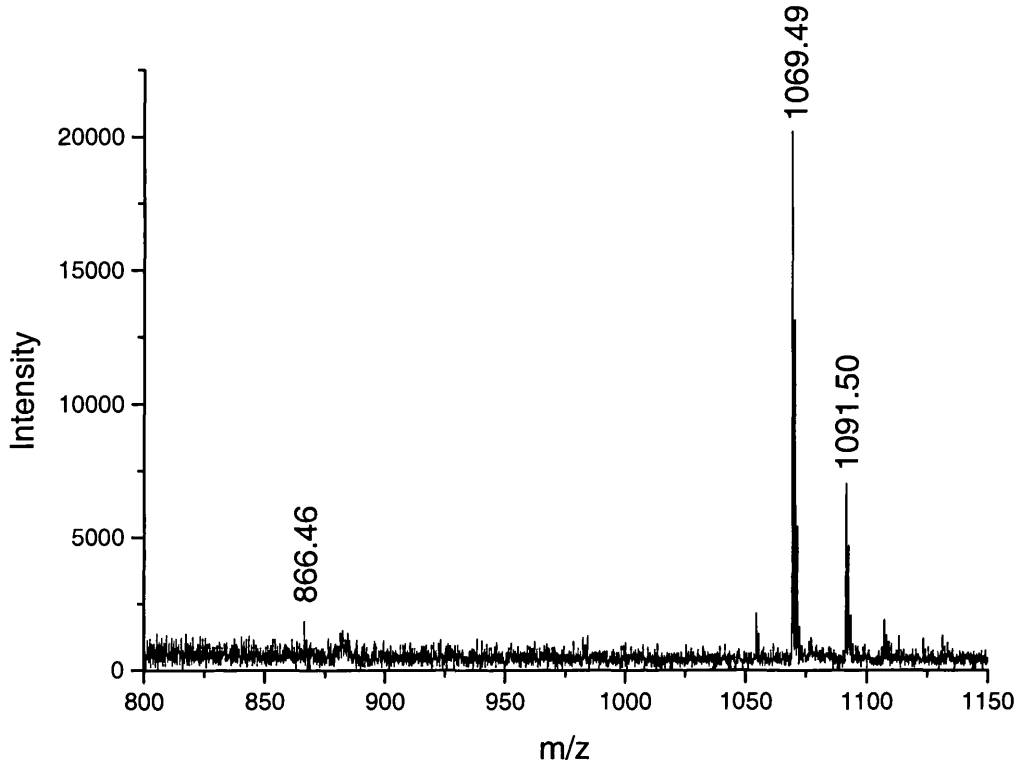


Figure 3.3: MALDI-MS spectrum of the GlcNAc-modified peptide G-CTD, acquired using DHB as the matrix.

The peaks in this spectrum correspond to the glycosylated parent ion at $[M + H]^+$ m/z 1069.49, $[M + Na]^+$ m/z 1091.50 and a very weak peak corresponding to the 'deglycosylated' peptide at m/z 866.46.

3.1.3. Negative Ion ESI-MS

The stability of the O-GlcNAc glycosidic bond was investigated in negative ion mode. Figure 3.4 shows a negative ion ESI-MS spectrum of the GlcNAc-modified peptide G-CTD. There are only two ions present in the spectrum $[M - H]^-$ m/z 1067.61, and an unidentified peak at m/z 1097.66. The peptide is observed in a singly charged state, unlike in positive ion mode, where it was a doubly charged species. This may be due to the lack of acidic residues in the peptide sequence that can become deprotonated. There is no sign of any deglycosylated ions formed by either prompt fragmentation or metastable decay. Despite increasing the cone voltage as high as 50 eV, there was still no

deglycosylation. This suggests that the glycosidic bond is more stable in negative ion mode. However, it is difficult to make a direct comparison to the positive ion ESI-MS spectrum, as singly-charged ions are generally more stable than their doubly-charged equivalents.

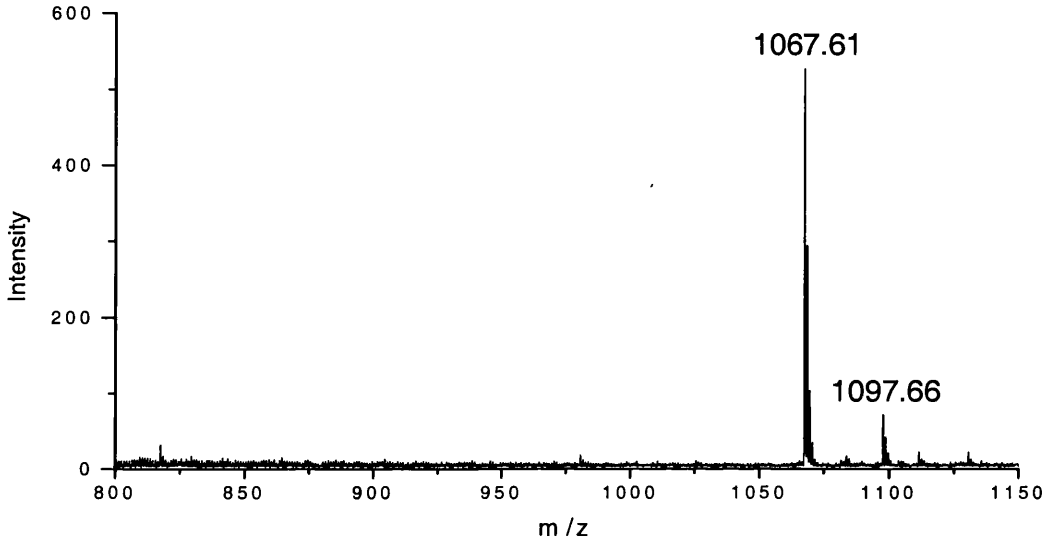


Figure 3.4: ESI-MS negative ion spectrum of the GlcNAc-modified G-CTD peptide. *The peak at m/z 1067.61 is the deprotonated G-CKII peptide. There are no peaks formed due to the loss of GlcNAc.*

3.2. Fragmentation Analysis of GlcNAc-Modified Peptides

The O-glycosidic link is dramatically more labile than bonds in the peptide backbone. This has caused problems in determining sites of modification using fragmentation analysis by mass spectrometry (see Section 1.3.3.2.)

3.2.1. MALDI-PSD

Figure 3.5 is a MALDI-PSD spectrum of the GlcNAc-modified synthetic peptide G-CKII, acquired using CHCA as the MALDI matrix. The identity of each labelled peak is given in Table 3.1. This is a typical PSD spectrum of a GlcNAc-modified peptide, where the dominant fragment ion is formed by the cleavage of the GlcNAc residue from the parent ion. The next most intense fragment is the 'deglycosylated' y_8 ion, which is

formed by a cleavage N-terminal to a proline residue, a favoured cleavage site in PSD and CID spectra[170]. In this spectrum the only glycosylated fragment ion is the $y8_G$ ion at m/z 1040.81. Hence, this spectrum narrows down the glycosylation site to one of the eight most C-terminal residues. As only serines and threonines can be O-GlcNAc-modified, this indicates one of the two consecutive serine residues is the site of modification. The high level of chemical noise in the MALDI-PSD spectrum dwarfs the low intensity fragment ions. Thus, low intensity glycosylated fragment ions are not visible.

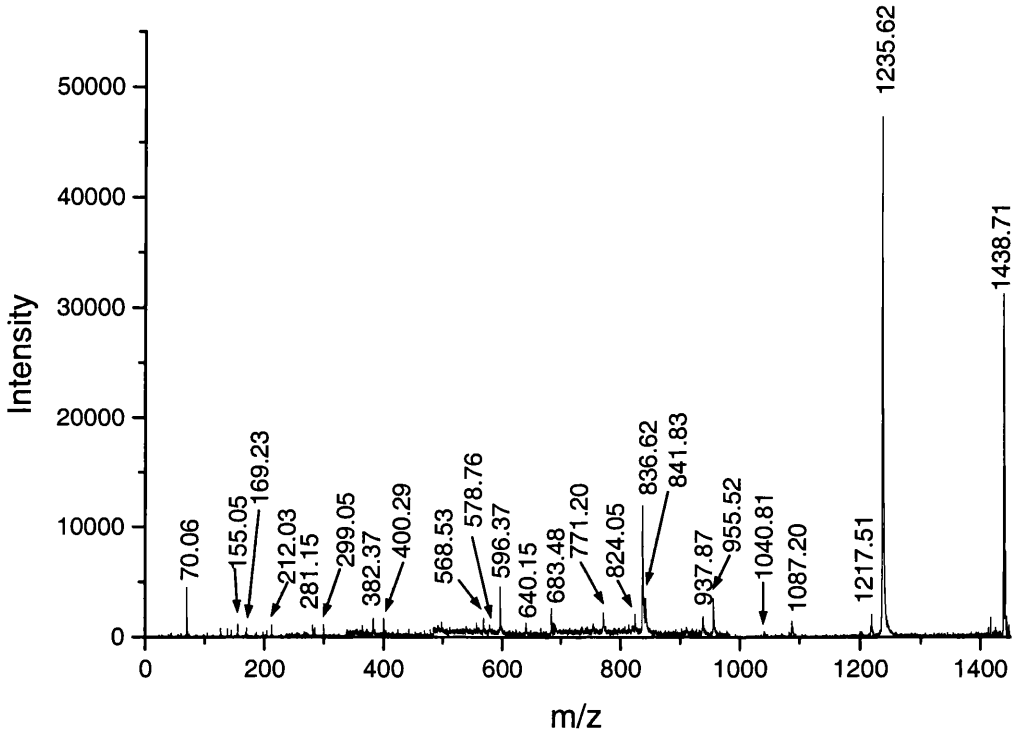
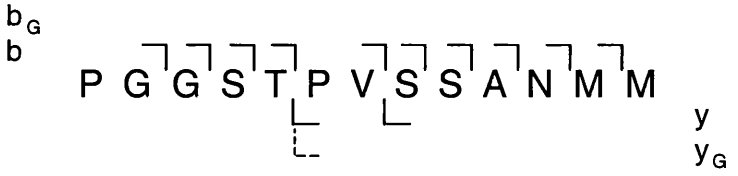


Figure 3.5: MALDI-PSD spectrum of the GlcNAc-modified G-CKII.

The major fragment ion in this spectrum is formed by the cleavage of the sugar moiety from the parent ion $[M + H]^+$ m/z 1235.62. All other fragment ions are 'deglycosylated', with the exception of the $y8_G$ ion at m/z 1040.81. Due to the lower resolution of PSD spectra, some of the higher mass peak labels are average masses.

Table 3.1: Identities of peaks observed in the MALDI-PSD spectrum of G-CKII.

This spectrum is shown in Figure 3.5.



Peak	Match	Peak	Match	Peak	Match
70.06	P	568.53	a7	841.83	b10
155.05	b2	578.76	b7-H_2O	937.87	b11-H_2O
169.23	PV_a	596.37	b7	955.52	b11
212.03	b3	640.15	y6	1040.22	$y8_G$
281.15	b4-H_2O	683.48	b8	1087.20	b12
299.05	b4	771.20	b9	1217.51	MH^+-NH_3
382.37	b5-H_2O	824.05	b10-H_2O	1235.62	MH^+
400.29	b5	836.62	y8	1438.71	MH_G^+

3.2.2. MALDI-CID-MS-MS of a Singly-Charged Peptide

A MALDI-CID-MS-MS spectrum of G-CKII is shown in Figure 3.6. Peak identities for this spectrum are given in Table 3.2. The pattern of fragmentation is exactly the same as that seen by MALDI-PSD (Figure 3.5). However, there is dramatically less chemical noise in the MALDI-CID-MS-MS spectrum, due to it being from an oa-TOF instrument. Also, due to the ability to adjust the collision energy, the peptide was fragmented using less energy, which should have increased the likelihood of detecting glycosylated fragment ions. This resulted in a fragmentation spectrum where the parent ion (m/z 1438.65) and the 'deglycosylated' parent ions (m/z 1235.58) are more intense, and immonium ions (a product of higher energy fragmentation) are weaker in intensity, compared to the MALDI-PSD spectrum.

A number of low intensity fragment ions are visible that could not be observed in the MALDI-PSD spectrum. However, the majority of these new ions are losses of water from 'b' ions, and only one new fragment ion is glycosylated, the $b11_G$ ion at m/z 1158.51. The other glycosylated fragment ion present is the $y8_G$ at m/z 1039.51, which is also observed by MALDI-PSD.

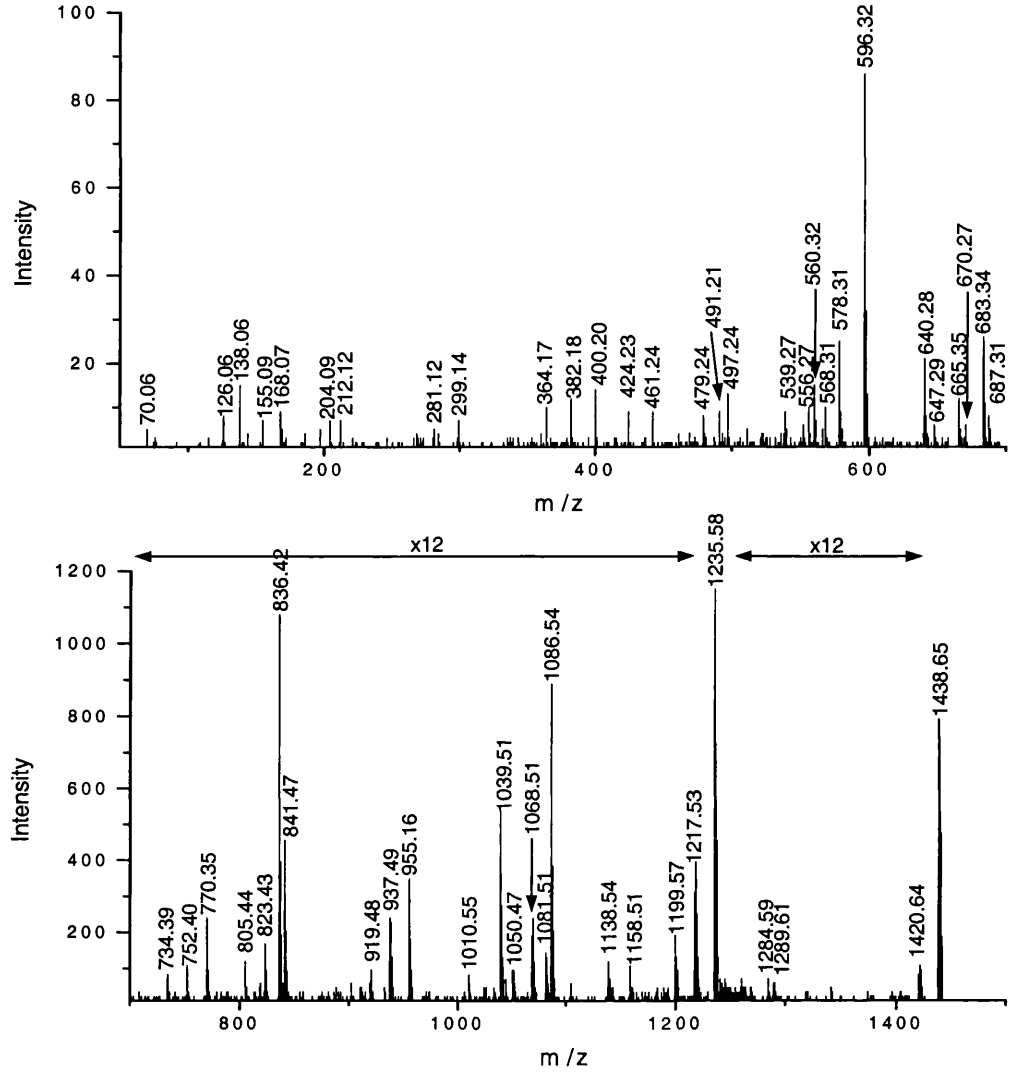
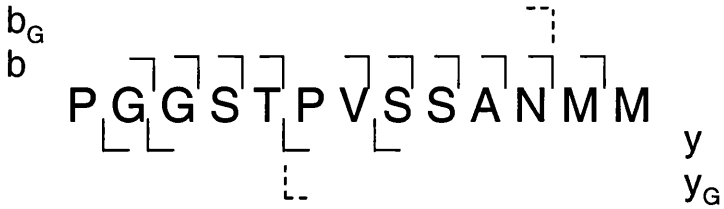


Figure 3.6: MALDI-CID-MS-MS spectrum of the singly-charged G-CKII peptide.

The major fragment in the spectrum is the 'deglycosylated' intact peptide at m/z 1235.58. There are a number of low intensity fragment ions in this spectrum, but only two are glycosylated fragments; the $y8_G$ ion at m/z 1039.47, and the $b11_G$ ion at m/z 1158.51.

Table 3.2: Peaks observed in the MALDI-CID-MS-MS spectrum of G-CKII.

This spectrum is shown in Figure 3.6.



Peak	Match	Peak	Match	Peak	Match
70.06	P	560.32	b7-2H_2O	937.49	y9 or b11-H_2O
126.06	GlcNAc fragment	568.31	a7	955.16	b11
138.06	GlcNAc fragment	578.31	b7-H_2O	1010.55	
155.09	b2	596.32	b7	1039.51	$y8_G$
168.07	GlcNAc fragment	640.28	y6	1050.47	b12-2H_2O
204.09	GlcNAc	647.29	b8-2H_2O	1068.51	b12-H_2O
212.12	b3	665.35	b8-H_2O	1081.51	y11
281.12	y2	670.27	PVSSANM-NH_3	1086.54	b12
299.14	b4	683.34	b8	1138.54	y12
364.17	b5-2H_2O	687.31	PVSSANM	1158.51	$b11_G$
382.18	b5-H_2O	734.39	b9-2H_2O	1199.57	MH^+-2H_2O
400.20	b5	752.40	b9-H_2O	1217.53	MH^+-H_2O
424.23	PVSSA-H_2O	770.35	b9	1235.58	MH^+
461.24	b6-2H_2O	805.44	b10-2H_2O	1284.59	y16
479.24	b6-H_2O	823.43	b10-H_2O	1420.64	MH_G-H_2O
497.24	b6	836.42	y8	1438.65	MH_G
539.27	PVSSAN-NH_3	841.47	b10		
556.27	PVSSAN	919.48	b11-2H_2O		

3.3. ESI-CID-MS-MS

3.3.1.1. ESI-CID-MS-MS of G-CKII

Electrospray MS-MS is generally of multiply charged ions. The extra charge gives rise to fragmentation spectra with a different appearance to those of singly charged ions, even though the types of fragment ions observed are essentially the same. An ESI-CID-MS-MS spectrum of a doubly charged ion of G-CKII is given in Figure 3.7. The immediate difference between the CID spectrum of a doubly charged glycosylated ion and the singly charged ions produced by MALDI is the formation of the ion at m/z 204.1. This is the characteristic oxonium ion of a HexNAc residue, and is always a dominant ion in positive ion ESI-CID spectra of GlcNAc-modified peptides. Also observed are ions at m/z 186.12 and m/z 168.10. These are fragment ions formed by successive losses of water from the GlcNAc oxonium ion. A full list of GlcNAc related fragment ions is presented in Appendix 1. As in the MALDI fragmentation spectrum, there is a major ion corresponding to the deglycosylated parent ion ($[M + H]^+$ m/z 1235.72, $[M + 2H]^{2+}$ m/z 618.41). However, by magnifying regions of this spectrum a number of low intensity fragment ions are visible, several of which are glycosylated. The identities of the fragment ions are described in Table 3.3. The most important of these ions are the $y6_G$ ion at m/z 843.52 and the $b8_G$ ion at m/z 886.61. These two ions permit the assignment of the GlcNAc modification to serine 8.

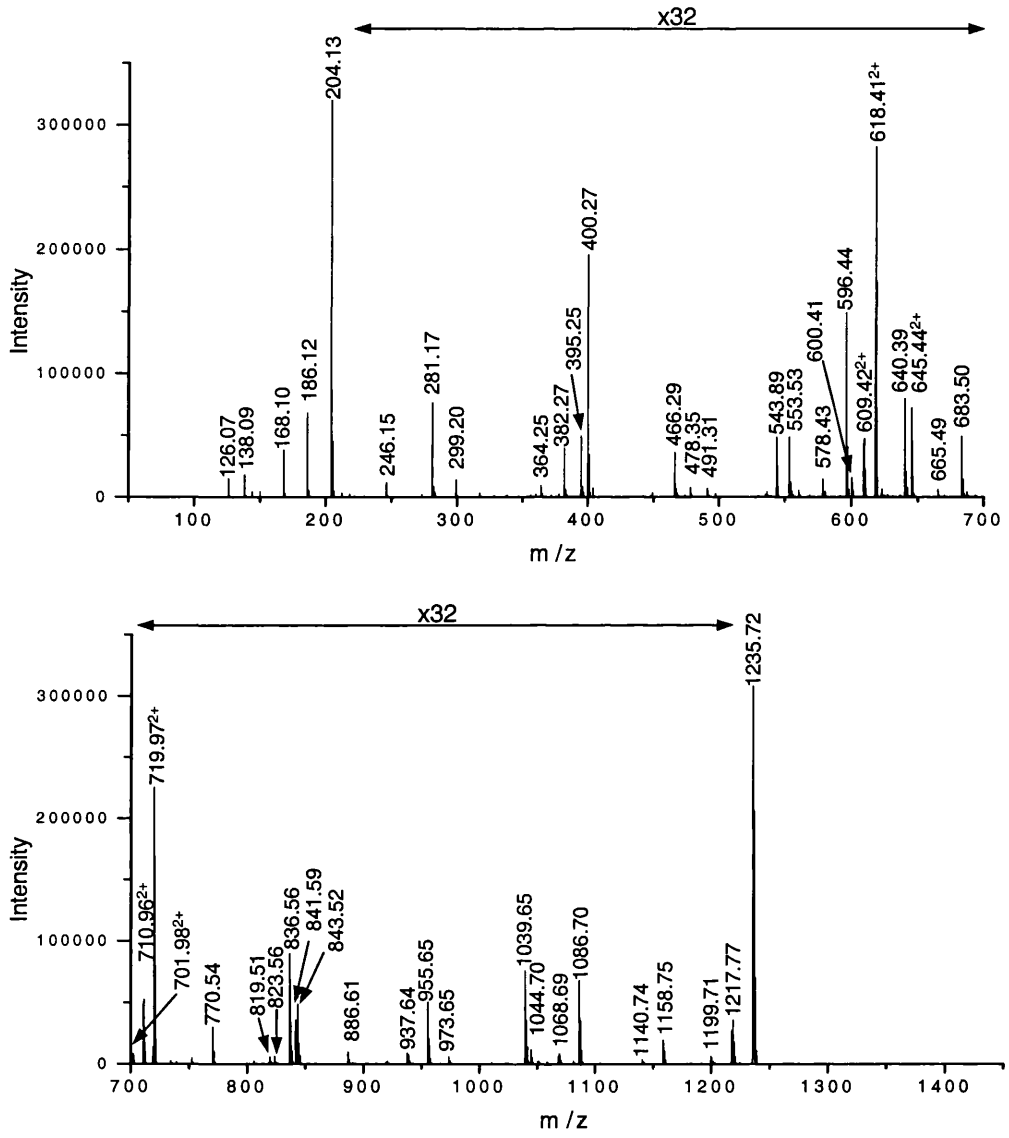
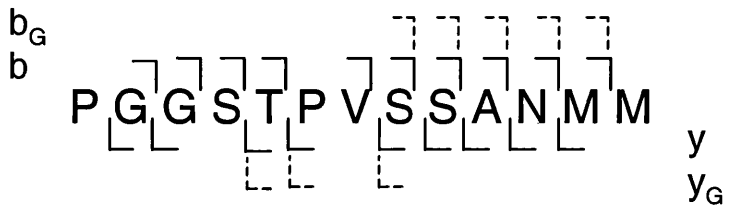


Figure 3.7: ESI-CID-MS-MS spectrum of the glycosylated G-CKII peptide.

The intense peaks in this spectrum include the doubly charged parent ion at m/z 719.97, 'deglycosylated' parent ion ($[M + H]^+$ m/z 1235.72, $[M + 2H]^{2+}$ m/z 618.41), and the GlcNAc oxonium ion at m/z 204.13. The spectrum also contains a number of low intensity ions, including many glycosylated ions. The glycosylated $y6_G$ ion at m/z 843.52 and $b8_G$ ion at m/z 886.61 are sufficient to identify serine 8 as being GlcNAc-modified.

Table 3.3: Peaks observed in the ESI-CID-MS-MS spectrum of G-CKII.

This spectrum is shown in Figure 3.7.



Peak	Match	Peak	Match	Peak	Match
126.07	GlcNAc fragment	553.53	y5	836.56	y8
138.09	GlcNAc fragment	578.43	b7-H_2O	841.59	b10
168.10	GlcNAc fragment	596.44	b7	843.52	$y6_G$
186.12	GlcNAc fragment	600.41^{2+}	MH_2^{2+}-$2H_2O$	886.61	$b8_G$
204.13	GlcNAc	609.42^{2+}	MH_2^{2+}-H_2O	937.64	y9
246.15	SSA	618.41^{2+}	MH_2^{2+}	955.65	b11
281.17	y2	640.39	y6	973.65	$b9_G$
299.20	b4	645.44^{2+}	$b12_G$	1039.65	$y8_G$
364.25	b5-$2H_2O$	665.49	b8-H_2O	1044.70	$b10_G$
382.27	b5-H_2O	683.50	b8	1068.69	b12-H_2O
395.25	y3	701.98^{2+}	$M_GH_2^{2+}$-$2H_2O$	1086.70	b12
400.27	b5	710.96^{2+}	$M_GH_2^{2+}$-H_2O	1140.74	$y9_G$
466.29	y4	719.97^{2+}	$M_GH_2^{2+}$	1158.75	$b11_G$
478.35^{2+}	$b11^{2+}$	770.54	b9	1199.71	MH^+-$2H_2O$
491.31	SSANM	819.51	y8-NH_3	1217.77	MH^+-H_2O
543.89	TPVSSA	823.56	b10-H_2O	1235.72	MH^+

3.3.1.2. ESI-CID-MS-MS of G-CTD

Figure 3.8 shows two ESI-CID-MS-MS spectra of the G-CTD peptide acquired using slightly different collision offsets. Both spectra contain the same ions, and the identities of these peaks are given in Table 3.4. G-CTD represents a typical tryptic peptide, containing a basic C-terminal residue. The charge is preferentially retained on this residue, causing the resulting fragmentation to be dominated by C-terminally derived 'y' type fragment ions. Hence, a GlcNAcylation site can only be implied by the lack of a glycosylated version of a 'y' ion. Glycosylated and 'deglycosylated' y4 fragment ions

are observed at m/z 621.34 and m/z 418.24 respectively. However, for the y3 ion only the unmodified version at m/z 331.20 is present. In the CID spectrum acquired using 20 eV of collision energy, the lack of a glycosylated y3 ion should not be used to assign the glycosylation site, as the glycosylated fragment ions are all less intense than their 'deglycosylated' counterparts. However, in the spectrum produced using a collision energy of 15 eV many of the glycosylated fragment ions are more intense than their 'deglycosylated' versions. Thus, the absence of a $y3_G$ in the presence of a y3 ion could be used to assign serine 5 as the GlcNAc modification site.

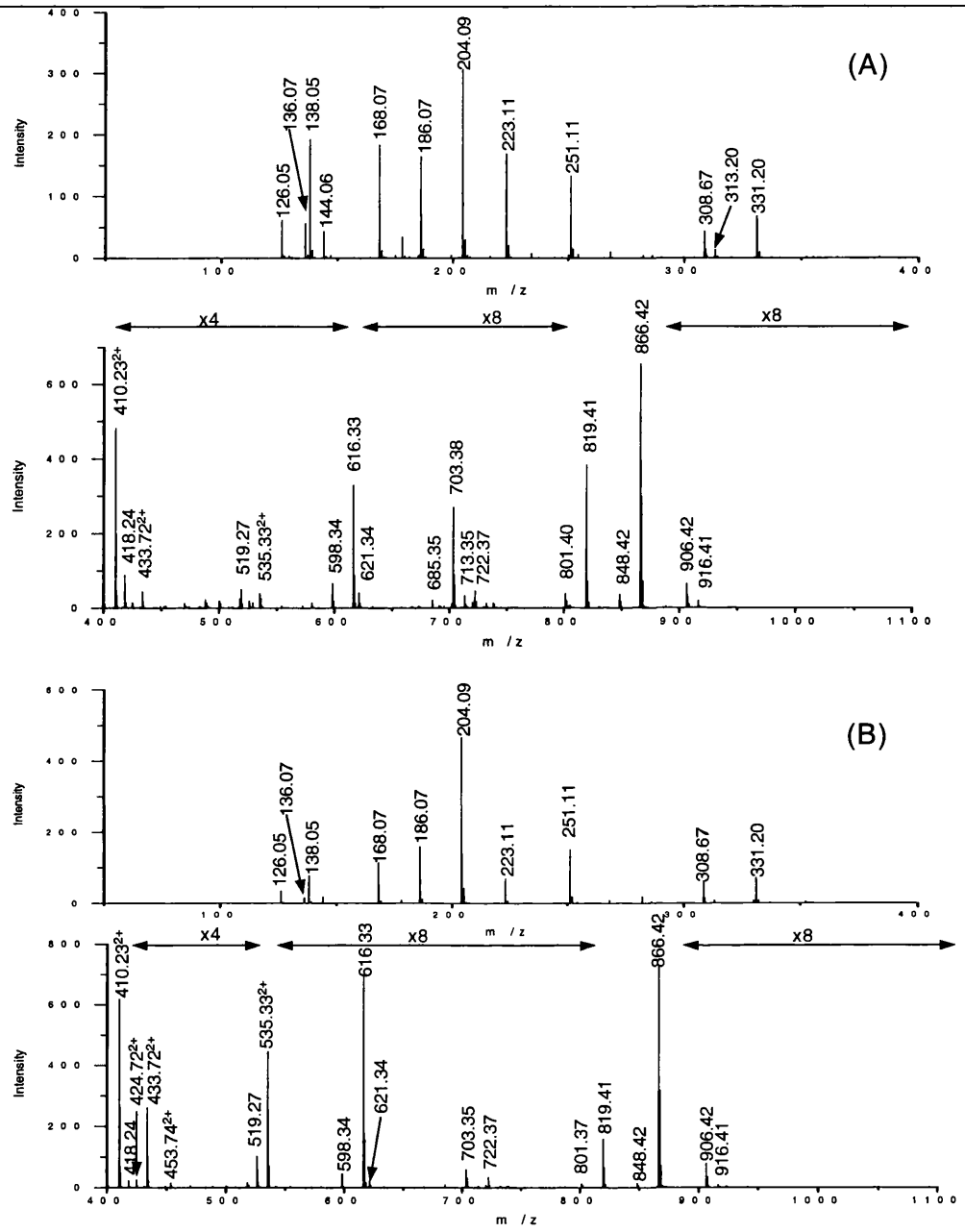
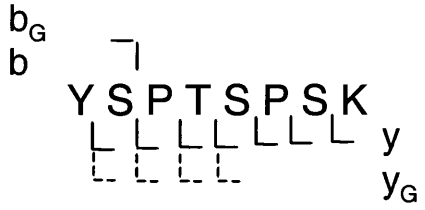


Figure 3.8: ESI-CID-MS-MS spectra of the glycosylated G-CTD peptide at two different collision offsets.

Using a collision energy of (A) 20 eV, intense peaks include the 'deglycosylated' parent ion (m/z 866.42), GlcNAc oxonium ion (m/z 204.09), $y6_G$ (m/z 819.41) and 'deglycosylated' y_6 (m/z 616.33). Using a collision energy of (B) 15 eV the doubly charged parent ion (m/z 535.27) is still present, and some glycosylated fragment ions are more intense than their 'deglycosylated' counterparts. There is a glycosylated $y6_G$ ion at m/z 621.34, but no glycosylated y_3 ion. This implies that serine 5 is GlcNAc-modified.

Table 3.4: Peaks observed in the ESI-CID-MS-MS spectra of G-CTD.

Spectra were acquired at two different collision offsets and are shown in Figure 3.8.



Peak	Match	Peak	Match	Peak	Match
126.05	GlcNAc fragment	410.23^{2+}	$y6_G$	703.38	y7
136.07	Y	418.24	y4	713.35	
138.05	GlcNAc fragment	424.72^{2+}	MH_2^{2+}-H_2O	722.37	$y5_G$
168.07	GlcNAc fragment	433.72^{2+}	MH_2^{2+}	801.40	$y6_G$-H_2O
186.07	GlcNAc fragment	453.74^{2+}	$y7_G$	819.41	$y6_G$
204.09	GlcNAc	519.27	y5	848.42	MH^+-H_2O
223.11	a2	535.33^{2+}	$M_GH_2^{2+}$	866.42	MH^+
251.11	b2	598.34	y6-H_2O	906.42	$y7_G$
308.67^{2+}	y6	616.33	y6	916.41	
313.20	y3-H_2O	621.34	$y4_G$		
331.20	y3	685.35	y7-H_2O		

3.3.1.3. ESI-CID-MS-MS in Negative Ion Mode

The glycosidic bond appeared to be more stable in negatively charged ions (see Section 3.1.3). This should increase the likelihood of observing glycosylated fragment ions. Thus, it was investigated whether negative ion ESI-CID-MS-MS was a useful technique for determining O-GlcNAc modification sites. The fragmentation spectrum of the G-CTD peptide $[M - H]^-$ m/z 1067.61 is presented in Figure 3.9. This spectrum looks completely different to the corresponding spectrum in positive ion mode. One of the significant differences is that instead of the formation of an ion due to the loss of 203 Da (GlcNAc) from the parent ion, there is an ion at m/z 846.51, formed by the cleavage of the GlcNAc and the hydroxyl group from the side chain of the serine to which the sugar residue is attached. This will have converted the glycoserine residue into a

dehydroalanine, and caused a loss of 18 Da in mass of this residue. Thus, if a series of ions form a sequence tag for this region, then this characteristic difference of 69 Da would enable the site of modification to be determined. However, the higher mass region of the spectrum contains no ions that differ by masses corresponding to amino acid residues. The only series of ions that give sequence information are the y_1 (m/z 145.08), y_2 (m/z 232.12) and y_3 (m/z 329.19) ions, and all the other ions in this region of the mass spectrum, would make this sequence difficult to interpret *de novo* if the peptide sequence was not already known.

The only other ion that can be assigned in this spectrum is the ion at m/z 202.10, which is probably the negatively charged GlcNAc oxonium ion. The majority of the ions could not be assigned to predicted peptide fragment ions.

3.3.1.4. Sensitivity of Site Determination by Nanospray-MS

The sensitivity at which a GlcNAc-modified peptide can be detected is much lower than the level at which one can determine which residue is modified. In positive ion ESI-MS GlcNAc-modified peptides give a peak that is approximately half the intensity of that from the corresponding unmodified peptide[59]. Nevertheless, low fmole amounts of glycopeptides can be observed by electrospray ionisation. However, the amount of sample required to identify the site of modification is significantly higher.

To determine the lowest level at which a site can be identified by nanospray-MS, G-CKII peptide was sprayed at decreasing concentrations. From a solution at 100 fmoles/µl, after five minutes spraying the $y6_G$ and $b8_G$ ions can both be observed (Figure 3.10). Although the flow rate for nanospray-MS is not directly measured, a typical flow rate for nanospray-MS is 10 - 20 nl/min. Hence, the site of modification of G-CKII can be determined from a consumption of 5 – 10 fmoles of peptide.

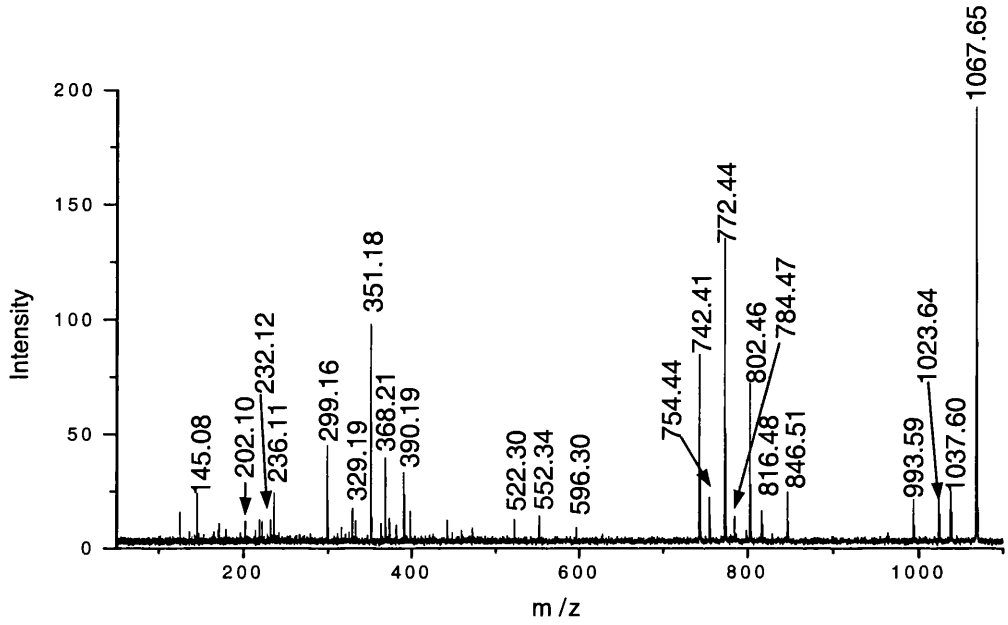


Figure 3.9: ESI-CID-MS-MS negative ion spectrum of G-CTD peptide.

The ion at m/z 846.51 corresponds to the loss of GlcNAc and the hydroxyl side chain of the modified serine residue. None of the higher mass fragment ions differ by masses corresponding to amino acid residues, making this spectrum very difficult to interpret. The majority of ions could not be assigned.

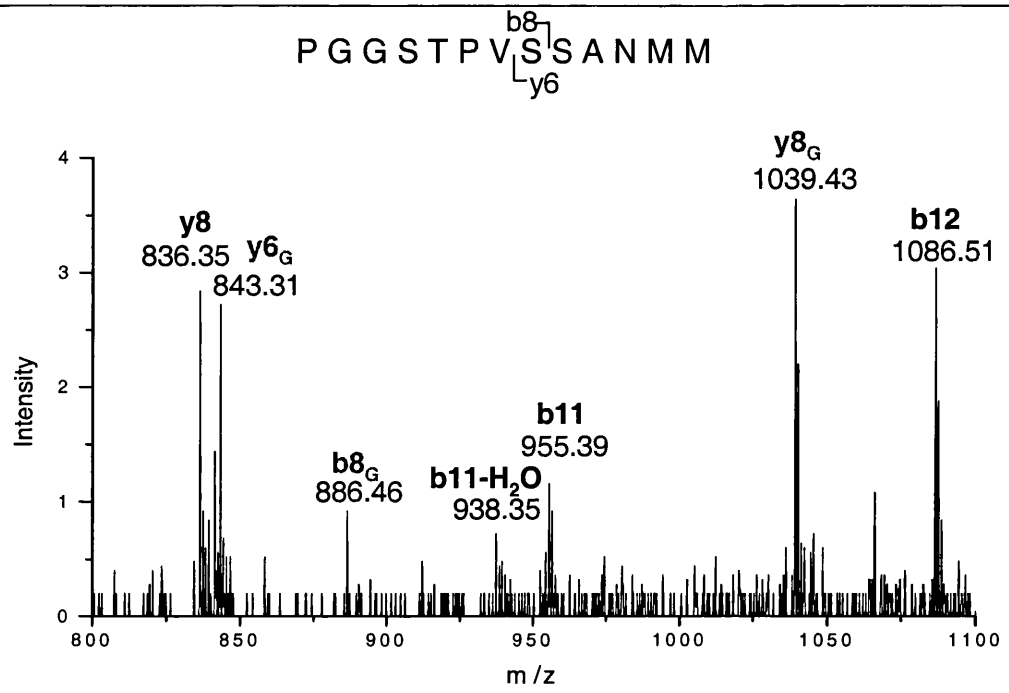


Figure 3.10: Magnified region of a nanospray-CID-MS-MS spectrum of the peptide G-CKII.

G-CKII was infused by nanospray-MS at a concentration of 100 fmoles/µl. This spectrum is the accumulation of five minutes of acquisition. The glycosylated $y6_G$ (m/z 843.52) and $b8_G$ (m/z 886.61) ions identify serine 8 as being GlcNAc-modified.

3.3.1.5. Sensitivity of Site Determination by LC-ESI-CID-MS-MS

Although it is possible to identify sites of modification at high sensitivity by nanospray-MS when a sample is pure, this is not the case when analysing a 'real' sample. Samples from protein digests, especially in-gel digests, contain salt and other contaminants that suppress peptide ion signals, or may even prevent the sample from spraying efficiently. For these samples, post-digest sample clean-up is required. This is usually carried out using C18 column packing material. Sample is loaded onto the packing material in low organic solution, then washed using more low organic solvent. Peptides are retained on the column, whilst salts are washed away. Peptides can then be eluted off using a higher organic solvent. The danger of a washing step is that it causes

loss of sample whilst getting rid of contaminants. Sample loss is minimised by on-line sample clean-up. This can easily be carried out prior to LC-MS. Sample is loaded and washed on a C18 guard column, then eluted directly onto a second analytical C18 column, where peptides are separated on the basis of hydrophilicity/hydrophobicity. As they elute off the column, peptides are fed directly into the mass spectrometer. This set-up not only cleans up the sample, but the column provides sample concentration and the peptide separation reduces peptide ion suppression by more abundant peptides present in the same mass spectrum. Thus, sample can be loaded at a lower concentration in a relatively large volume, then concentrated and eluted in much smaller volumes.

Figure 3.11 shows the magnified region containing the $y6_G$ and $b8_G$ ions from ESI-CID-MS-MS spectra from two LC-MS analyses of G-CKII. For the top spectrum, 100 fmoles of G-CKII peptide were loaded, whereas the bottom spectrum was acquired from 50 fmoles of peptide. In both spectra, peaks for the $y6_G$ and $b8_G$ ions are present. However, in the spectrum of 50 fmoles of sample the $b8_G$ ion is present at a signal to noise of only 2:1. Hence, assignment of this peak is tenuous. These results suggest a limit of site determination by LC-ESI-CID-MS-MS of 50 – 100 fmoles of peptide loaded.

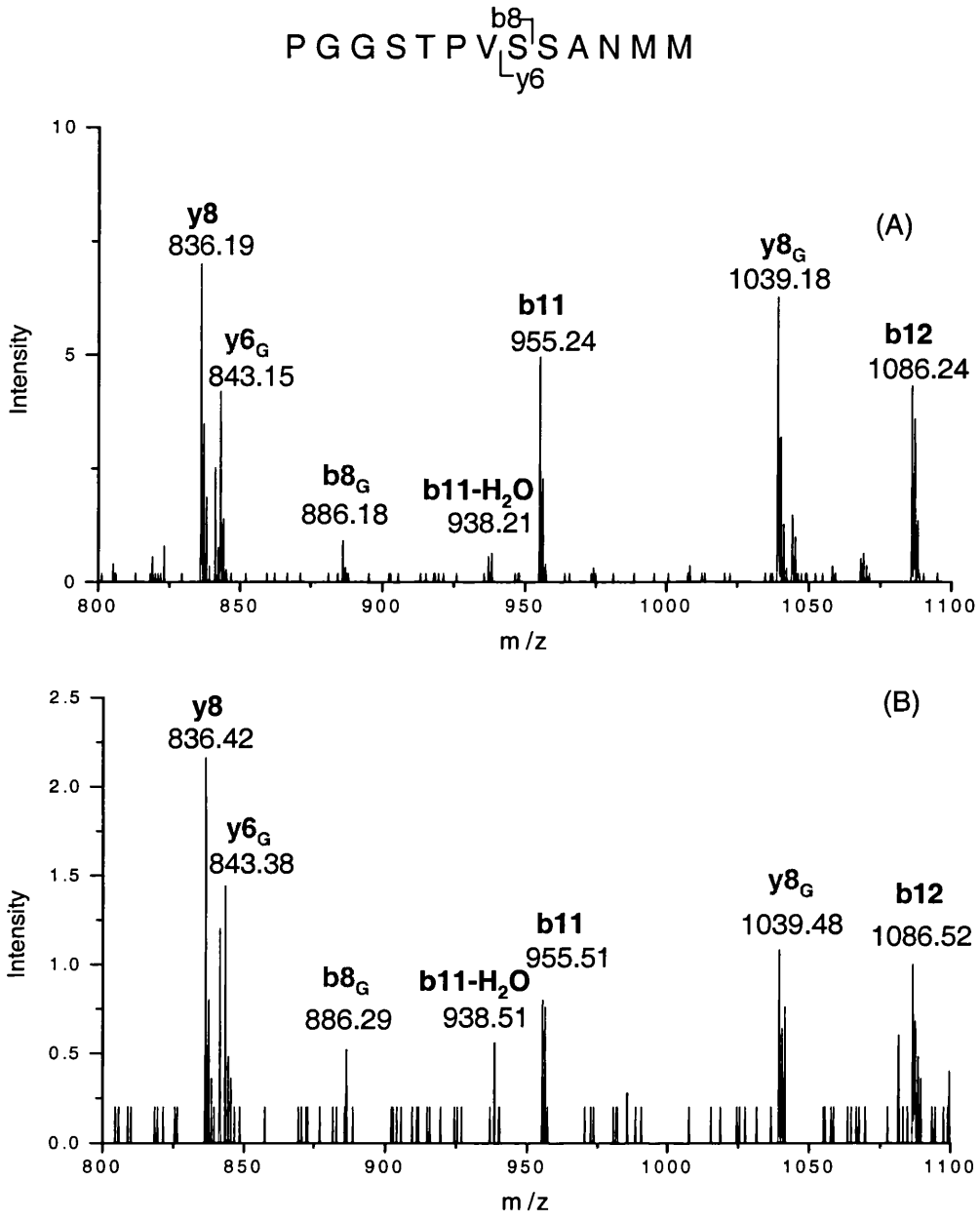


Figure 3.11: Magnified regions of LC-ESI-CID-MS-MS spectra of the G-CKII peptide.

(A) 100 fmoles and (B) 50 fmoles of G-CKII were loaded onto the HPLC. The glycosylated $y6_G$ (m/z 843.15) and $b8_G$ ions (m/z 886.18) identify serine 8 as being modified.

3.3.2. Fragmentation of GlcNAc-Gal-Modified Peptides

It is common practice to derivatise GlcNAc residues enzymatically with radiolabelled galactose for purposes of locating glycosylated proteins and for isolating modified peptides from a complex mixture using the lectin *Ricinus communis* (RCA I) [169, 171]. Hence, the effect the additional galactose residue has on the ability to identify the modified residue by tandem mass spectrometry was investigated. To this end, synthetic GlcNAc-modified peptides were *in vitro* galactosylated, and their CID spectra were acquired. Figure 3.12 shows an ESI-CID-MS-MS spectrum of the galactosylated G-CKII and Table 3.4 lists the identities of peaks observed. As previously reported, diagnostic fragment ions from the GlcNAc-Gal disaccharide moiety (m/z 366.23) and GlcNAc monosaccharide residue (m/z 204.13) are observed [157, 169]. As a result, the fragment ion spectrum contains three fragment ions for each peptide backbone cleavage; a GlcNAc-Gal modified peptide, a GlcNAc-modified peptide and a fully 'deglycosylated' peptide. For the y_6 fragment this trio appear at m/z 1005.58, m/z 843.45 and m/z 640.39 respectively, and for the b_8 ion they are at m/z 1048.68, m/z 886.61 and m/z 683.50.

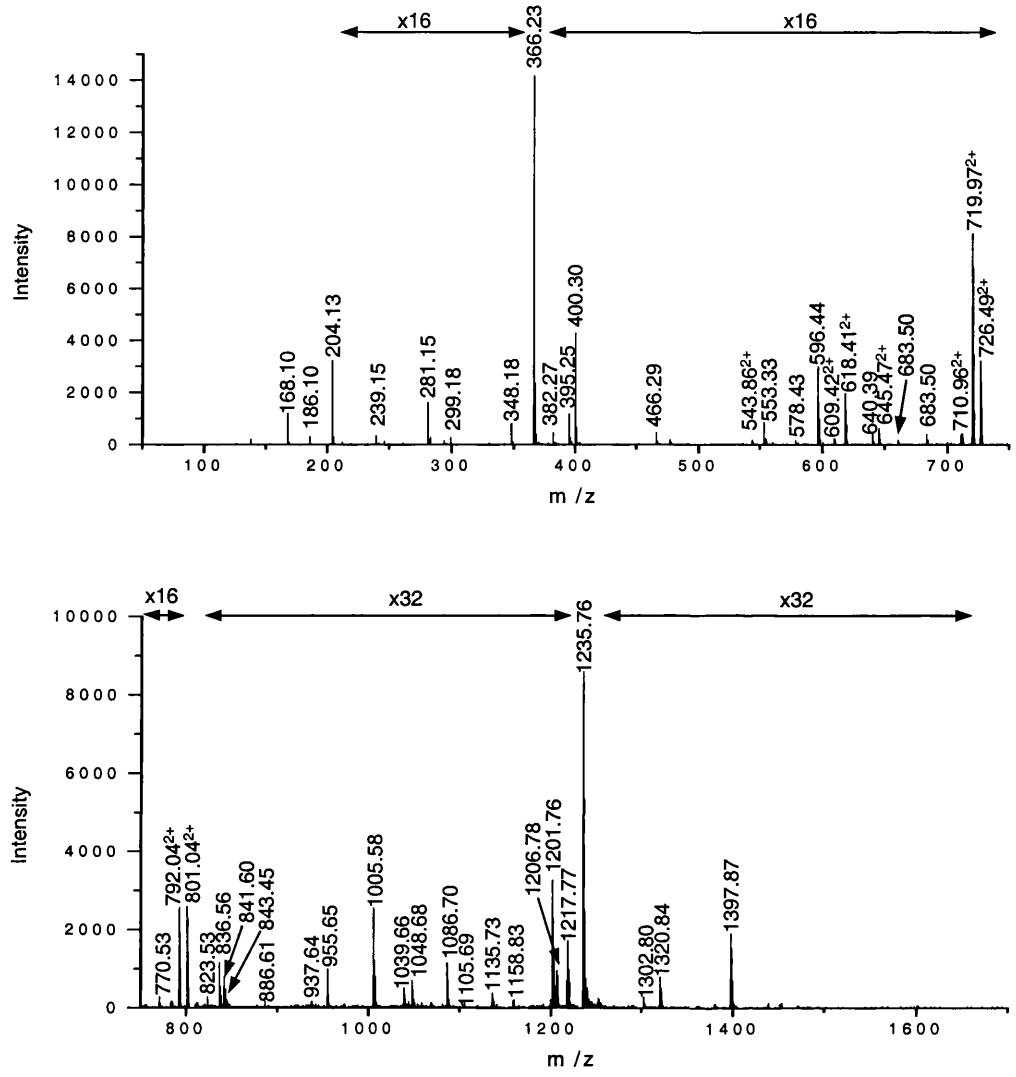
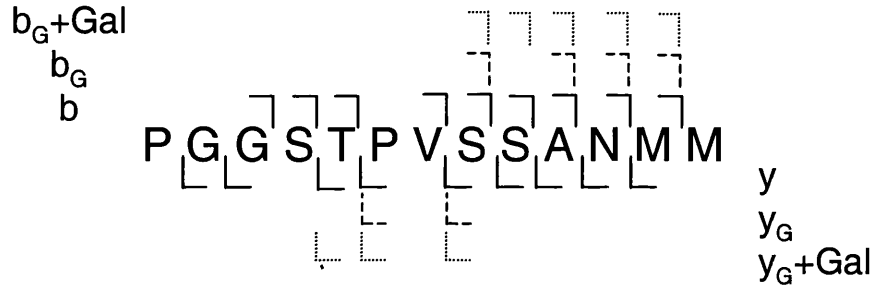


Figure 3.12: ESI-CID-MS-MS spectrum of galactosylated G-CKII peptide.

The major ions in the spectrum are the GlcNAc-Gal modified parent ion at $[M + 2H]^{2+}$ m/z 801.00, and ions due to the cleavage of the Gal ($[M + 2H]^{2+}$ m/z 719.97) and GlcNAc-Gal ($[M + H]^{+}$ m/z 1235.76) residues. Also present are the GlcNAc and GlcNAc-Gal oxonium ions at m/z 204.13 and m/z 366.23 respectively. Many glycosylated fragment ions are present including GlcNAc- and GlcNAc-Gal-modified fragment ions of b8 (m/z 886.61 and m/z 1048.68) and y6 (m/z 843.45 and m/z 1005.58) respectively, which identify serine 8 as the glycosylated residue.

Table 3.5: Identities of peaks observed in the ESI-CID-MS-MS spectrum of galactosylated G-CKII.

This spectrum is shown in Figure 3.12.



Peak	Match	Peak	Match	Peak	Match
168.10	GlcNAc fragment	609.42^{2+}	MH_2^{2+}-H_2O	937.64	y9
186.10	GlcNAc fragment	618.41^{2+}	MH_2^{2+}	955.65	b11
204.13	GlcNAc	640.39	y6	1005.58	$y6_G$Gal
239.15		645.47^{2+}	$b12_G$	1039.66	$y8_G$
281.15	y2	683.50	b8	1048.68	$b8_G$Gal
299.18	b4	710.96^{2+}	$M_GH_2^{2+}$-H_2O	1086.70	b12
348.18	GlcNAcGal-H_2O	719.97^{2+}	$M_GH_2^{2+}$	1104.63	$y7_G$Gal
366.23	GlcNAcGal	726.49^{2+}		1135.73	$b9_G$Gal
382.27	b5-H_2O	770.53	b9	1158.83	$b11_G$
395.25	y3	792.04^{2+}	$M_GGalH_2^{2+}$-H_2O	1201.76	$y8_G$Gal
400.30	b5	801.04^{2+}	$M_GGalH_2^{2+}$	1206.78	$b10_G$Gal
466.29	y4	823.53	b10-H_2O	1217.77	MH^+-H_2O
543.86	TPVSSA	836.56	y8	1235.76	MH^+
553.33	y5	841.60	b10	1302.80	$b11_G$Gal –H_2O or $y9_G$Gal
578.43	b7-H_2O	843.45	$y6_G$	1320.84	$b11_G$Gal
596.44	b7	886.61	$b8_G$	1397.87	$MHGal^+$

3.3.3. β-Elimination followed by MALDI-PSD

O-linked sugars can be eliminated from attached peptides using a strong base, and this reaction is a standard protocol to release of O-linked sugar chains for subsequent glycosylation analysis. In a strongly basic environment, sugar chains are relatively stable, but peptides are easily degraded, making their subsequent analysis more difficult.

Figure 3.13 shows spectra of G-CTD before and after undergoing a β-elimination reaction in ammonium hydroxide, showing the production of two β-eliminated peaks at $[M + H]^+$ m/z 848.55, $[M + Na]^+$ m/z 870.53.

The protonated species was then selected for MALDI-PSD analysis, and the resulting spectrum is given in Figure 3.14. Also presented is a MALDI-PSD spectrum of the unmodified N-CTD peptide. Comparing these spectra, one sees that fragment ions y_4 - y_7 differ in mass by 18Da with respect to each other, whereas the y_3 ions in both spectra are at m/z 331.2. The difference in mass between the y_3 and y_4 ions is 69 Da, which corresponds to a dehydroalanine residue, and shows this residue was previously a glycosylated serine residue.

Thus, for this peptide, β-elimination was effective at determining the site of O-GlcNAcylation. However, for another synthetic GlcNAc-modified peptide, the β-elimination reaction was not as successful. The G-Myc peptide is a much larger peptide and is glycosylated on a threonine residue. This produced very little β-eliminated product, even after very long incubations (data not shown).

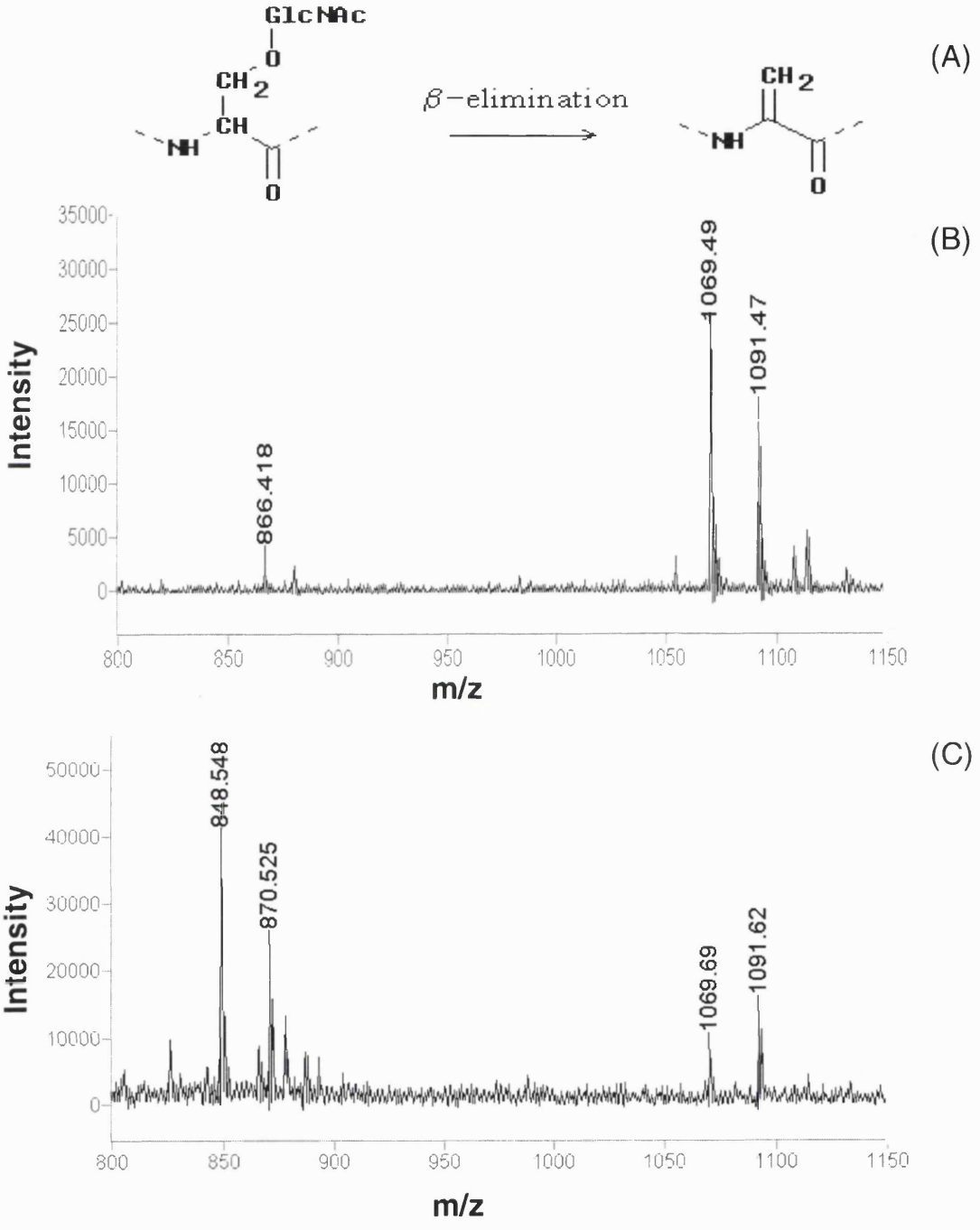


Figure 3.13: MALDI spectra of the G-CTD peptide before and after β-elimination.

The β-elimination reaction (A) converts the glycoserine to a dehydroalanine. This converts the GlcNAc-modified peptide observed in (B) at $[M + H]^+$ m/z 1069.49,$[M + Na]^+$ m/z 1091.47 into the β-eliminated peptide observed in (C) at $[M + H]^+$ m/z 848.55, $[M + Na]^+$ m/z 870.53. The peak at m/z 866.418 in (B) is 'deglycosylated' G-CTD.

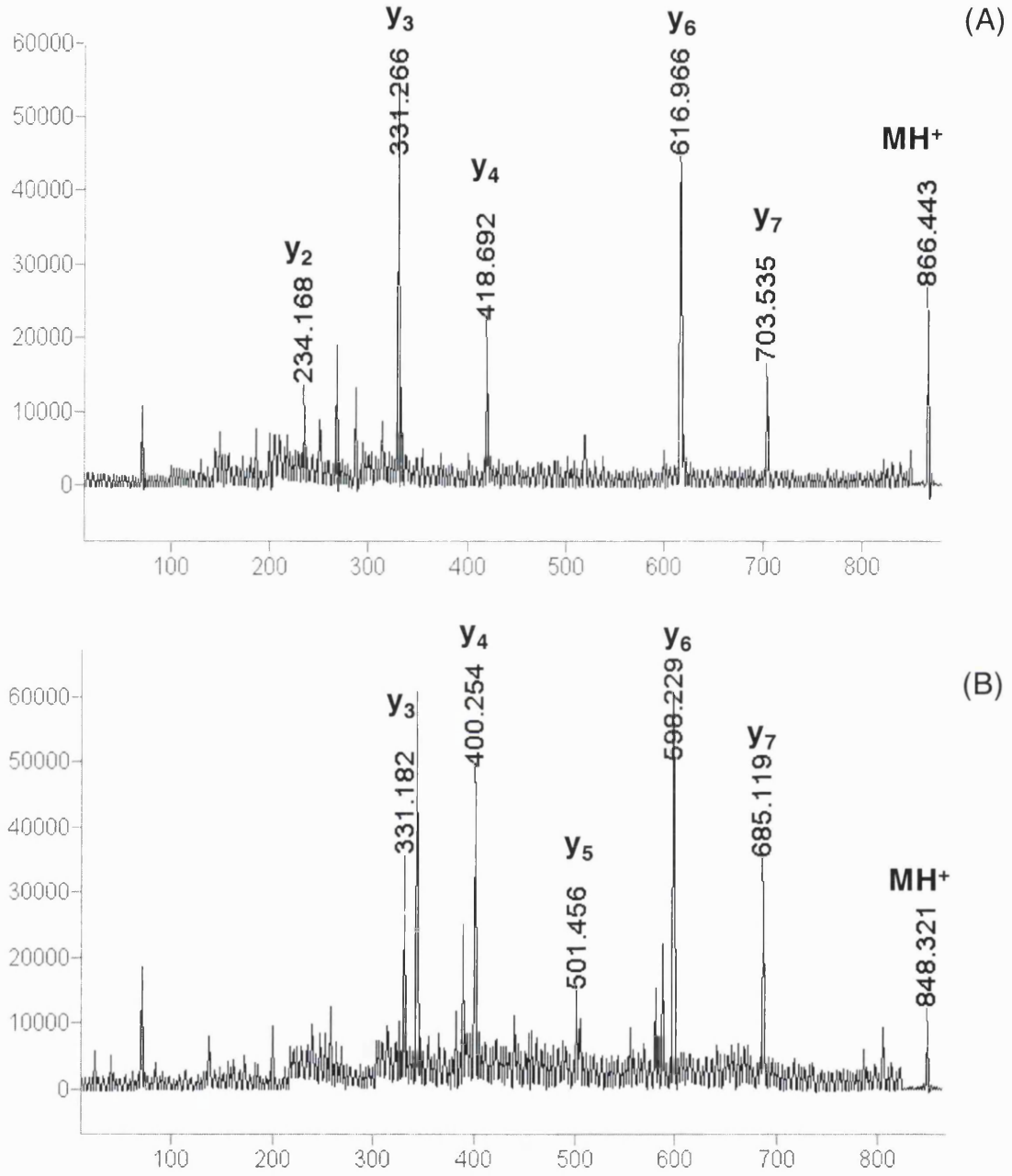


Figure 3.14: MALDI-PSD spectra of the unmodified N-CTD peptide and the β-eliminated product of the G-CTD peptide.

The parent ions and fragment ions y4 -y7 all differ by 18 Da between (A) the N-CTD and (B) the β-eliminated product of G-CTD. However, the y3 ions are the same mass. Hence, the G-CTD derived peptide contains a dehydroalanine at residue serine 5, indicating this is the site of glycosylation.

3.3.4. Precursor Ion Scanning

The most common use for precursor ion scanning in proteomics is for locating phosphopeptides. However, precursor ion scanning for the HexNAc oxonium ion is also an established technique[66]. Unfortunately, there is an added complication as the HexNAc oxonium ion is not as stable as the PO_3^- ion, and readily loses water groups to form ions at m/z 186.08 and m/z 168.07. Hence, it can be difficult to determine the optimal amount of collision energy to employ for precursor ion scanning. Enough energy must be applied to fragment the peptide causing the loss of the sugar moiety, but too much energy will cause the fragmentation of the GlcNAc oxonium ion and leads to a loss of sensitivity in locating the glycopeptides. As larger peptides generally require more collision energy to fragment than short peptides, there is not a single collision energy that will be effective for locating all glycopeptides. Hence, a range of collision offsets will often be required.

Figure 3.15 shows a mass spectrum and a precursor ion spectrum of a digest of BSA spiked with G-CKII. A large number of peaks are observed in the survey spectrum, including two minor peaks corresponding to G-CKII at $[M + 2H]^{2+}$ m/z 719.77, $[M + H + Na]^{2+}$ m/z 730.76. However, when a precursor ion scan for peptides that fragment to produce the GlcNAc oxonium ion at m/z 204.09 was performed, there were only two peaks in the spectrum (Figure 3.15B). These correspond to G-CKII at m/z 720.25 and G-CKII containing an oxidised methionine residue at m/z 729.04. The precursor ion spectrum was produced from the quadrupole scan, which was scanning over a 500 Da mass range in five seconds. Hence, the resolution is low, and is unable to resolve the isotopes of these doubly charged species and the peak labels are average masses, rather than monoisotopic masses.

Precursor ion scanning can also be performed using MALDI-CID-MS-MS. However, this technique is generally of little use for locating GlcNAc-modified peptides. This is due to the fragmentation pattern of singly charged glycopeptides. Nearly all ions produced by MALDI-MS carry a single charge. In MALDI-CID-MS-MS spectra of singly charged GlcNAc-modified peptides, the characteristic GlcNAc oxonium ion is not observed, as the charge is always retained on the peptide fragment ions. However,

MALDI-CID-MS-MS of doubly charged glycosylated ions does produce the ion at m/z 204.09, as demonstrated below.

G-Myc is atypical to peptides that are normally analysed by fragmentation analysis in that it has four basic residues; two at the N-terminus and two at the C-terminus. This allows it to carry higher charge states than is typical of a peptide of its size (by ESI-MS it is observed at 4+ and 5+ charge states). It was observed as singly and doubly charged ions at comparable intensities when analysed by MALDI-MS on the QSTAR (not shown). MALDI-MS on the QSTAR is performed at close to atmospheric pressure, whereas traditional MALDI-TOF instruments ionise in a higher vacuum environment. The higher pressure provides collisional cooling during the ionisation process, leading to softer ionisation. Hence, multiply charged ions have a higher chance of surviving the MALDI process in the QSTAR. Figure 3.16 shows the low-mass region of MALDI-CID-MS-MS spectra of the singly and doubly charged G-Myc peptide. The difference between the two spectra is the presence of ions at m/z 204.09, m/z 186.07, m/z 138.06 and m/z 126.05 in the spectra derived from the doubly charged species. These peaks correspond to the GlcNAc oxonium ion and its fragments (Appendix 1).

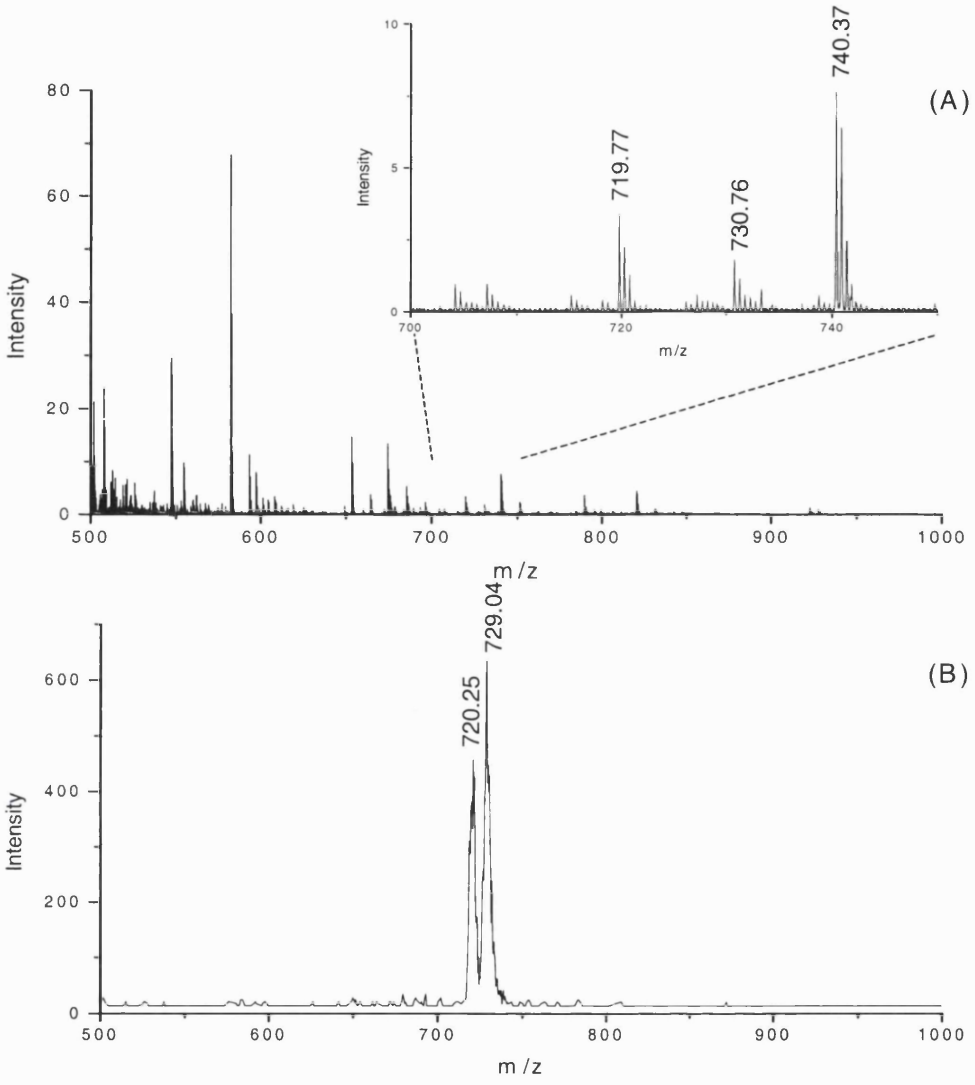


Figure 3.15: Precursor ion scanning locates GlcNAc-modified peptides in a complex mixture.

A tryptic digest of bovine serum albumin was spiked with the G-CKII peptide and analysed by nanospray-MS. The mass spectrum of the mixture is (A), and the magnified region (inset) contains the G-CKII peptide peaks $[M + H]^+$ m/z 719.77, $[M + Na]^+$ m/z 730.76, and the BSA derived tryptic peptide peak of residues 421-433 at m/z 740.37. A precursor ion scan for parents of m/z 204.1 (B) produces two peaks: a peak at m/z 720.25 corresponding to the G-CKII peptide, and a peak at m/z 729.04 formed by a methionine oxidised G-CKII peptide.

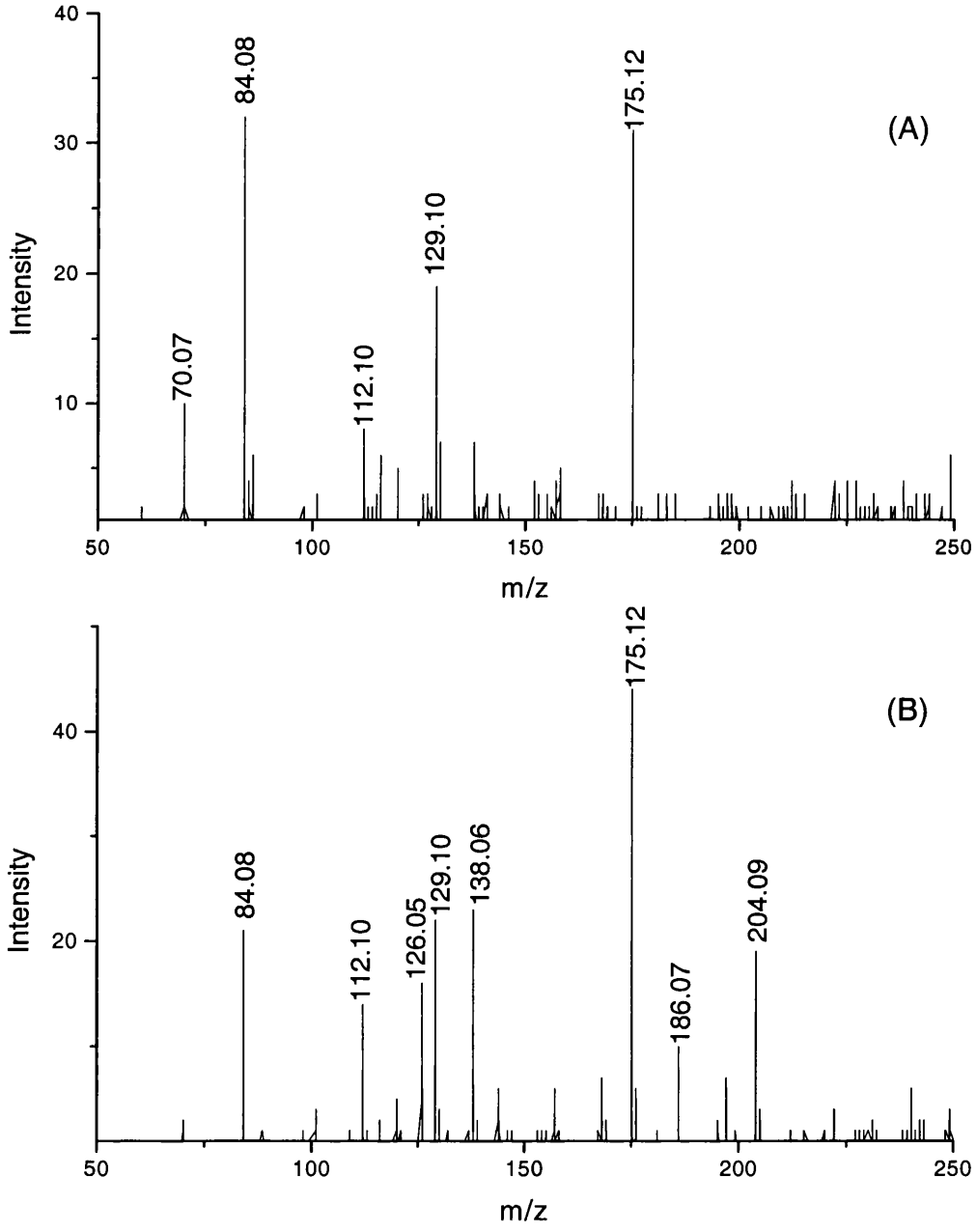


Figure 3.16: Low mass region of MALDI-CID-MS-MS spectra of singly charged and doubly charged GlcNAc-modified G-Myc peptide.

The CID-MS-MS spectrum of the (B) doubly charged version of the peptide contains the characteristic GlcNAc oxonium ion at m/z 204.09 and its fragment ions, but these are absent from the spectrum of the (A) singly charged parent ion.

3.4. Discussion

The use of the 'hot' matrix CHCA to analyse GlcNAc-modified peptides caused a significant loss of the GlcNAc residue during ionisation. This cleavage will compromise the sensitivity for detecting modified peptides and make estimations of stoichiometry of modification inaccurate. DHB was a better choice of matrix, but still caused limited sugar loss during ionisation. Electrospray ionisation provided the softest ionisation, and appeared to be the most appropriate ionisation method for observing GlcNAc-modified peptides.

MALDI-PSD was unable to identify sites of GlcNAc modification. In the PSD spectrum of G-CKII, which contained a relatively high level of chemical noise, only one glycosylated fragment ion was observed, which was sufficient to discount one possible site of modification, but could not differentiate between two serine residues that could be modified. The spectrum acquired by MALDI low energy CID on the QSTAR produced a similar spectrum to that produced by PSD. Despite the lower chemical noise leading to the observation of several more low intensity fragment ion peaks, only two glycosylated fragment ions were observed, and these provided no further information on the site of modification. This showed that the high level of chemical noise in the MALDI-PSD spectrum was not the main factor preventing GlcNAcylation site identification; it was the fragmentation mechanism itself. These results are consistent with previously published analysis of O-glycosylation using PSD[63] where very few glycosylated fragment ions could be observed, although in this case the site could be determined from the glycosylated ions observed.

Following on from the demonstration that O-glycosylation modification sites can be identified using QTOF mass spectrometry [64], ESI-CID-MS-MS of the G-CKII peptide on a QTOF instrument was able to unequivocally determine the site of GlcNAc modification. Many glycosylated fragment ions were observed, and overlapping GlcNAc-modified 'b' and 'y' ion series were able to assign the modification to serine 8. Identifying the site of modification of the G-CTD peptide was less straightforward. The G-CTD peptide has a basic C-terminal residue, which caused the observation of mainly C-terminally derived fragment ions, with only one 'b' ion. Hence, overlapping 'b' and

'y' ion series could not be obtained. However, for this peptide it was possible to produce glycosylated 'y' ions of a higher intensity than their 'deglycosylated' counterparts. This allowed the lack of a glycosylated y3 ion, when an unmodified y3 ion and glycosylated y4 ion were observed to assign the modification to serine 5. However, for larger, more stable peptides such as G-CKII, it will not be possible to achieve the higher intensity glycosylated fragment ions. In these cases a lack of glycosylated ion would not provide valid evidence for assigning a modification site. Hence, in order to identify sites of GlcNAc modification it will generally be necessary to produce overlapping 'b' and 'y' ion series. Thus, for identifying GlcNAc modification sites, trypsin or lysine endopeptidase (Lys-C), which both produce peptides containing basic C-terminal residues, will generally not be appropriate choices of enzyme for protein digestion to analyse glycopeptides.

Previous efforts to identify sites of GlcNAc modification using ESI-CID-MS-MS on triple quadrupole instruments were unsuccessful [157, 169]. The results in this chapter using ESI-CID-MS-MS on a QTOF instrument demonstrate that a key factor in the ability to identify GlcNAcmodification sites is the high dynamic range and low chemical noise of oa-TOF MS-MS spectra.

The addition of the extra galactose residue to create the GlcNAc-Gal disaccharide did not have a significant effect on the stability of the GlcNAc-peptide glycosidic link. Thus, it did not affect the ability to identify glycosylation sites. It caused the formation of additional glycosylated fragment ions, with a GlcNAc- and a GlcNAc-Gal-modified version of each fragment ion, increasing the complexity of the spectrum. The extra ions provided further confirmation of the assignment of glycosylated peaks, but will have a negative impact on the sensitivity of detection by distributing the fragment ions formed due to a given peptide backbone cleavage amongst three fragment ions instead of two, thus reducing the intensity of a given ion.

The production of the GlcNAc oxonium ion and related fragment ions from the doubly charged G-Myc peptide demonstrated that the charge state of the parent ion is at least as important in the determination of the fragmentation pattern than the ionisation method employed. A restricting factor in the use of MALDI for peptide glycosylation analysis is the lack of formation of doubly charged ions by MALDI. The G-Myc peptide

is unusual in its formation of an intense doubly charged ion, due to it containing four basic residues. The formation of multiply-charged MALDI ions in the QSTAR is facilitated by the ionisation process occurring at much higher pressure than in conventional axial MALDI-TOF instruments. This causes collisional cooling, making it a softer ionisation method and allowing more multiply charged ions to survive. Despite this, multiply-charged MALDI ions are unusual, and when observed are generally of low intensity. Nevertheless, the observation of the GlcNAc oxonium ions suggests that if multiply charged ions could be consistently formed by MALDI, precursor ion scanning for glycopeptides could be employed, and it would make it a more useful ionisation method for glycosylation analysis.

The ESI-CID-MS-MS fragmentation of the negatively charged G-CTD peptide produced a fragmentation spectrum for which the majority of the fragmentation ions observed could not be assigned. Hence, it was difficult to assess the relative lability of the GlcNAc-peptide glycosidic bond. A fragment formed by the loss of GlcNAc and the hydroxyl group from the side chain of the glycoserine residue was observed and a GlcNAc oxonium ion was also detected, but neither of these were intense fragment ions. However, as little peptide sequence could be determined from the fragmentation spectrum, there was no information accrued on the site of modification. The fragmentation pattern of peptides in negative ion spectra makes it an unattractive approach to determining GlcNAc modification sites. Even if the spectra could be interpreted, it is likely to be significantly less sensitive due to the inherent lower sensitivity of mass spectrometry in negative ion mode.

Precursor ion scanning is an effective method for detecting GlcNAc modified peptides in a complex mixture [66, 169]. In this chapter the G-CKII peptide was specifically detected in a spiked BSA digest. Precursor ion scanning also detected an oxidised methionine version of the G-CKII peptide that was not visible in the survey mass spectrum. In fact, this oxidised version of the peptide gave a larger peak than the unmodified version. Also there was no peak in the precursor ion scan corresponding to the sodiated version of G-CKII, which was observed in the TOF-MS spectrum at m/z 730.76. These results are likely to be due to the effect these modifications have on the stability of the peptide. The sodium adduct versions of peptides are generally more stable

than protonated peptides, and require more collision energy to fragment[63]. Hence, at the collision energy used in this experiment the sodiated peptide may not have fragmented. Conversely, methionine oxidised peptides more readily fragment, most notably by losing the oxidised methionine side chain[172], and this would explain why proportionately more GlcNAc oxonium ion was produced from this peptide at a low collision energy.

The fragility of the GlcNAc oxonium ion itself means that care must be taken to choose a collision energy at which the GlcNAc oxonium ion is formed from GlcNAc-modified peptides, but there is minimal fragmentation of the GlcNAc oxonium ion to form the fragment ions presented in Appendix 1. Combining precursor ion scans for masses m/z 204.1, 186.1 and 168.1 would allow compensation for using a higher collision energy than ideal for observing the GlcNAc oxonium ion, and this would reduce the requirement for optimising collision offsets for each sample. However, scanning multiple masses will increase the number of false positive results from fragments at these masses that are not glycosylation-related.

β-Elimination followed by mass spectrometry is a useful alternative method for determining sites of GlcNAc modification [157, 158]. For the serine modified G-CTD peptide, the site of modification was confirmed from a PSD spectrum of the β-eliminated product. However, using the same conditions there was no eliminated product formed from the threonine-modified G-Myc peptide. The protocol applied used ammonium hydroxide to effect the β-elimination, as this is volatile, so is easily removed after the reaction. It may be necessary to use harsher, more basic conditions in order to eliminate efficiently O-GlcNAc moietites from some modified threonine residues. Yet, one would have to be careful to minimise degradation of the peptide under these conditions. Even so, this was a less sensitive method of site determination than using ESI-CID-MS-MS on a QTOF instrument.

4. ALPHA CRYSTALLIN

Having established that sites of O-GlcNAc modification can be determined using CID fragmentation on a Q-TOF instrument (Chapter 3), these techniques were then applied to study a known site of GlcNAc modification in a previously characterised GlcNAc-modified protein.

4.1. Introduction

α-Crystallin is a protein composed of two polypeptide chains, αA-crystallin and αB-crystallin. They are members of the small heat shock protein (sHSP) family of 20 – 40 kDa proteins that contain a conserved domain of 80 to 100 amino acids referred to as the α-crystallin core[173]. α-Crystallins make up over 30% of the total protein content in the lens of the eye and are responsible for the eye's optical properties. However, the expression of crystallins is not restricted to the eye; they are expressed in many non-ocular tissues.

In the lens, crystallins combine to form large non-covalent complexes of 300 - 1000 kDa[174]. Elsewhere, in response to stress, they combine with other small heat shock proteins to form heat shock granules[175]. Mutations in α-crystallin have been linked to cataract formation in the eye[176], and sHSP over-expression has been linked to a number of diseases[173].

As well as being phosphoproteins, both chains have been reported to be O-GlcNAc-modified in a wide variety of species[74]. The site of modification of the A chain was determined to be serine 162[74], and the B chain is modified on threonine 170[75]. In both cases, the site was identified using enzymatic radiolabelling with [^{3}H]galactose, followed by multiple rounds of HPLC purification and screening fractions for radioactivity. The peptide was analysed by mass spectrometry to obtain a molecular weight, from which the peptide could be identified from a theoretical digest of the protein. Finally sites were identified using Edman degradation analysis by monitoring for the release of the radioactivity [74, 75].

Recently, it was investigated whether GlcNAc-modified peptides could be located in a tryptic digest of α-crystallins in an automated manner using LC-ESI-MS precursor ion

scanning on a triple quadrupole instrument[169]. The protein was labelled *in vitro* with galactose, and precursor ion scanning for the GlcNAc-Gal disaccharide oxonium ion was carried out. It was found that the stoichiometry of modification was too low in the crude tryptic digest to detect modified peptides from either chain by precursor ion scanning. However, after an affinity enrichment step using the lectin *Ricinnus communis* (RCA I), which binds terminal galactose residues, followed by LC-ESI-MS precursor ion scanning for the GlcNAc-Gal oxonium ion at m/z 366, Haynes *et al.* were able to locate the modified peptide from αA-crystallin [169]. This peak was automatically selected for CID-MS-MS analysis, and a fragmentation spectrum was acquired. However, they were unable to elicit any information on the site of modification, as the dynamic range of their triple quadrupole CID-MS-MS spectrum was not good enough to see low intensity glycosylated fragment ions. Also, they employed a collision energy optimised for peptide sequencing, which is likely to be higher than the optimal energy for observation of glycosylated fragment ions.

If this fragmentation spectrum was acquired on a Q-TOF instrument, the likelihood of being able to identify the site of modification would be increased (Section 3.2.3).

4.2. *Identification of a GlcNAc Modification Site from a Solution Digest*

α-Crystallin was digested with trypsin and a MALDI mass fingerprint was acquired (Figure 4.1). The peak list from this spectrum was submitted to a MS-FIT search[44], and the search result is shown in Table 4.1. High sequence coverages (90%) for both the A chain and B chain are observed. Unmodified versions of the peptides from each chain containing the previously identified GlcNAc modification sites are observed at m/z 1641.91 for the A chain and m/z 1822.08 and m/z 1950.14 for the B chain. However, peaks corresponding to GlcNAc-modified versions of peptides from either chain are not detected. This is probably due to the stoichiometry of modification being too low to detect them in the background of more abundant unmodified peptides.

This digest was then analysed by LC-ESI-MS. As peptides eluted into the mass spectrometer they were automatically selected for fragmentation analysis. These CID-MS-MS spectra were submitted for a MASCOT search[47]. Table 4.2 lists the MS-MS spectra whose identities were determined. A large number of CID-MS-MS spectra are matched to peptides from both A and B chains, as well as two peptides from βB-crystallin B1. Detecting βB-crystallin in this sample demonstrates the power of MS-MS sequence analysis at locating low abundance contaminating proteins whose presence would not be identified from mass fingerprint data. ESI-CID-MS-MS spectra are assigned to unmodified peptides of sequences containing the proposed sites of GlcNAc modification from both peptides. However, modified versions of these peptides were not selected for MS-MS analysis.

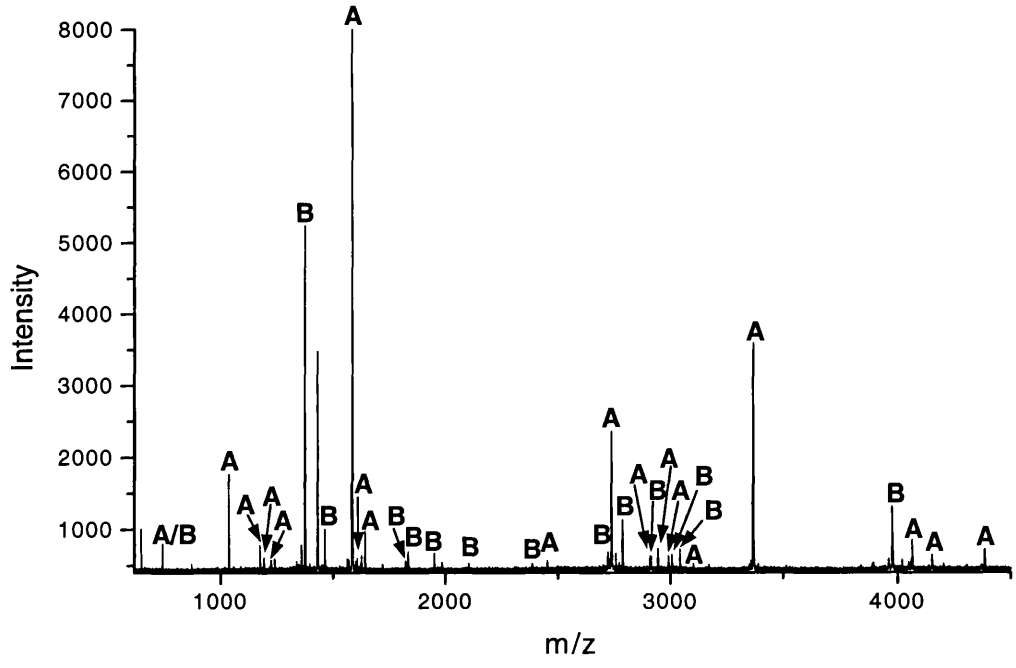


Figure 4.1: MALDI mass spectrum of a solution digest of α-crystallin.

Peak labels correspond to masses matched to (A) αA-crystallin and (B) αB-crystallin in the MS-FIT search of this spectrum (Table 4.1).

Table 4.1: MS-FIT search result from a MALDI mass spectrum of a solution tryptic digest of α-crystallin.

1. 19/46 matches (41%). 19790.2 Da, pI = 5.78. Acc. # P02470. BOVIN. ALPHA CRYSTALLIN A CHAIN.

m/z submitted	MH^+ matched	Delta ppm	start	end	Peptide Sequence (Click for Fragment Ions)
744.4176	744.3905	36.4253	113	117	(R)EFHRR(Y)
1037.5754	1037.5420	32.2071	13	21	(R)TLGPFYPSR(L)
1175.6650	1175.6272	32.1879	55	65	(R)TVLDSGISEVR(S)
1193.6826	1193.6431	33.0594	12	21	(K)RTLGPFYPSR(L)
1224.6106	1224.5973	10.8812	146	157	(K)IPSGVDAGHSER(A)
1581.8710	1581.8641	4.3541	66	78	(R)SDRDKFVIFLDVK(H)
1626.7312	1626.7373	-3.7484	100	112	(K)HNERQDDHGYISR(E)
1641.9083	1641.8448	38.6617	158	173	(R)AIPVSREEKPSSAPSS(-)
2454.4174	2454.2305	76.1452	79	99	(K)HFSPEDLTVKVQEDFVEIHGK(H)
2735.4515	2735.4408	3.8823	66	88	(R)SDRDKFVIFLDVKHFSPEDLTVK(V)
2738.4644	2738.4729	-3.1078	55	78	(R)TVLDSGISEVRSDRDKFVIFLDVK(H)
2908.3148	2908.3727	-19.8830	89	112	(K)VQEDFVEIHGKHNERQDDHGYISR(E)
2944.4438	2944.4185	8.6144	118	145	(R)YRLPSNVDQSALSCSLSADGMLTFSGPK(I)
2990.5509	2990.4761	25.0209	79	103	(K)HFSPEDLTVKVQEDFVEIHGKHNER(Q)
3100.6219	3100.5196	33.0110	117	145	(R)RYRLPSNVDQSALSCSLSADGMLTFSGPK(I)
3364.6414	3364.6458	-1.3019	22	49	(R)LFDQFFGEGLFEYDLLPFLSSTISPYYR(Q)
4061.7703	4061.9494	-44.1062	79	112	(K)HFSPEDLTVKVQEDFVEIHGKHNERQDDHGYISR(E)
4149.8787	4149.9974	-28.5937	118	157	(R)YRLPSNVDQSALSCSLSADGMLTFSGPKIPSGVDAGHSER(A)
4383.1419	4383.1694	-6.2807	13	49	(R)TLGPFYPSRLFDQFFGEGLFEYDLLPFLSSTISPYYR(Q)

27 unmatched masses: 650.4687 872.6019 1374.7233 1427.6623 1430.7655 1462.7196 1562.8850 1569.1100 1583.8527 1822.0753 1832.8949 1950.1384 1984.9289 2104.0630 2387.3136 2722.6279 2757.5451 2771.4463 2786.4239 2913.5705 3006.5519 3041.5229 3891.8535 3958.3822 3973.8762 4016.8435 4200.8093

The matched peptides cover **90%** (156/173 AA's) of the protein.

2. 15/46 matches (32%). 20036.9 Da, pI = 6.76. Acc. # P02510. BOVIN. ALPHA CRYSTALLIN B CHAIN (ALPHA(B)-CRYSTALLIN).

m/z submitted	MH^+ matched	Delta ppm	start	end	Peptide Sequence (Click for Fragment Ions)
744.4176	744.3640	72.0134	151	157	(K)QASGPER(T)
1374.7233	1374.7071	11.7930	12	22	(R)RPFFPFHSPSR(L)
1462.7196	1462.7000	13.3510	57	69	(R)APSWIDTGLSEMR(L)
1562.8850	1562.8542	19.7095	70	82	(R)LEKDRFSVNLDVK(H)
1822.0753	1822.0438	17.2506	158	174	(R)TIPITREEKPAVTAAPK(K)
1832.8949	1832.9217	-14.6087	57	72	(R)APSWIDTGLSEMRLEK(D)
1950.1384	1950.1388	-0.1850	158	175	(R)TIPITREEKPAVTAAPKK(-)
2104.0630	2104.0497	6.3108	57	74	(R)APSWIDTGLSEMRLEKDR(F)
2387.3136	2387.2723	17.2929	73	92	(K)DRFSVNLDVKHFSPEELKVK(V)
2722.6279	2722.4627	60.6848	124	150	(R)IPADVDPLAITSSLSSDGVLTVNGPRK(Q)
2786.4239	2786.3974	9.4953	93	116	(K)VLGDVIEVHGKHEERQDEHGFISR(E)
2913.5705	2913.5322	13.1584	122	149	(K)YRIPADVDPLAITSSLSSDGVLTVNGPR(K)
3006.5519	3006.5359	5.3238	57	82	(R)APSWIDTGLSEMRLEKDRFSVNLDVK(H)
3041.5229	3041.6271	-34.2657	121	149	(R)KYRIPADVDPLAITSSLSSDGVLTVNGPR(K)
3041.5229	3041.6271	-34.2657	122	150	(K)YRIPADVDPLAITSSLSSDGVLTVNGPRK(Q)
3973.8762	3974.0169	-35.4077	23	56	(R)LFDQFFGEHLLESDLFPASTSLSPFYLRPPSFLR(A)

31 unmatched masses: 650.4687 872.6019 1037.5754 1175.6650 1193.6826 1224.6106 1427.6623 1430.7655 1569.1100 1581.8710 1583.8527 1626.7312 1641.9083 1984.9289 2454.4174 2735.4515 2738.4644 2757.5451 2771.4463 2908.3148 2944.4438 2990.5509 3100.6219 3364.6414 3891.8535 3958.3822 4016.8435 4061.7703 4149.8787 4200.8093 4383.1419

The matched peptides cover **90%** (159/175 AA's) of the protein.

Figure 4.2 is a combined mass spectrum of several scans acquired during this run. Peaks are observed that correspond to the unmodified peptide spanning residues 158 - 173 of αA-crystallin at $[M + 3H]^{3+}$ m/z 547.93, $[M + 2H]^{2+}$ m/z 821.39. Eluting at the same time is a low intensity peak corresponding to the GlcNAc-modified version of this peptide at $[M + 3H]^{3+}$ m/z 615.62. Thus, the modified peptide was detected, but was not of sufficient intensity to be automatically selected for fragmentation analysis. This spectrum is the combined spectrum over the period during which the glycosylated and unmodified peptides eluted and illustrates the low stoichiometry of modification. Extracted ion chromatograms of these two peaks (Figure 4.3) show the GlcNAc-modified peptide eluted fractionally before the unmodified, which is what is expected, as the sugar moiety will make the peptide more hydrophilic. However, both peaks eluted at the very start of the run, as the peptide itself is already very hydrophilic.

Table 4.2: List of ESI-CID-MS-MS spectra assigned to proteins by MASCOT from the solution tryptic digest of α-crystallin.

Alpha A Crystallin

Observed	Mr (expt)	Mr (calc)	Delta	Miss	Peptide	Mod
490.78	979.54	979.57	-0.04	0	FVIFLDVK	
519.23	1036.44	1036.53	-0.10	0	TLGPFYPSR	
586.78	1171.55	1171.59	-0.04	0	HFSPEDLTVK	
588.28	1174.55	1174.62	-0.07	0	TVLDSGISEVR	
612.81	1223.60	1223.59	0.01	0	IPSGVDAGHSER	
650.82	1299.61	1299.64	-0.04	0	VQEDFVEIHGK	
714.33	1426.65	1426.71	-0.05	0	MDIAIQHPWFK	N-Acetyl
722.33	1442.65	1440.70	-0.05	0	MDIAIQHPWFK	N-Acetyl Met-Ox
767.38	1532.74	1532.78	-0.04	1	TVLDSGISEVRSDR	
527.94	1580.80	1580.96	-0.06	2	SDRDKFVIFLDVK	
792.38	1582.75	1582.81	-0.06	1	MDIAIQHPWFKR	N-Acetyl
533.93	1598.77	1598.80	-0.03	1	MDIAIQHPWFKR	N-Acetyl Met-Ox
542.93	1625.76	1625.73	0.03	1	HNERQDDHGYISR	
821.42	1640.83	1640.84	-0.01	1	AIPVSREEKPSSAPSS	
592.96	1775.86	1775.90	-0.04	2	TVLDSGISEVRSDRDK	
982.11	2943.30	2943.41	-0.11	1	YRLPSNVDQSALSCSLSADGMLTFSGPK	

Alpha B Crystallin

Observed	Mr (expt)	Mr (calc)	Delta	Miss	Peptide	Mod
461.25	920.48	920.50	-0.01	0	FSVNLDVK	
493.75	985.48	985.49	-0.01	0	HFSPEELK	
607.33	1212.64	1212.65	-0.01	1	HFSPEELKVK	
458.90	1373.67	1373.70	-0.03	0	RPFFPFHSPSR	
715.85	1429.69	1429.73	-0.04	0	MDIAIHHPWIR	N-Acetyl
482.91	1445.69	1445.72	-0.03	0	MDIAIHHPWIR	N-Acetyl Met-Ox
731.83	1461.65	1461.69	-0.05	0	APSWIDTGLSEMR	
521.61	1561.82	1561.85	-0.03	2	LEKDRFSVNLDVK	
608.02	1821.03	1821.04	-0.01	1	TIPITREEKPAVTAAPK	
611.63	1831.86	1831.91	-0.06	1	APSWIDTGLSEMRLEK	
865.43	2593.28	2593.36	-0.08	0	IPADVDPLAITSSLSSDGVLTVNGPR	
971.81	2912.41	2912.52	-0.11	1	YRIPADVDPLAITSSLSSDGVLTVNGPR	

Beta B Crystallin B1

Observed	Mr (expt)	Mr (calc)	Delta	Miss	Peptide	Mod
733.36	1464.70	1464.74	-0.03	0	LVVFEQENFQGR	
860.39	1718.76	1718.81	-0.04	0	VSSGTWVGYQYPGYR	

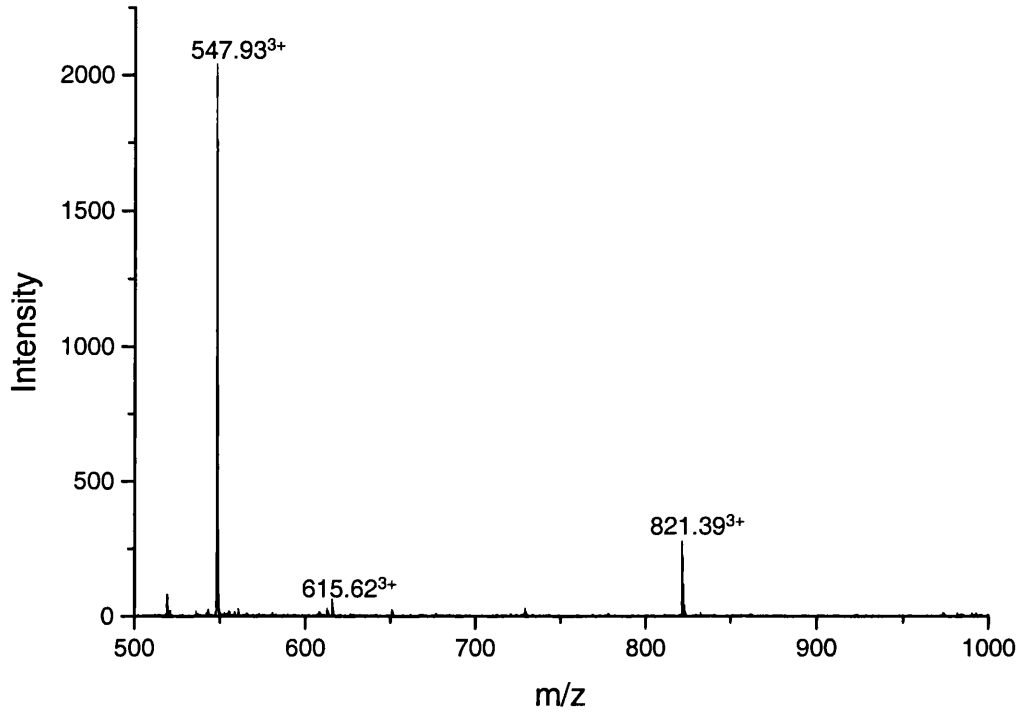


Figure 4.2: Combined mass spectrum over the period during which the triply-charged m/z 547.93 and m/z 615.62 eluted.

Peaks observed correspond to unmodified doubly and triply charged versions of the peptide residues 158 - 173 of αA-crystallin at $[M + 3H]^{3+}$ m/z 547.93 and $[M + 2H]^{2+}$ m/z 821.39, and the triply charged GlcNAc-modified peptide at $[M + 3H]^{3+}$ m/z 615.62. This is the combined spectrum of the period during which the GlcNAc-modified and unmodified peptides eluted.

Figure 4.4 is the combined mass spectrum of the time during which the potential GlcNAc-modified peptide, corresponding to residues 158 - 175 from the B chain, eluted. The unmodified peptide is observed at $[M + 4H]^{4+}$ m/z 488.30. A quadruply-charged glycosylated peptide would have been observed at m/z 539.05, and this region is magnified in the inset to Figure 4.4. A peak for the GlcNAc-modified peptide is not observed above the background chemical noise, although the peaks at m/z 539.30 and m/z 539.57 are the correct masses for the second and third isotopes of this peptide. The

extracted ion chromatogram for a peak at this mass rises above the baseline around the time the peak would be expected to have eluted (Figure 4.4B), but this data is not sufficient to confirm the presence of the modified peptide. However, another peak corresponding to the peptide residues 158 - 174 was also observed at $[M + 3H]^{3+}$ m/z 608.04, and this is present in Figure 4.5. The GlcNAc-modified version of this peptide would have appeared at $[M + 3H]^{3+}$ m/z 675.71, and although this is not immediately visible, in the magnified region inset into Figure 4.5A there is a triply-charged peak at m/z 675.69, which is probably the modified peptide. Also, the extracted ion chromatogram of the potential GlcNAc-modified peak shows a slight increase in the extracted ion chromatogram just before the unmodified peptide eluted.

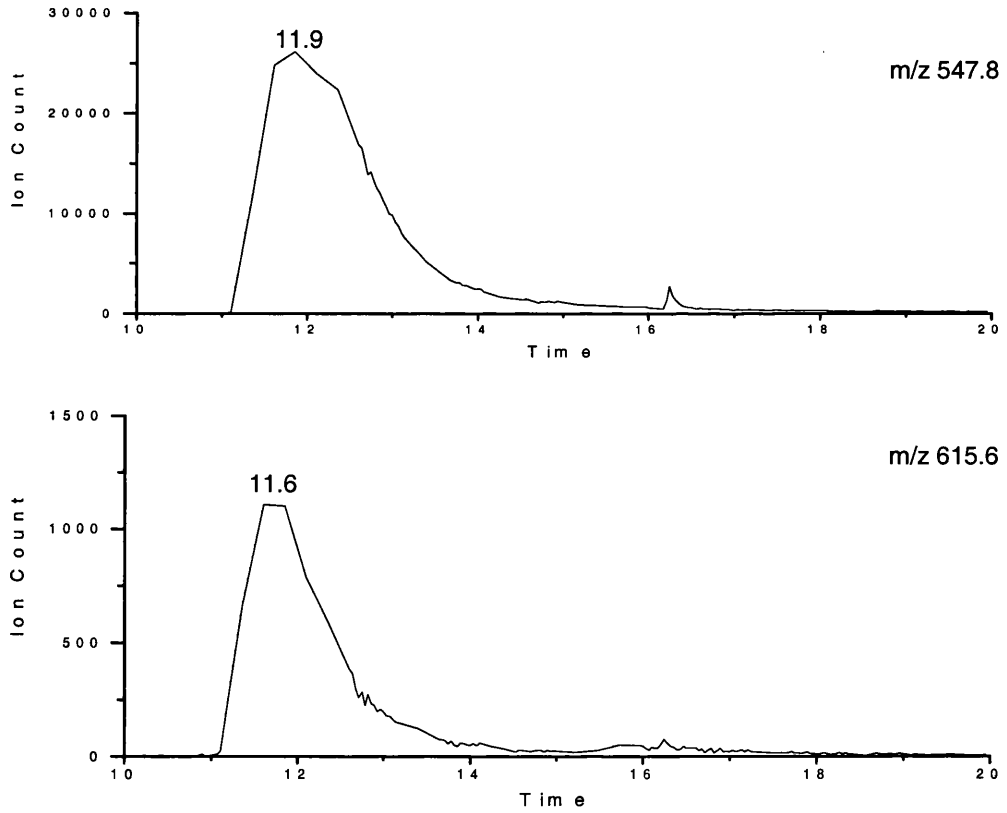


Figure 4.3: Extracted ion chromatograms of the unmodified and glycosylated peptide residues 158 - 173 from an LC-MS analysis of αA-crystallin.

(A) The unmodified peptide elutes marginally later than (B)the glycosylated peptide, showing the sugar moiety increases the hydrophilicity of the peptide.

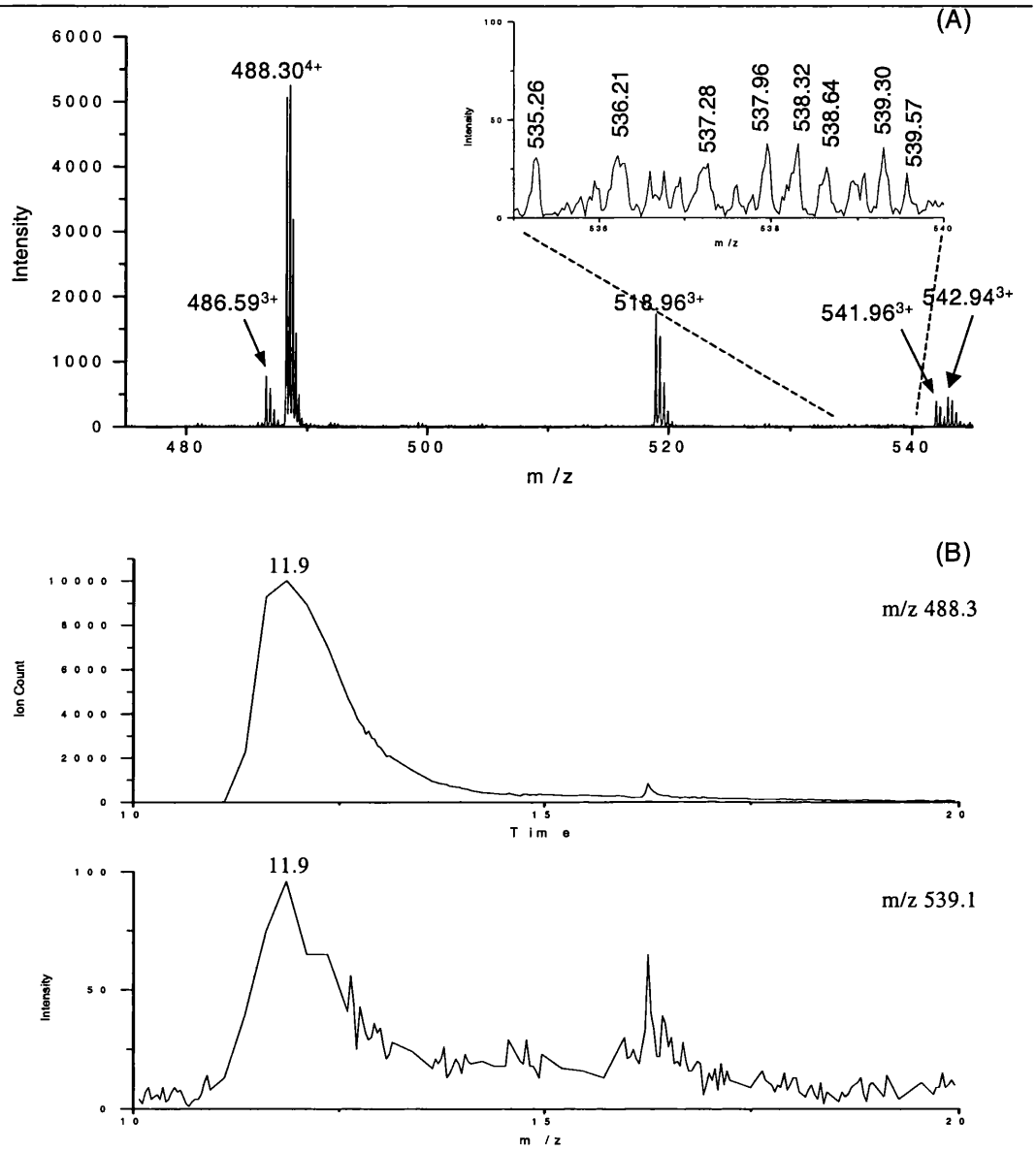


Figure 4.4: Combined mass spectrum over the period during which peaks are observed in the extracted ion chromatograms of m/z 488.3 and m/z 539.1.

(A) The peak observed at $[M + 4H]^{4+}$ m/z 488.30, corresponds to residues 158 - 175 of αB-crystallin. The inset shows the region that would contain the GlcNAc-modified version of this peptide at $[M + 4H]^{4+}$ m/z 539.05. (B) Extracted ion chromatograms corresponding to the unmodified (m/z 488.3) and GlcNAc-modified (m/z 539.1) peptides. The chromatogram at the mass of the GlcNAc-modified peptide contains a peak at the same retention time as the unmodified peptide elutes (11.9 minutes), suggesting there may be modified peptide present.

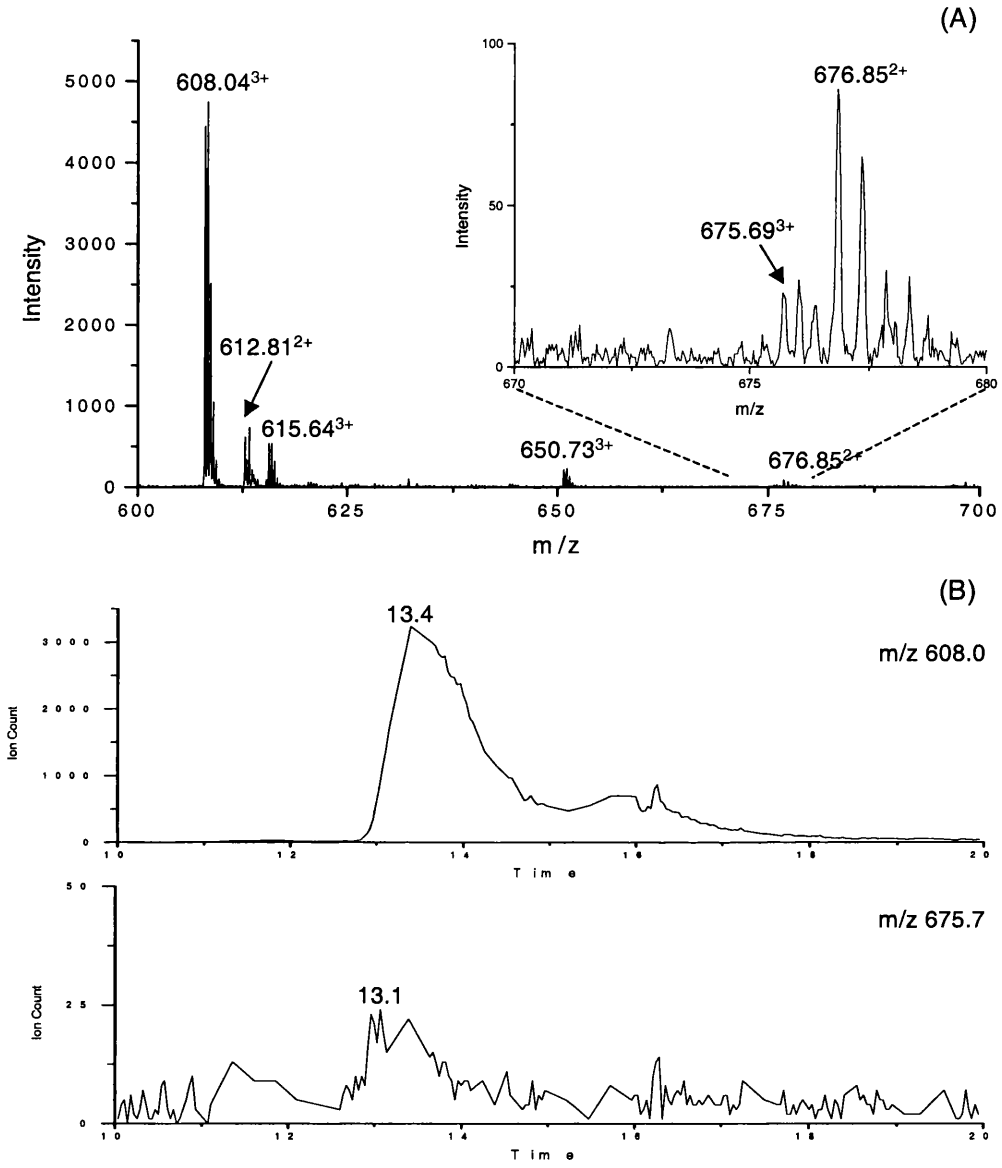


Figure 4.5: Combined mass spectrum over the period during which peaks are observed in the extracted ion chromatograms of m/z 608.3 and m/z 675.7.

(A) The peak at $[M + 3H]^{3+}$ m/z 608.04, corresponds to residues 158 - 174 of αB-crystallin. The magnified region contains a peak at $[M + 3H]^{3+}$ m/z 675.69, corresponding in mass to the GlcNAc-modified peptide 158 - 174. (B) Extracted ion chromatograms corresponding to the unmodified (m/z 608.0) and glycosylated (m/z 675.7) peptides. The extracted ion chromatogram of the GlcNAc-modified peptide contains a very weak peak eluting at 13.1 minutes, just before the unmodified peptide.

The fragmentation spectra of the doubly and triply charged species of the unmodified peptide residues 158 - 173 from αA-crystallin are shown in Figures 4.6 and 4.7 respectively, and the identities of the peaks observed are given in Tables 4.3 and 4.4. The doubly-charged peptide fragmented considerably less than the triply charged and the parent ion dominates the spectrum. The low intensity fragment ions include many 'b' and 'y' ions. The triply charged parent ion is not present in the CID spectrum acquired using the same collision offsets. This spectrum contains many more fragment ions than that of the doubly-charged species. Most notably, more internal fragment ions are present. Both of these spectra were acquired from four MS-MS scans: two each at collision offsets of 28 eV and 32 eV. These spectra show that the triply charged species is significantly more fragile, and can be fragmented at a lower collision energy than the doubly-charged version. As was demonstrated in Chapter 3, low collision offsets should be utilised for observing GlcNAc-modified fragment ions. Therefore, the triply charged version of the GlcNAc-modified peptide should give a better chance of observing glycosylated fragment ions than the doubly-charged species.

This sample was analysed again by LC-MS, but this time it was specified to select the triply charged GlcNAc-modified species from the A chain for CID-MS-MS analysis. As the fragmentation spectra of the unmodified triply charged peptide showed comprehensive fragmentation and no remaining parent ion at a collision energy of 28 eV, collision offsets of 22 eV and 25 eV were employed for analysis of the glycosylated species. The fragmentation spectrum of the triply-charged GlcNAc modified peptide is given in Figure 4.8, and the peak identities are explained in Table 4.5. The dominant ions in this fragmentation spectrum are the deglycosylated doubly-charged parent ion (m/z 821.42), doubly-charged y_{14} and y_{14G} (m/z 729.40 and m/z 830.95 respectively), the GlcNAc oxonium ion at m/z 204.08 and the a_2 and b_2 ions at m/z 157.13 and m/z 185.15. The spectrum, as a whole, contains an unusually large number of internal fragment ions. This is due to the presence of three proline residues in the sequence, and cleavage N-terminal to proline residues is highly favoured by CID[170]. However, there are also more high mass 'b' ions in this spectrum than in the triply-charged unmodified peptide spectrum (Figure 4.7) as a result of using lower collision offsets. There are a number of glycosylated fragment ions in the spectrum. However, the most informative ions for

determining the site of modification are at m/z 1029.50 and weaker ion at m/z 901.39. These correspond to glycosylated internal ions $PVSREEK_G$ and $PVSREE_G$. As there is only one possible residue in these fragments that can bear the glycosylation, serine 162 is unequivocally confirmed as the site of modification.

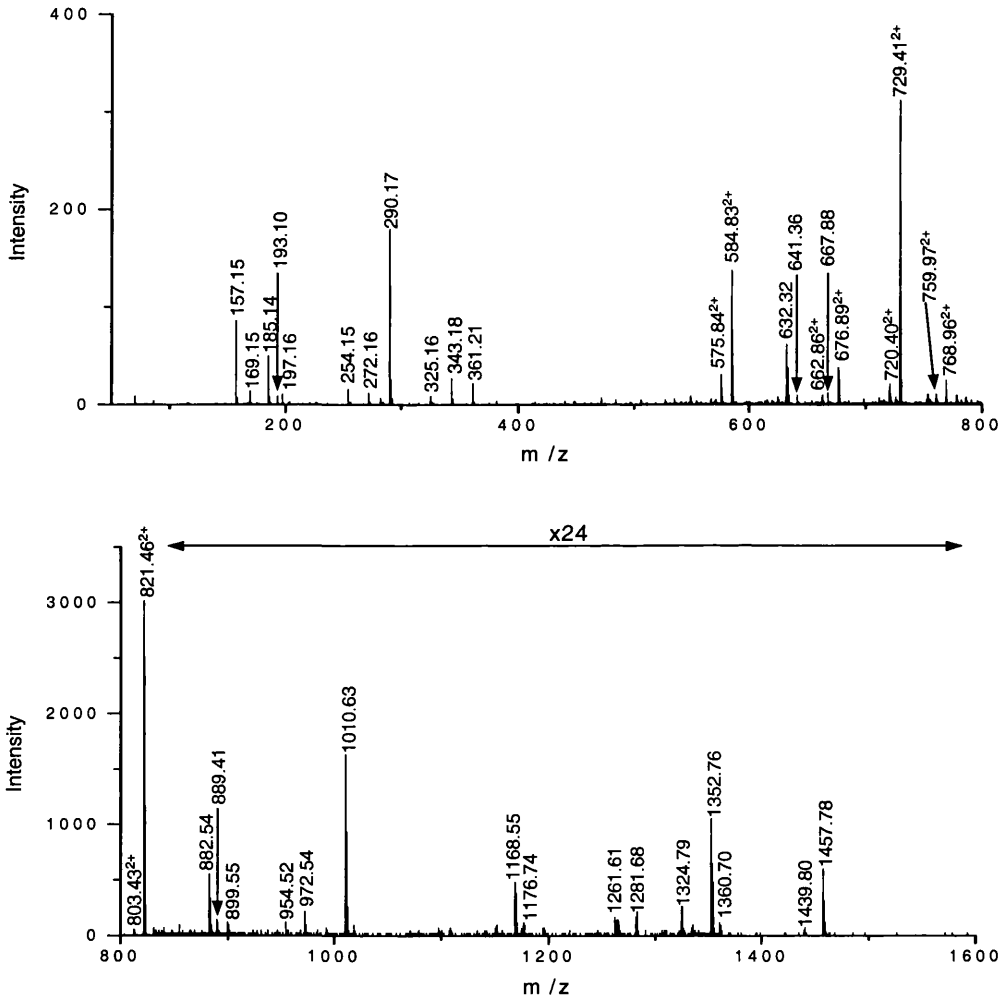


Figure 4.6: A CID-MS-MS spectrum of $[M + 2H]^{2+}$ m/z 821.39, corresponding to residues 158 - 173 of αA-crystallin.

The identities of peaks labelled in this spectrum are given in Table 4.3. This spectrum was acquired using the same collision offsets as the triply charged spectrum in Figure 4.7.

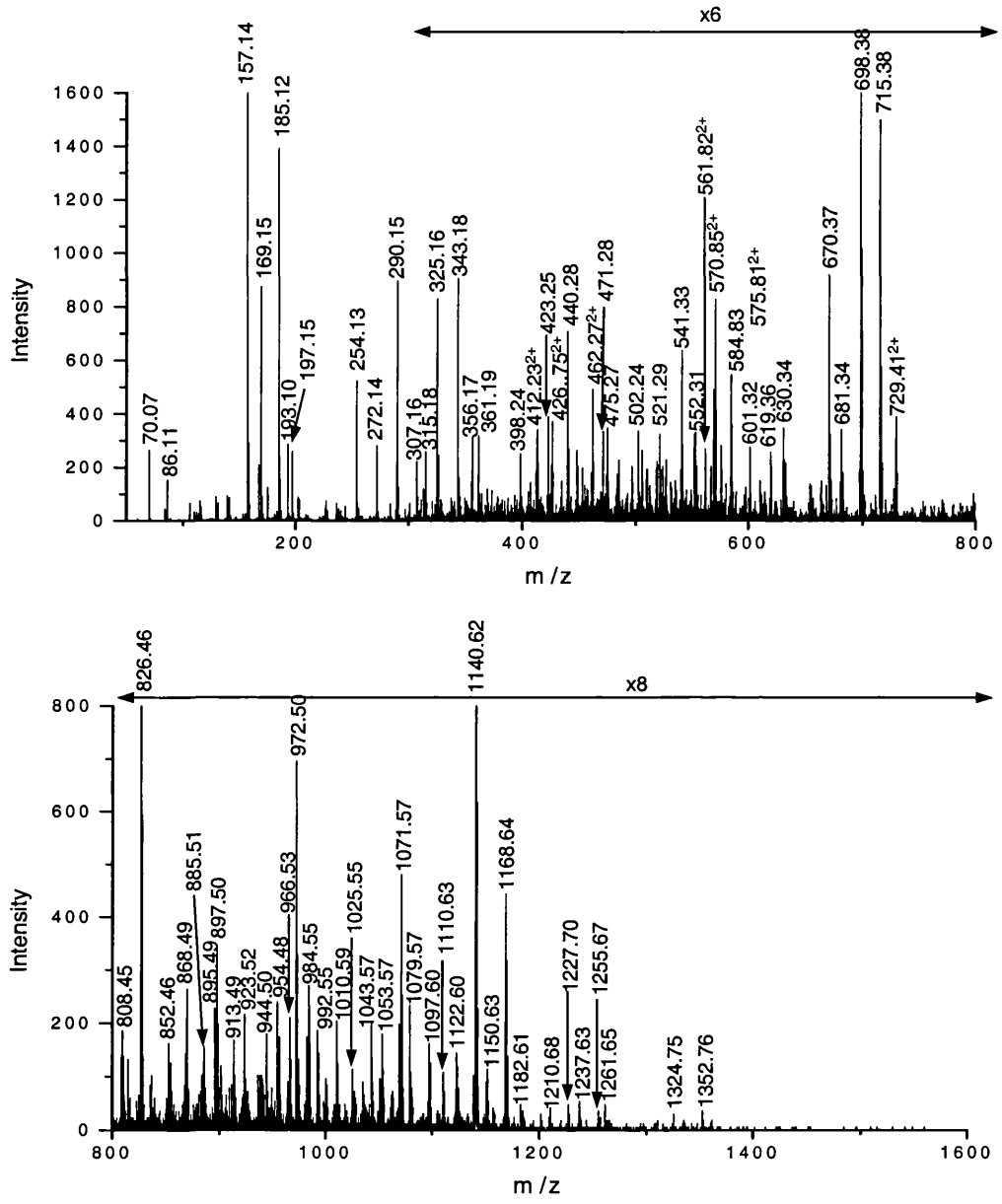
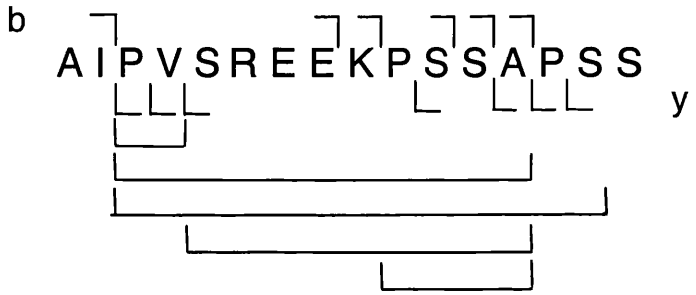


Figure 4.7: A CID-MS-MS spectrum of [M + 3H]$^{3+}$ m/z 547.93, corresponding to residues 158-173 of αA-crystallin.

The identities of peaks labelled in this spectrum are given in Table 4.4. This spectrum was acquired using the same collision offsets as for the doubly-charged spectrum in Figure 4.6.

Table 4.3: Fragment ions observed in the CID-MS-MS spectrum of $[M + 2H]^{2+}$ m/z 821.46.

The spectrum is shown in Figure 4.6, and is the unmodified peptide spanning residues 158 – 173 of αA-crystallin.



Peak	Match	Peak	Match	Peak	Match
157.15	a2 or PS-28	632.32	y6	899.55	
169.15	AP	641.36		954.52	SREEKPSSA-H_2O
185.14	b2 or PS	662.86^{2+}		972.54	SREEKPSSA
193.10	y2	667.88		1010.63	b9
197.16	PV	676.89^{2+}	b13	1168.55	PVSREEKPSSA
254.15	y3-$2H_2O$	720.40^{2+}	y14-H_2O	1176.74	b11-H_2O
272.16	y3-H_2O	729.41^{2+}	y14	1261.61	y12
290.17	y3	759.97^{2+}	b15-H_2O	1281.68	b12
325.16	y4-$2H_2O$	768.96^{2+}	b15	1324.79	PVSREEKPSSAPS-28 or a13
343.18	y4-H_2O or PSSA	803.43^{2+}	MH_2^{2+}-$2H_2O$	1352.76	PVSREEEKPSSAPS or b13
361.21	y4	821.46^{2+}	MH_2^{2+}	1360.70	y13
575.84^{2+}	PVSREEKPSSA-H_2O	882.54	b8	1439.80	y14-H_2O
584.83^{2+}	PVSREEKPSSA	889.41	y9	1457.78	y14

Table 4.4: Fragment ions observed in the CID-MS-MS spectrum of $[M + 3H]^{3+}$ m/z 547.93.

The spectrum is shown in Figure 4.7 and corresponds to the unmodified peptide spanning residues 158 – 173 of αA-crystallin.

Peak	Match	Peak	Match	Peak	Match
70.07	P	521.29		954.48	SREEKPSSA-H_2O
86.11	L/I	541.33	PVSRE-28	966.53	
157.14	a2 or PS-28	552.31	PVSRE-NH_3	972.50	SREEKPSSA
169.15	AP	561.82^{2+}	IPVSREEKPS	984.55	
185.12	b2 or PS	570.85^{2+}	PVSREEKPSSA-28	992.55	b9-H_2O
193.10	y2	575.81^{2+}	PVSREEKPSSA-H_2O	1010.59	b9
197.15	PV	584.83^{2+}	PVSREEKPSSA	1025.55	VSREEKPSSA-28-H_2O
254.13	y3-2H_2O	601.32	VSREE	1043.57	VSREEKPSSA-28
272.14	y3-H_2O	619.36		1053.57	VSREEKPSSA-H_2O
290.15	y3	630.34	SREEK	1071.57	VSREEKPSSA
307.16	PSSA-2H_2O	670.37	PVSREE-28	1079.57	PVSREEKPSS-H_2O
315.18	PSSA-28	681.34	PVSREE-NH_3	1097.60	PVSREEKPSS
325.16	y4-2H_2O or PSSA-H_2O	698.38	PVSREE	1110.63	
343.18	y4-H_2O or PSSA	715.38		1122.60	PVSREEKPSSA-28-H_2O
356.17	SRE-NH_3	729.41^{2+}	y14	1140.62	PVSREEKPSSA-28
361.19	y4	808.45	PVSREEK-H_2O	1150.63	PVSREEKPSSA-H_2O
398.24	REE-NH_3	826.46	PVSREEK	1168.64	PVRSEEKPSSA
412.23^{2+}		852.46		1182.61	IPVSREEKPSS-28
423.25	PVSR-NH_3	868.49	REEKPSSA-NH_3	1210.68	IPVSREEKPSS
426.75^{2+}		885.51	REEKPSSA	1227.70	VSREEKPSSAPS-28
440.28	PSSAP or PVSR	895.49	VSREEKPS-H_2O or PVSREEKP-28	1237.63	PVSREEKPSSAP-28
462.27^{2+}	PVSREEKP	897.50		1255.67	VSREEKPSSAPS
471.28	KPSSA	913.49	VSREEKPS	1261.65	y12
475.27		923.52	PVSREEKP	1324.75	PVSREEKPSSAPS-28 or a13
502.24	SREE	944.50	SREEKPSSA-28	1352.76	PVSREEKPSSAPS or b13

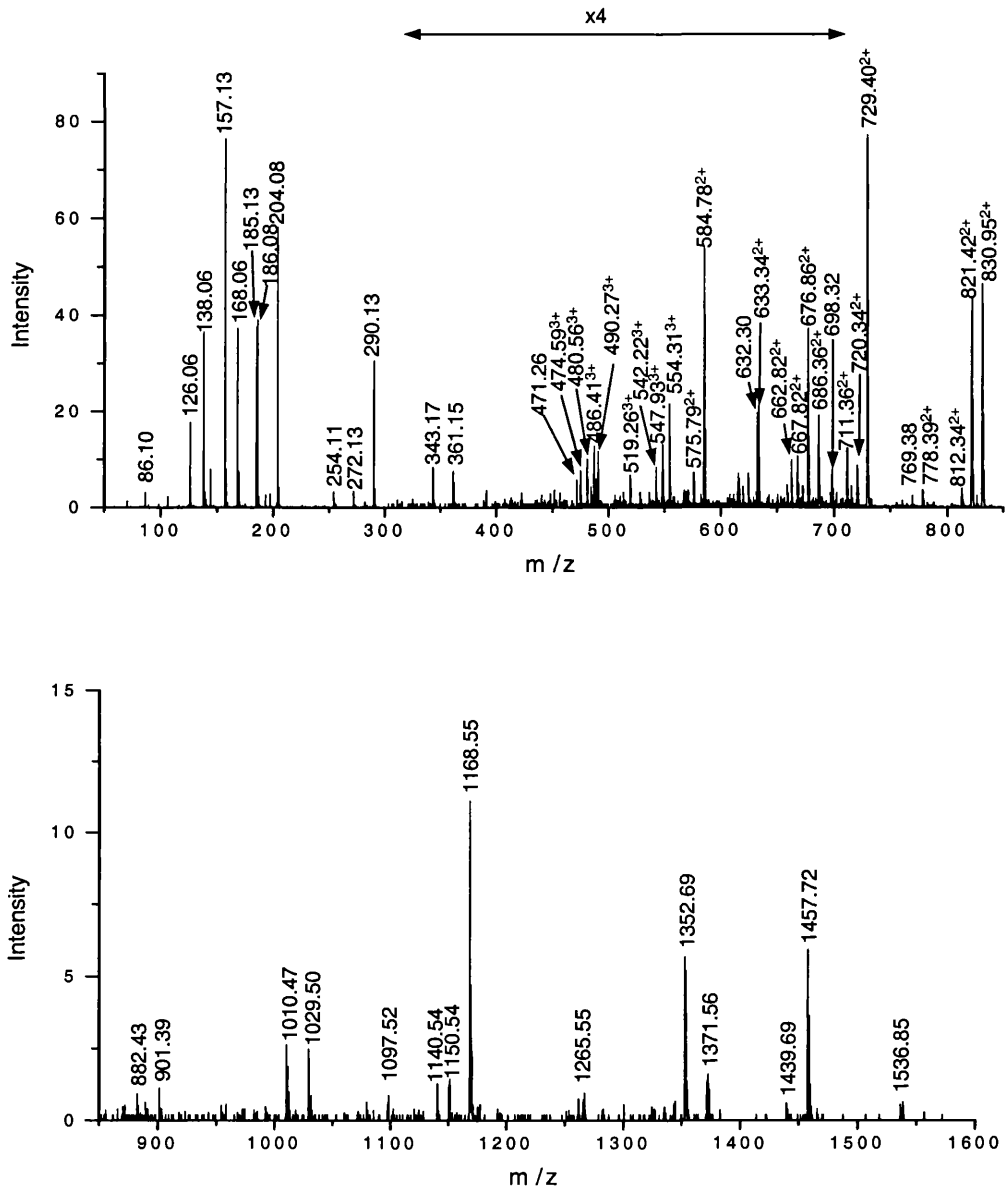
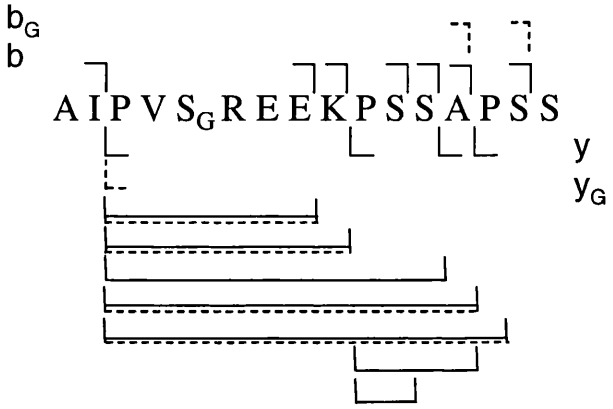


Figure 4.8: A CID-MS-MS spectrum of the GlcNAc-modified peptide $[M + 3H]^{3+}$ m/z 615.62.

The identities of the labelled peaks are given in Table 4.5. This spectrum is from an LC-MS run of a solution tryptic digest of αA-crystallin. The glycosylated internal fragment ions $PVSREE_G$ (m/z 901.39) and $PVSREEK_G$ (m/z 1029.50) determine the site of GlcNAc modification as serine 162.

Table 4.5: Fragment ions observed in the CID-MS-MS spectrum of $[M + 3H]^{3+}$ m/z 615.62.

This peptide corresponds to the GlcNAc-modified peptide spanning residues 158 – 173 of αA-crystallin and the fragmentation spectrum is Figure 4.8.. Several GlcNAc-modified b ions and internal ions are observed. The internal ions PVSREE$_G$ and PVSREEK$_G$ determine the site of GlcNAc modification as serine 162.



Peak	Match	Peak	Match	Peak	Match
86.10	L / I	519.26^{3+}	$b13_G$	812.34^{2+}	MH_2-H_2O
126.06	GlcNAc Fragment	542.22^{3+}	$y14_G$-$2H_2O$	821.42^{2+}	MH_2
138.06	GlcNAc Fragment	547.93^{3+}	$y14_G$-H_2O	830.95^{2+}	$y14_G$
157.13	a2	554.31^{3+}	y14	882.43	b8
168.06	GlcNAc Fragment	575.79^{2+}	PVSREEKPSSA-H_2O	901.39	PVSREE$_G$
185.13	b2	584.78^{2+}	PVSREEKPSSA	1010.47	b9
186.08	GlcNAc Fragment	632.30	y7	1029.50	PVSREEK$_G$
204.08	GlcNAc	633.34^{2+}	PVSREEKPSSAP	1097.52	PVSREEKPSS
254.11	y3-$2H_2O$	662.82^{2+}	SREEKPSSA	1140.54	PVSREEKPSSA-28
272.13	y3-H_2O	667.82^{2+}	b13	1150.54	PVSREEKPSSA-H_2O
290.13	y3	676.86^{2+}	PVSREEKPSSAPS	1168.55	PVSREEKPSSA
343.17	y4-H_2O	686.36^{2+}	PVSREEKPSSA$_G$	1265.55	PVSREEKPSSAP
361.15	y4	698.32	PVSREE	1352.69	b13
471.26	KPSSA	711.36^{2+}	y14-$2H_2O$	1371.56	PVSREEKPSSA$_G$
474.59^{3+}	y14-$2H_2O$	720.34^{2+}	y14-H_2O	1439.69	y14-H_2O
480.56^{3+}	y14-H_2O	729.40^{2+}	y14	1457.72	y14
486.41^{3+}	y14	769.38	b15	1536.85	b15
490.27^{3+}	PVSREEKPSSAP$_G$	778.39^{2+}	$b13_G$		

The GlcNAc modification site was determined from a spectrum acquired from 50% of a digest of 200 ng of α-crystallin, which would correspond to 2.5 pmoles of αA-crystallin digest loaded onto the HPLC. However, as less than 10% of the protein appears to be modified, this spectrum is acquired from, at most, 250 fmoles.

4.3. Identification of a GlcNAc Modification Site from an In-gel Digest

It was investigated whether the GlcNAc modification site of αA-crystallin could be identified from an in-gel digest, as separation of proteins by gel electrophoresis is a common method of purifying proteins prior to analysis. α-Crystallin was run by 1D SDS-PAGE, stained using Coomassie, then the lower band, corresponding to the A chain, was excised and subjected to in-gel tryptic digestion. The sample was analysed by LC-MS and automatic selection of peaks for fragmentation analysis was performed. The list of identified CID-MS-MS spectra matched using MASCOT is given in Table 4.6. Several spectra are successfully matched to peptides from αA-crystallin, including the doubly-charged unmodified peptide at m/z 821.40, but the modified peptide was not selected for fragmentation analysis. However, the GlcNAc-modified peptide is present (Figure 4.9). The modified peptide was not automatically selected as the more intense unmodified peptide effectively co-eluted, so was selected for fragmentation during the time the modified peptide was observed. The ratio of the intensity of the glycosylated peptide peak at $[M + 3H]^{3+}$ m/z 615.51 in relation to the unmodified ion at $[M + 3H]^{3+}$ m/z 547.81 is the same as that observed in the solution digest (Figure 4.3). This suggests that the in-gel digestion protocol, which includes extensive washing, and extraction in a strongly acidic solution, does not cause any deglycosylation.

The in-gel tryptic digest was then submitted for a second LC-MS-MS run to specifically select the triply charged glycopeptide for fragmentation analysis. The resulting spectrum (Figure 4.10) is essentially identical to that acquired from the solution digest, and the glycosylated internal fragment ions $PVSREE_G$ (m/z 901.21) and $PVSREEK_G$ (m/z 1029.31) define the glycosylation site. 400 ng of α-crystallin was loaded onto the gel, which corresponded to 200 ng of A chain and 200 ng of B chain.

The MS-MS spectrum was acquired from 50% of the digest of the A chain, which corresponded to up to 100 ng (5 pmoles) of protein, or 500 fmoles of modified protein if 10% of the protein was modified.

Table 4.6: List of ESI-CID-MS-MS spectra assigned by MASCOT from an in-gel digest of αA-crystallin band.

Ten spectra are matched to αA-crystallin.

Alpha A Crystallin						
Observed	**Mr (expt)**	**Mr (calc)**	**Delta**	**Miss**	**Peptide**	**Mod**
519.27	1036.53	1036.53	-0.01	0	TLGPFYPSR	
537.23	1072.45	1072.46	-0.01	0	QDDHGYISR	Pyro-Glu
545.75	1089.49	1089.48	0.01	0	QDDHGYISR	
586.79	1171.57	1171.59	-0.01	0	HFSPEDLTVK	
588.30	1174.59	1174.62	-0.03	0	TVLDSGISEVR	
612.79	1223.57	1223.59	-0.02	0	IPSGVDAGHSER	
650.82	1299.63	1299.65	-0.02	0	VQEDFVEIHGK	
714.33	1426.65	1426.71	-0.06	0	MDIAIQHPWFK	N-Acetyl
722.33	1442.64	1442.70	-0.06	0	MDIAIQHPWFK	N-Acetyl Met-Ox
821.40	1640.78	1640.84	-0.06	1	AIPVSREEKPSSAPSS	

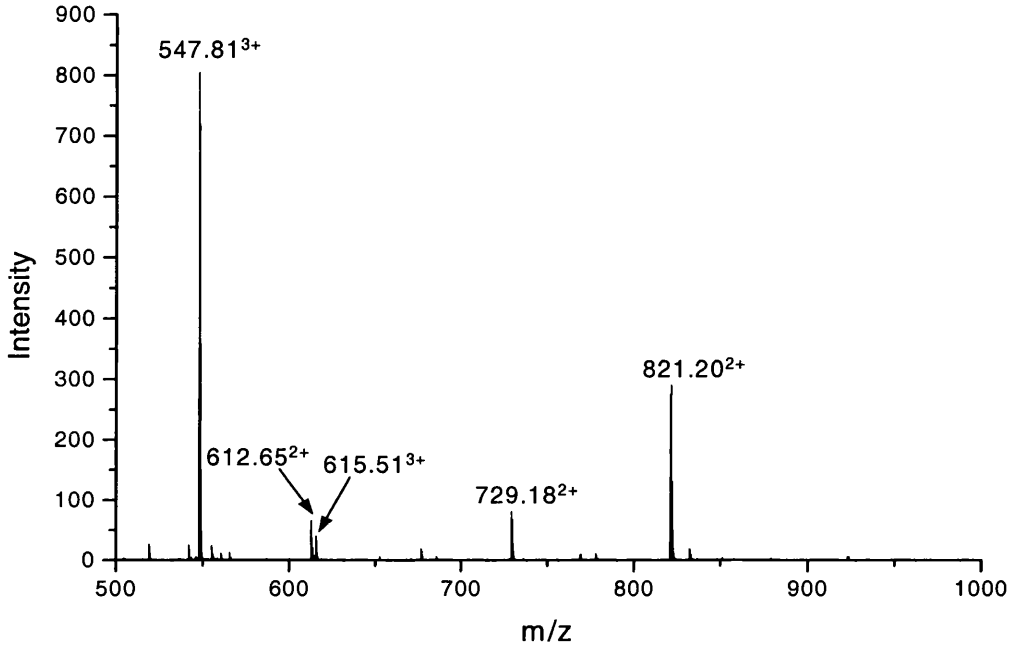


Figure 4.9: Combined mass spectrum over the period during which the triply charged m/z 547.93 and m/z 615.62 eluted.

Peaks observed include the doubly and triply charged unmodified peptide of residues 158 - 173 ($[M + 3H]^{3+}$ m/z 547.81 and $[M + 2H]^{2+}$ m/z 821.20) and the GlcNAc-modified peptide at $[M + 3H]^{3+}$ m/z 615.51.

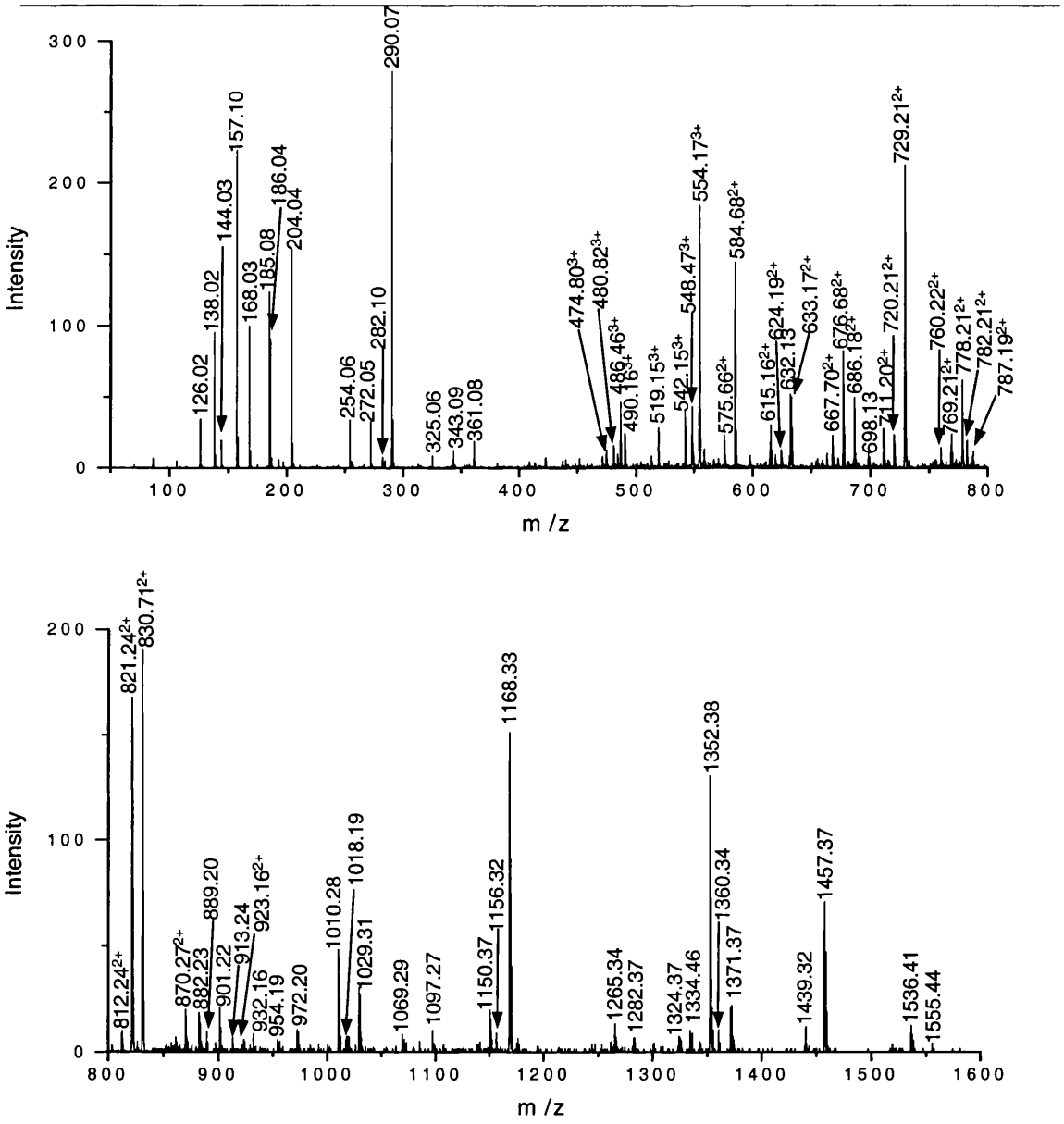


Figure 4.10: CID-MS-MS spectrum of the GlcNAc-modified peptide $[M + 3H]^{3+}$ m/z 615.51 from an in-gel digest of αA-crystallin.

The glycosylated fragment ions $PVSREE_G$ at m/z 901.21 and $PVSREEK_G$ at m/z 1029.31, identify the GlcNAc modification site as serine 162.

4.4. Precursor Ion Scanning to Locate GlcNAc-modified Peptides

Precursor ion scanning is usually carried out on a triple quadrupole instrument, as this is ideally designed for this type of experiment. The QSTAR quadrupole-TOF instrument can also perform precursor ion scanning. However, it is less sensitive for precursor ion scanning than a triple quadrupole instrument, due to the duty cycle of the TOF, as the instrument has to wait for the previous TOF separation to be completed before the next packet of ions can be accelerated[177]. Nevertheless, the QSTAR does have an advantage over triple quadrupole instruments in one respect for precursor ion scanning. As it uses time-of-flight to separate the fragment ions rather than a quadrupole, it has considerably higher resolution for product ions. This allows it to differentiate between product fragment ions of nominally the same mass. The QSTAR has been reported to be able to differentiate between a phosphotyrosine immonium ion at m/z 216.043 and the dipeptide ion TN at m/z 216.098[178]. This high resolution is also an advantage for precursor ion scanning for the GlcNAc oxonium ion.

A tryptic digest of α-crystallin, as well as producing the GlcNAc-modified peptide residues 158 to 173 from the A chain, should produce, amongst others, peptides of residues 89 to 99 of the A chain and 93 to 103 of the B chain. Both of these peptides have the sequence GK at their C-termini, so fragment ion spectra of these two peptides produce y_2 ions at a theoretical mass of m/z 204.135. The GlcNAc oxonium ion has a mass of m/z 204.086. A triple quadrupole instrument could not discriminate between these two nominally identical mass fragment ions.

Figure 4.11 is an electrospray mass spectrum of a tryptic digest of α-crystallin, which contained many peaks. The region between m/z 575 and m/z 660 is magnified in Figure 4.11(B). In this magnified mass range the GlcNAc-modified peptide from the A chain at $[M + 3H]^{3+}$ m/z 615.60 is just visible above the baseline. Also present are peaks at $[M + 2H]^{2+}$ m/z 583.33 and $[M + 2H]^{2+}$ m/z 650.83, which correspond to residues 93 – 103 of the B chain, and 158 – 173 of the A chain; the two aforementioned tryptic peptides ending with the sequence GK. In this spectrum both of these ions are significantly more intense than the GlcNAc-modified peptide ion. This sample was then analysed by

precursor ion scanning on the QSTAR. The top spectrum in Figure 4.12(A) shows product ions of mass m/z 204.1 ± 0.1 Th, and the resulting spectrum contains three peaks, corresponding to the GlcNAc-modified peptide at m/z 616.05 and the two tryptic peptides forming y_2 fragment ions of nominally the same mass. However, the spectrum in Figure 4.12(B) detects products of mass of m/z 204.08 ± 0.03 Th, and only a peak corresponding to the GlcNAc-modified peptide at m/z 616.39 is observed. Hence, using these parameters it is possible to differentiate between the glycosylated peptide and all other peptides in the digest.

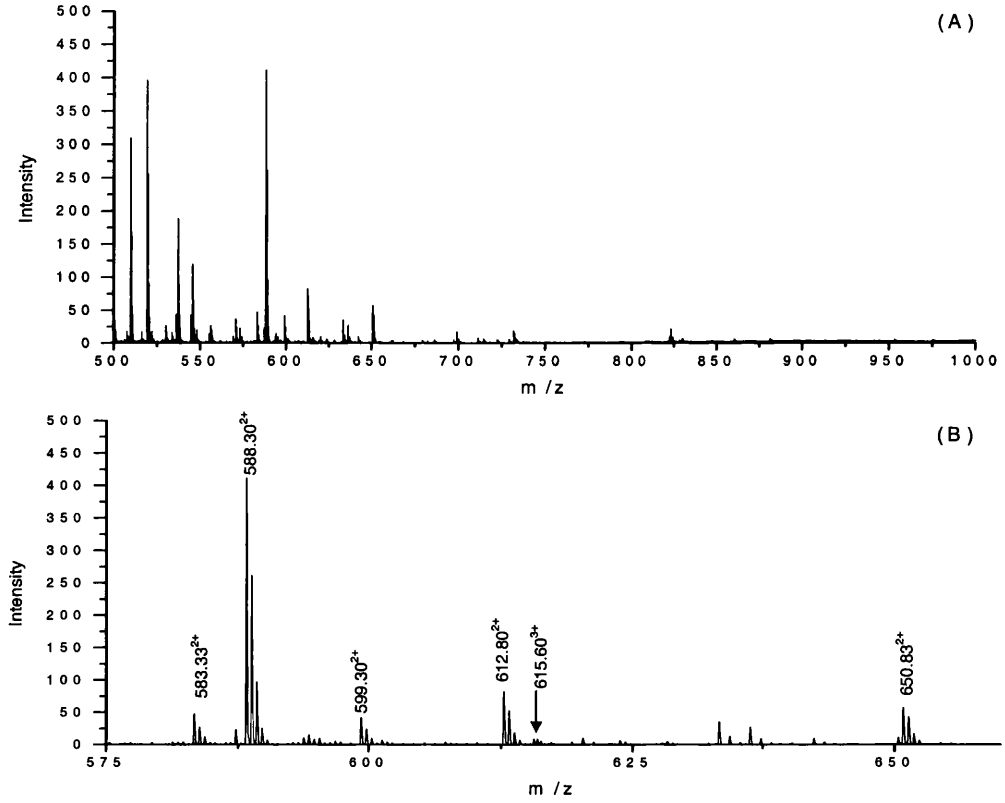


Figure 4.11: Nanospray-MS spectrum of a tryptic digest of α-crystallin.

(A) The complete spectrum contains a large number of peaks. (B) Peaks in the magnified region include $[M + 3H]^{3+}$ m/z 615.60, corresponding to the GlcNAc-modified peptide residues 158 – 173 from αA-crystallin; $[M + 2H]^{2+}$ m/z 583.33, corresponding to residues 93 – 103 of αB-crystallin; and $[M + 2H]^{2+}$ m/z 650.83, corresponding to residues 89 – 99 of αA-crystallin.

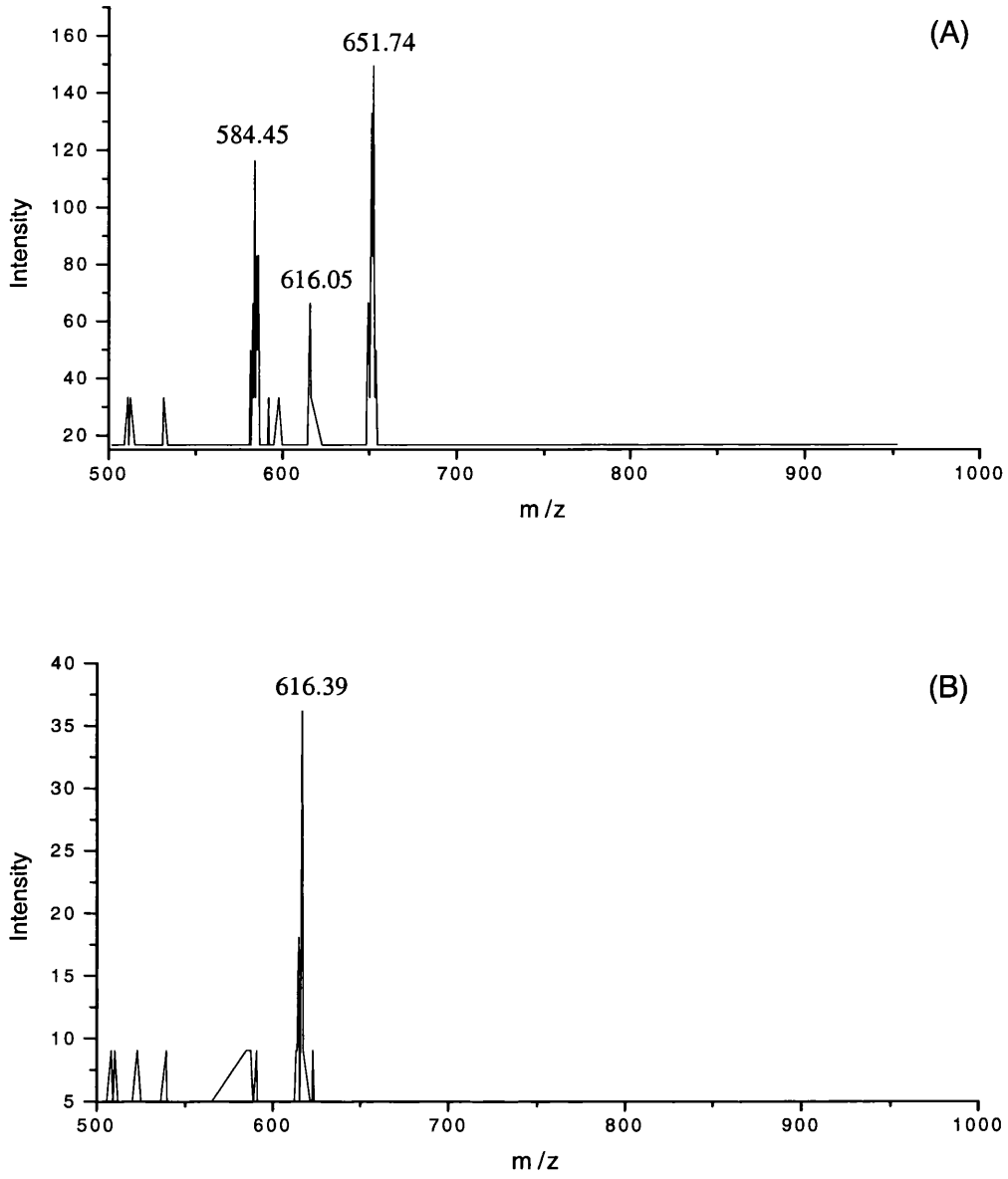


Figure 4.12: Precursor ion scanning for fragments of m/z 204 from a tryptic digest of α-crystallin.

A scan for precursors of m/z 204.1 ± 0.1 Th (A) produces a spectrum containing three peaks, corresponding to the GlcNAc-modified peptide (m/z 616.05) and the tryptic peptides of residues 93 to 103 of the B chain (m/z 584.45) and residues 89 to 99 of the A chain (m/z 651.74) that produce y_2 ions at a theoretical m/z 204.14. However, a scan for parents of m/z 204.08 ± 0.03, produces only one peak at m/z 616.39, corresponding to the GlcNAc-modified peptide.

4.5. Discussion

The results in this chapter were the first demonstration that a GlcNAc modification site of a protein can be determined from an enzymatic digest of the protein using mass spectrometry [107]. The site of modification of αA-crystallin was determined from a solution digest of 2.5 pmoles of α-crystallin. The stoichiometry of modification is known to be low [74, 169]. The intensity of the GlcNAc-modified triply-charged peptide observed during LC-MS analysis of the digest was less than 5% of the unmodified triply-charged peptide (Figure 4.2). It has been shown that GlcNAc-modification of a peptide halves the intensity of the peak by ESI-MS in comparison to an unmodified counterpart [59], but even so, this spectrum suggests that at most 10% of the protein was modified at this site. Hence, even if there were no sample losses during digestion and loading onto the HPLC, the site has been determined from 250 fmoles of modified peptide.

The GlcNAc-modified peptide from αB-crystallin was also detected, but was present at extremely low stoichiometry. The triply charged modified peptide was approximately 0.5% of the intensity of its unmodified counterpart (Figure 4.5). Assuming the addition of the GlcNAc moiety was not significantly altering the charge state of the peptide but was halving the peak intensity [59], this suggested a stoichiometry of modification of around 1% at this site.

The site of GlcNAc modification was also determined from an in-gel tryptic digest of αA-crystallin. The site was determined from 50% of a digest of 200 ng αA-crystallin. A similar quality spectrum was obtained to that acquired from a solution digest of half the amount of protein. This suggested approximately 50% sample loss during gel running, in-gel digestion and peptide extraction.

Another important result was the demonstration that the in-gel digestion of a protein from a gel band or spot did not cause any noticeable deglycosylation, as the ratio of the GlcNAc-modified and unmodified triply-charged peptide from αA-crystallin were the same in solution and in-gel digests (compare Figures 4.2 and 4.9). This makes gel electrophoresis an attractive and convenient protein purification technique prior to glycosylation analysis.

Finally, the demonstration that precursor ion scanning on the QSTAR could differentiate between the GlcNAc oxonium ion and other fragment ions of the same nominal mass demonstrated that using this instrument provides a more specific mass spectrometric scanning method for identifying GlcNAc-modified peptides in a complex mixture than on triple quadrupole instruments [66].

The major drawback of precursor ion scanning is its relative lack of sensitivity in comparison to the ability to detect intact peptides by mass spectrometry. However, if there is a significant amount of sample, and the requirement is to find glycosylated peptides within a complex mixture, then precursor ion scanning on the QSTAR is the most specific mass spectrometric scanning method available.

5. SERUM RESPONSE FACTOR

5.1. Introduction

The serum response element (SRE) is a regulatory sequence found upstream of the genes of many proteins transiently expressed upon growth factor stimulation[179]. The SRE binds the ubiquitously expressed serum response factor (SRF). SRF is a member of the MADS box family of transcription factors, which are named after the four original members of the family: MCM1, AG, DEFA and SRF. Members of this family contain a conserved domain of 56 amino acids. The N-terminal part of the domain defines the protein's DNA binding specificity, whilst the C-terminal part effects protein dimerisation. Dimers of SRF form ternary complexes on the SRE along with accessory ternary complex factors (TCFs) such as Elk-1, Sap-1 and Sap-2. SRF can be activated by serum, lysophosphatidic acid, aluminium fluoride and G proteins via the GTPase RhoA[180], and triggers the expression of key proteins involved in cell cycle progression, differentiation and development [181-183]

SRF is a 50 kDa protein containing a DNA binding domain which spans residues 133 –222 [184], of which the second half controls its binding as a dimer to the SRE. The MADS box sequence is located between residues 143 and 197. A schematic of SRF is given in Figure 5.1, and the sequence of the protein is presented in Appendix 3.

SRF is phosphorylated on serine 103 in response to stimulation by MAPKAP-K1 ($p90^{rsk}$) [185], in response to stress by MAPKAP-2 [186] and by calcium/calmodulin dependent kinases II and IV [187]. It is also phosphorylated at serine 83 by casein kinase II [188] and serine 435 by DNA activated protein kinase [189].

SRF is also O-GlcNAc-modified [190], and Reason *et al.*[92] characterised a number of GlcNAc modification sites. Using recombinant SRF over-expressed in baculovirus [164], they identified serine 283 as a major site of GlcNAc modification. Serine 316 was also found to be glycosylated, and one of serine 307 or serine 309 was modified at very low stoichiometry. Sites were identified by first carrying out sequential enzymatic digestion of the protein using a combination of enzymes. Peptides were then separated by HPLC and fractions were collected. These were analysed by fast atom bombardment

mass spectrometry (FAB-MS) to look for fractions containing peptides that differed in mass by 203 Da, which could correspond to unmodified and GlcNAc-modified versions of the same peptide. Selected fractions were subjected to *in vitro* [^{3}H]galactose labelling to tag the GlcNAc residue, and sites of modification were finally determined by Edman degradation analysis and monitoring for the release of the radioactively modified amino acid. This work was carried out starting with 5 - 10 nmoles of highly purified protein. This large amount of starting material was partly required due to the low ionisation efficiency of peptides by FAB-MS. Indeed, for some of the peptides they propionylated the hydroxyl groups to improve the ionisation efficiency in FAB-MS [191], but the limit of detection of peptides is still two to three orders of magnitude lower using modern cutting-edge MALDI-MS and ESI-MS instruments. Also, their approach contained a large number of steps, increasing the chances of sample loss through the analysis.

The aim of the work in this chapter was to show that using mass spectrometry sites of O-GlcNAc modification could be identified at much higher sensitivity, *i.e.* starting with only 5 – 10 pmoles of protein. It was also hoped that previously undetected sites of modification would be identified.

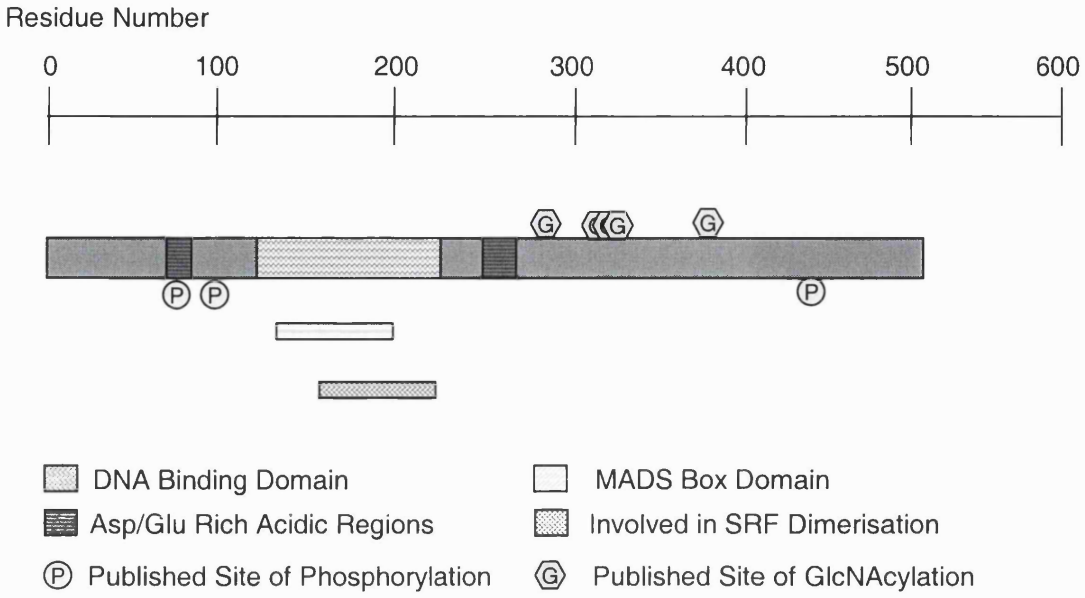


Figure 5.1: Graphical representation of structural features of SRF.

The region spanning residues 133 – 222 binds to DNA. Contained within this region is a MADS box domain (residues 143 – 197), which is a DNA binding region conserved across a number of proteins. The region 168 – 222 is important for the dimerisation of SRF. There are also two highly acidic regions of the protein spanning residues 80 – 90 and 242 – 258.

5.2. Tryptic Digestion

5.2.1. MALDI-MS

An in-gel tryptic digest of SRF was carried out, from which the MALDI mass fingerprint shown in Figure 5.2 was obtained. A large number of peaks were observed in the mass range up to m/z 7000. Unfortunately, peaks above m/z 4500 were not resolved, so monoisotopic masses could only be estimated to the nearest dalton in mass.

The resulting peak list was submitted for a search using MS-FIT[44] and the search result is presented in Table 5.1. Only 20 of the 77 labelled peaks were assigned to theoretical tryptic peptides from SRF, corresponding to 57% sequence coverage. This sequence coverage is low, and does not permit exhaustive analysis of the protein's post-

translational state. However, examination of the sequence of SRF shows there are no tryptic cleavage sites between residues 374 and 506. Thus, even with no missed cleavages this region would produce a peptide of m/z 13548.62, higher than the mass range examined here. An attempt was made to look for this higher mass peptide (data not shown), but it could not be observed by MALDI-MS. This region alone comprises 26% of the protein and partly explains the disappointingly low sequence coverage. The other disturbing feature of this result is the large number of non-matching peptides. This sample has been highly purified [164] and further run on a 1D gel, and therefore should not contain significant contamination. Several of the unmatched peaks can be assigned to trypsin autolysis products, and others are known matrix contaminants, but this still leaves many unassigned ions. A second MS-FIT search was performed, looking for potential serine- or threonine-phosphorylated peptides (Table 5.2). This type of search gives many false matches, as it allows for phosphorylation of all serine and threonine residues regardless of sequence, so caution should be applied when interpreting this data. However, if a potential phosphorylated peptide is matched where the non-phosphorylated version of the same peptide is also observed, this makes the assignment more probable. Four potential phosphorylated peaks fit this criterion, and these are highlighted in Table 5.2.

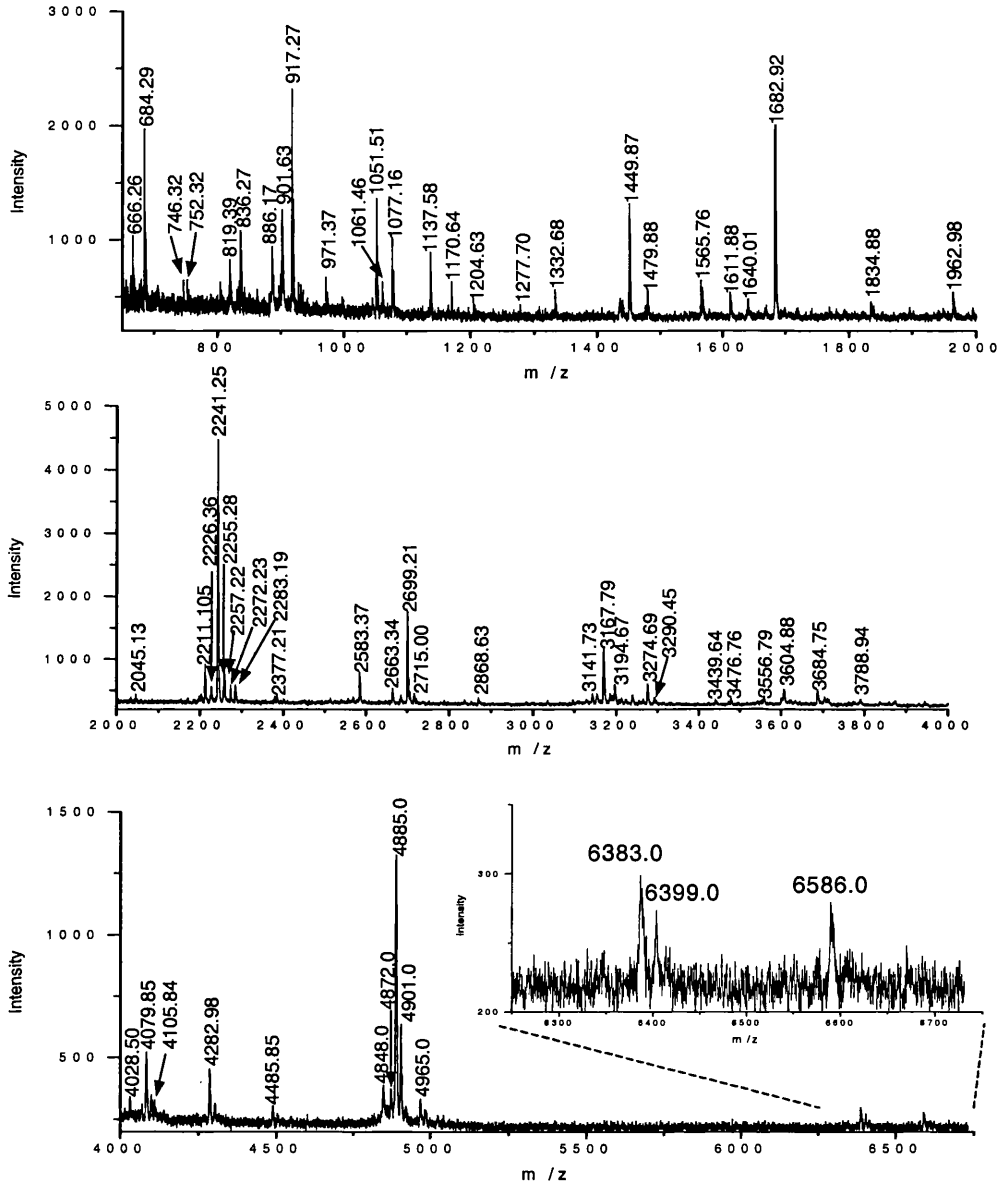


Figure 5.2: MALDI mass spectrum of a tryptic in-gel digest of SRF.

Peak labels are monoisotopic masses. Peak labels above m/z 4500 are not sufficiently resolved to accurately label monoisotopic peaks, and are thus labelled to the monoisotopic peak is labelled to the nearest dalton. The mass range m/z 6250 – 6750 is magnified in the inset, and these peaks are discussed in the text. The MS-FIT search result of this spectrum is Table 5.1.

Two of these peptides (m/z 3556.79 and m/z 3684.75) contain the published phosphorylation site at serine 103 [185-187]. The other two peptides, corresponding to residues 213 – 235 and 325 – 373 do not contain known phosphorylated residues.

Table 5.1: MS-FIT search result of the peaks observed in a MALDI mass fingerprint of a tryptic digest of SRF.

(Met-ox = oxidised methionine, Cys-am = acrylamide modified cysteine)

1. 20/77 matches (29%). 51594.1 Da, pI = 7.08. Acc. # 11831. HUMAN. Serum Response Factor.

m/z submitted	MH+ Matched	Delta ppm	Start	End	Peptide Sequence	Mod
746.319	746.3725	-71.654	158	163	(R)YTTFSK(R)	
1051.563	1051.5649	-1.7747	44	55	(R)VPGNGAGLGPGR(L)	
1137.579	1137.5978	-16.5312	146	154	(K)IKMEFIDNK(L)	
1170.635	1170.598	31.6443	25	37	(R)TPTGRPGGGGGTR(G)	
1204.628	1204.6247	2.6989	202	212	(K)LQPMITSETGK(A)	
1332.683	1332.7197	-27.5466	201	212	(R)KLQPMITSETGK(A)	
1449.871	1449.7926	54.053	44	58	(R)VPGNGAGLGPGRLER(E)	
1611.878	1611.8165	38.1769	213	227	(K)ALIQTCLNSPDSPPR(S)	
1682.92	1682.8435	45.4584	25	43	(R)TPTGRPGGGGGTRGANGGR(V)	
1682.92	1682.8536	39.4689	213	227	(K)ALIQTCLNSPDSPPR(S)	1Cys-am
2583.374	2583.2473	49.0455	213	235	(K)ALIQTCLNSPDSPPRSDPTTDQR(M)	1Cys-am
2699.21	2699.1518	21.5657	236	260	(R)MSATGFEETDLTYQVSESDSSGETK(D)	
2715.002	2715.1467	-53.2956	236	260	(R)MSATGFEETDLTYQVSESDSSGETK(D)	1Met-ox
2868.625	2868.4599	57.5435	202	227	(K)LQPMITSETGKALIQTCLNSPDSPPR(S)	1Cys-am
3141.726	3141.6472	25.0807	172	200	(K)AYELSTLTGTQVLLLVASETGHVYTFATR(K)	
3476.756	3476.7041	14.9151	101	138	(R)SLSEMEIGMVVGGPEASAAATGGYGPVSGAVSGAKPGK(K)	
3604.878	3604.7991	21.8853	101	139	(R)SLSEMEIGMVVGGPEASAAATGGYGPVSGAVSGAKPGKK(T)	
4885	4885.4729	-96.7992	325	373	(K)STGSGPVSSGGLMQLPTSFTLMPGGAVAQQVPVQAIQVHQAPQQASPSR(D)	
4901	4901.4678	-95.4458	325	373	(K)STGSGPVSSGGLMQLPTSFTLMPGGAVAQQVPVQAIQVHQAPQQASPSR(D)	1Met-ox
6383	6383.2496	-39.1036	261	324	(K)DTLKPAFTVTNLPGTTSTIQTAPSTSTTMQVSSGPSFPITNYLAPVSASVSPSA VSSANGTVLK(S)	
6399	6399.2445	-38.2112	261	324	(K)DTLKPAFTVTNLPGTTSTIQTAPSTSTTMQVSSGPSFPITNYLAPVSASVSPSA VSSANGTVLK(S)	1Met-ox

1	11	21	31	41	51	61	71	
MLPTQAGAAA	ALGRGSALGG	SLNRTPTGRP	GGGGGTRGAN	GGRVPGNGAG	LGPGRLEREA	AAAAATTPAP	TAGALYSGSE	
81	91	101	111	121	131	141	151	
GDSESGEEEE	LGAERRGLKR	SLSEMEIGMV	VGGPEASAAA	TGGYGPVSGA	VSGAKPGKKT	RGRVKIKMEF	IDNKLRRYTT	
161	171	181	191	201	211	221	231	
FSKRKTGIMK	KAYELSTLTG	TQVLLLVASE	TGHVYTFATR	KLQPMITSET	GKALIQTCLN	SPDSPPRSDP	TTDQRMSATG	
241	251	261	271	281	291	301	311	
FEETDLTYQV	SESDSSGETK	DTLKPAFTVT	NLPGTTSTIQ	TAPSTSTTMQ	VSSGPSFPIT	NYLAPVSASV	SPSAVSSANG	
321	331	341	351	361	371	381	391	
TVLKSTGSGP	VSSGGLMQLP	TSFTLMPGGA	VAQQVPVQAI	QVHQAPQQAS	PSRDSSTDLT	QTSSSGTVTL	PATIMTSSVP	
401	411	421	431	441	451	461	471	
TTVGGHMMYP	SPHAVMYAPT	SGLGDGSLTV	LNAFSQAPST	MQVSHSQVQE	PGGVPQVFLT	ASSGTVQIPV	SAVQLHQMAV	
481	491	501						
IGQQAGSSSN	LTELQVVNLD	TAHSTKSE						

The matched peptides cover **57%** (290/508AA's) of the protein

A similar search was performed searching for potential O-GlcNAc-modified peptides. The only peak match for which an unmodified version of the same peptide was also assigned was at m/z 6586 and is shown in the magnified region of Figure 5.2, along with the peak of the unmodified peptide at m/z 6383. This peptide corresponds to residues 261 - 324 and contains four published O-GlcNAc modification sites[92].

Table 5.2: MS-FIT search result of the peaks observed in a MALDI mass fingerprint of a tryptic digest of SRF, allowing for possible phosphorylated peptides.

m/z Submitted	MH^+ Matched	Delta ppm	Start	End	Peptide Sequence	Mod
746.319	746.3725	-71.654	158	163	(R)YTTFSK(R)	
1051.563	1051.5649	-1.7747	44	55	(R)VPGNGAGLGPGR(L)	
1137.579	1137.5978	-16.5312	146	154	(K)IKMEFIDNK(L)	
1170.635	1170.598	31.6443	25	37	(R)TPTGRPGGGGGTR(G)	
1204.628	1204.6247	2.6989	202	212	(K)LQPMITSETGK(A)	
1332.683	1332.7197	-27.5466	201	212	(R)KLQPMITSETGK(A)	
1449.871	1449.7926	54.053	44	58	(R)VPGNGAGLGPGRLER(E)	
1611.878	1611.8165	38.1769	213	227	(K)ALIQTCLNSPDSPPR(S)	
1682.92	1682.8435	45.4584	25	43	(R)TPTGRPGGGGGTRGANGGR(V)	
1682.92	1682.8536	39.4689	213	227	(K)ALIQTCLNSPDSPPR(S)	1Cys-am
2241.252	2241.1774	33.298	1	24	(-)MLPTQAGAAAALGRGSALGGSLDR(T)	
2257.217	2257.1723	19.8089	1	24	(-)MLPTQAGAAAALGRGSALGGSLDR(T)	1Met-ox
2583.374	2583.2473	49.0455	213	235	(K)ALIQTCLNSPDSPPRSDPTTDQR(M)	1Cys-am
2663.337	2663.2136	46.3223	213	235	(K)ALIQTCLNSPDSPPRSDPTTDQR(M)	1PO4 1Cys-am
2699.21	2699.1518	21.5657	236	260	(R)MSATGFEETDLTYQVSESDSSGETK(D)	
2715.002	2715.1467	-53.2956	236	260	(R)MSATGFEETDLTYQVSESDSSGETK(D)	1Met-ox
2836.3	2836.2042	33.765	15	43	(R)GSALGGSLDRTPTGRPGGGGGTRGANGGR(V)	3PO4
2868.625	2868.4599	57.5435	202	227	(K)LQPMITSETGKALIQTCLNSPDSPPR(S)	1Cys-am
3141.726	3141.6472	25.0807	172	200	(K)AYELSTLTGTQVLLLVASETGHVYTFATR(K)	
3194.673	3194.9447	-85.0368	236	260	(R)MSATGFEETDLTYQVSESDSSGETK(D)	6PO4 1 Met-Ox
3236.715	3236.4539	80.6769	201	227	(R)KLQPMITSETGKALIQTCLNSPDSPPR(S)	3PO4 1Cys-am
3274.692	3274.911	-66.8776	236	260	(R)MSATGFEETDLTYQVSESDSSGETK(D)	7PO4 1Met-ox
3476.756	3476.7041	14.9151	101	138	(R)SLSEMEIGMVVGGPEASAAATGGYGPVSGAVSGAKPGK(K)	
3556.787	3556.6705	32.7624	101	138	(R)SLSEMEIGMVVGGPEASAAATGGYGPVSGAVSGAKPGK(K)	1PO4
3604.878	3604.7991	21.8853	101	139	(R)SLSEMEIGMVVGGPEASAAATGGYGPVSGAVSGAKPGKK(T)	
3684.745	3684.7654	-5.5467	101	139	(R)SLSEMEIGMVVGGPEASAAATGGYGPVSGAVSGAKPGKK(T)	1PO4
3788.939	3788.6412	78.5982	59	96	(R)EAAAAAATTPAPTAGALYSGSEGDSESGEEEELGAERR(G)	1PO4
4028.494	4028.5402	-11.4709	59	96	(R)EAAAAAATTPAPTAGALYSGSEGDSESGEEEELGAERR(G)	4PO4
4282.983	4282.8609	28.5096	97	138	(R)GLKRSLSEMEIGMVVGGPEASAAATGGYGPVSGAVSGAKPGK(K)	4PO4 2Met-Ox
4885	4885.4729	-96.7992	325	373	(K)STGSGPVSSGGLMQLPTSFTLMPGGAVAQQVPVQAIQVHQAPQQASPSR(D)	
4901	4901.4678	-95.4458	325	373	(K)STGSGPVSSGGLMQLPTSFTLMPGGAVAQQVPVQAIQVHQAPQQASPSR(D)	1Met-ox
4965	4965.4392	-88.4595	325	373	(K)STGSGPVSSGGLMQLPTSFTLMPGGAVAQQVPVQAIQVHQAPQQASPSR(D)	1PO4
6383	6382.9093	14.2051	213	260	(K)ALIQTCLNSPDSPPRSDPTTDQRMSATGFEETDLTYQVSESDSSGETK(D)	14PO4 1Cys-am
6383	6383.2496	-39.1036	261	324	(K)DTLKPAFTVTNLPGTTSTIQTAPSTSTTMQVSSGPSFPITNYLAPVSASVSPSA VSSANGTVLK(S)	
6399	6398.9042	14.9643	213	260	(K)ALIQTCLNSPDSPPRSDPTTDQRMSATGFEETDLTYQVSESDSSGETK(D)	14PO4 1Met-Ox 1Cys-am
6399	6399.2445	-38.2112	261	324	(K)DTLKPAFTVTNLPGTTSTIQTAPSTSTTMQVSSGPSFPITNYLAPVSASVSPSA VSSANGTVLK(S)	1Met-ox

5.2.2. LC-ESI-CID-MS-MS

The tryptic digest was then analysed by LC-MS with automatic function switching to select peaks for MS-MS analysis, allowing for up to two precursors to be selected for fragmentation analysis at any given time. Figure 5.3 shows the total ion chromatograms (TIC's) for the ESI-MS spectra, and the two functions of ESI-CID-MS-MS spectra. Fragmentation spectra were automatically submitted for a MASCOT MS-MS search[47]. A summary of the output is given in Table 5.3. MASCOT has assigned 20 CID-MS-MS spectra to 16 different peptides from SRF, comprising 36% sequence coverage. However, it has also matched spectra to trypsin, bovine serum albumin (BSA) and keratins. BSA was a molecular weight marker in a neighbouring lane on the 1D gel the SRF was purified on immediately prior to digestion, and runs at nominally the same mass as SRF. Hence, a small amount of contamination must have been introduced during sample loading onto the gel. The keratins are introduced post-electrophoresis due to contamination from the laboratory environment during sample handling. These matches help to explain several of the unassigned peaks in the MALDI mass fingerprint. Despite this, there were still many MS-MS spectra that MASCOT was unable to match to tryptic peptides. Therefore, each CID-MS-MS spectrum was manually analysed to see if other peptides could be identified.

Manual interpretation permitted assignment of a further 22 CID-MS-MS spectra to peptides derived from SRF (Table 5.4). There are a several reasons why these spectra were not matched by MASCOT. For some peaks the second isotope was selected, so the molecular mass for the parent ion was wrong. For a number of peptides spanning residues 44 – 55 or longer missed cleavage products of this region, the asparagine at residue 47 had undergone deamidation to aspartic acid. Deamidation of asparagine to aspartic acid is a common reaction upon peptide storage[192].

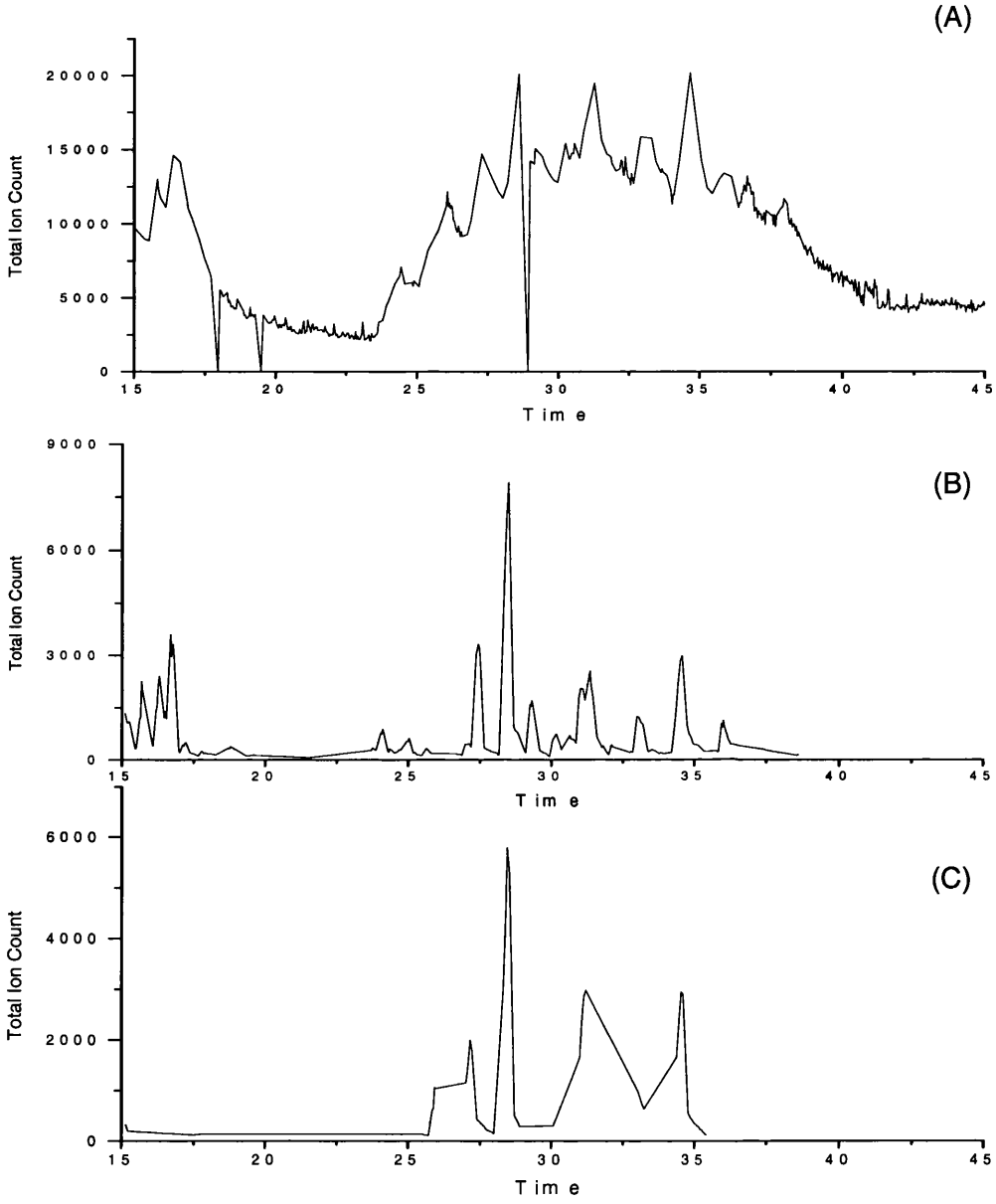


Figure 5.3: TICs of the ESI-MS spectra, and two ESI-CID-MS-MS functions.

(A) TIC of ESI-MS spectra. (B) and (C) TICs of MS-MS functions.

Table 5.3: Summary of a MASCOT search of the CID-MS-MS spectra produced in the LC-MS analysis of tryptic SRF.

1. gi|4507205 **Mass:** 51561 **Total score:** 550 **Peptides matched:** 15
serum response factor (c-fos serum response element-binding transcription factor) [Homo sapiens]

Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Rank	Peptide
448.73	895.44	895.41	0.03	0	12	6	MEFIDNK
466.27	930.52	930.49	0.03	0	48	1	GSALGGSLNR
466.28	930.54	930.49	0.06	0	-16	1	GSALGGSLNR
526.31	1050.6	1050.56	0.04	0	65	1	VPGNGAGLGPGR
569.31	1136.61	1136.59	0.02	1	63	1	IKMEFIDNK
598.85	1195.68	1195.67	0.02	0	57	1	LPTQAGAAAALGR
602.82	1203.63	1203.62	0.02	0	-21	1	LQPMITSETGK
602.83	1203.65	1203.62	0.03	0	27	1	LQPMITSETGK
666.88	1331.75	1331.71	0.04	1	69	1	KLQPMITSETGK
666.89	1331.76	1331.71	0.05	1	-46	1	KLQPMITSETGK
725.41	1448.8	1448.78	0.01	1	4	2	VPGNGAGLGPGRLER
806.42	1610.83	1610.81	0.02	0	79	1	ALIQTCLNSPDSPPR
1350.04	2698.06	2698.14	-0.08	0	104	1	MSATGFEETDLTYQVSESDSSGETK
901.96	3603.82	3603.79	0.03	1	13	1	SLSEMEIGMVVGGPEASAAATGGYGPVSGAVSGAKPGKK
1222.11	4884.4	4884.46	-0.06	0	8	1	STGSGPVSSGGLMQLPTSFTLMPGGAVAQQVPVQAIQVHQAPQQASPSR

2. gi|14783624 **Mass:** 65999 **Total score:** 294 **Peptides matched:** 6
keratin 1 (epidermolytic hyperkeratosis) [Homo sapiens]

Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Rank	Peptide
517.28	1032.55	1032.51	0.04	0	36	1	TLLEGEESR
590.32	1178.63	1178.59	0.04	0	53	1	YEELQITAGR
633.34	1264.65	1264.63	0.02	0	33	1	TNAENEFVTIK
639.38	1276.74	1276.7	0.03	0	64	1	LALDLEIATYR
697.37	1392.73	1392.72	0.01	1	40	1	TNAENEFVTIKK
738.4	1474.79	1474.78	0.02	0	71	1	FLEQQNQVLQTK

3. gi|999627 **Mass:** 8814 **Total score:** 173 **Peptides matched:** 3
Chain B, Porcine E-Trypsin (E.C.3.4.21.4)

Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Rank	Peptide
421.77	841.52	841.5	0.02	0	48	1	VATVSLPR
523.3	1044.59	1044.56	0.04	0	58	1	LSSPATLNSR
737.72	2210.13	2210.1	0.04	0	67	1	LGEHNIDVLEGNEQFINAAK

4. gi|2190337 **Mass:** 69278 **Total score:** 153 **Peptides matched:** 3
(X58989) serum albumin [Bos taurus]

Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Rank	Peptide
464.27	926.52	926.49	0.03	0	37	2	YLYEIAR
653.38	1304.74	1304.71	0.03	0	58	1	HLVDEPQNLIK
740.41	1478.81	1478.79	0.02	0	58	1	LGEYGFQNALIVR

5. gi|6678643 **Mass:** 65183 **Total score:** 143 **Peptides matched:** 3
keratin complex 2, basic, gene 1 [Mus musculus]

Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Rank	Peptide
633.34	1264.65	1264.63	0.02	0	33	1	TNAENEFVTIK
697.37	1392.73	1392.72	0.01	1	40	1	TNAENEFVTIKK
738.4	1474.79	1474.81	-0.02	1	71	1	FLEQQNKVLQTK

Table 5.4: Additional CID-MS-MS spectra assigned to SRF after manual interpretation of spectra.

** = peptide is a product of a non-specific enzymatic cleavage.*

Observed	Monoistopic Mass	Match
666.90	1332.80	201-212
526.81	1052.62	44-55 N⇒D
725.93	1450.86	44-58 N⇒D
484.29	1450.88	44-58 N⇒D
655.33	1964.00	38-58 N⇒D
598.88	1196.76	2-14
861.77	2583.32	213-235 Cys-Am
841.94	1682.89	25-43
888.387	2663.16	213-235 Cys-Am Phos
758.07	2272.21	485-506*
956.80	2868.40	202-227 Cys-Am
900.40	2699.20	236-260
907.73	2721.219	236-260 Na^+ Adduct
753.05	2257.16	Novel N-terminus Met-Ox
1129.08	2257.14	Novel N-terminus Met-Ox
1228.87	3684.60	101-139 Phos
747.73	2241.20	Novel N-terminus
1121.059	2241.12	Novel N-terminus
1360.62	4079.86	374-413*
1079.54	3236.62	374-406*
1106.99	2212.98	374-395*
448.74	896.49	148-154
1056.53	3167.58	202-230*

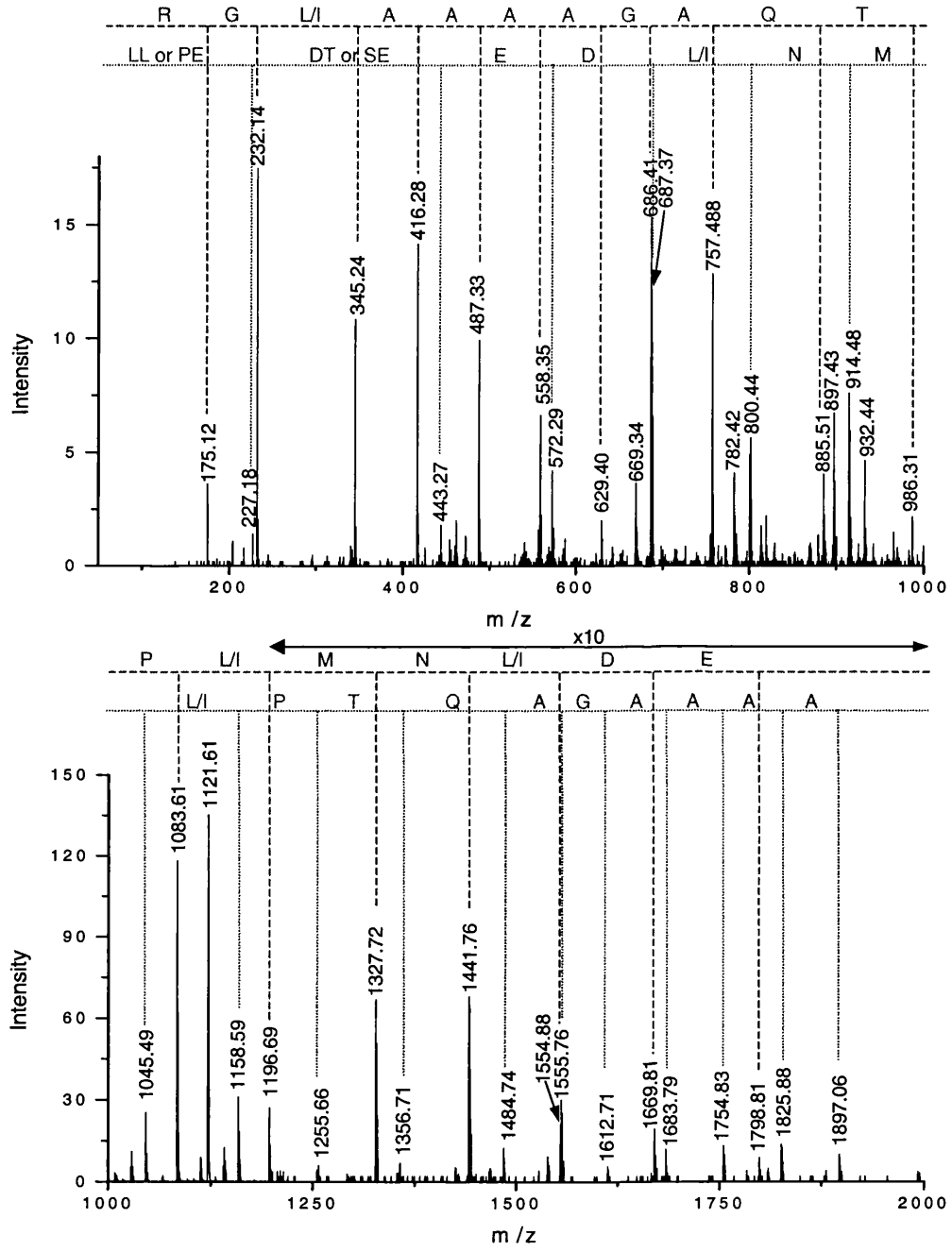


Figure 5.4: ESI-CID-MS-MS spectrum of $[M + 2H]^{2+}$ m/z 1121.06 from tryptic SRF. *This intense peak was not assigned to a predicted tryptic peptide of SRF. Almost complete ,b' and ,y' ion series are observed, and from this spectrum a sequence of (227.18)(216.09)ED[L/I]NM[L/I]PTQAGAAAA[L/I]GR can be assigned.*

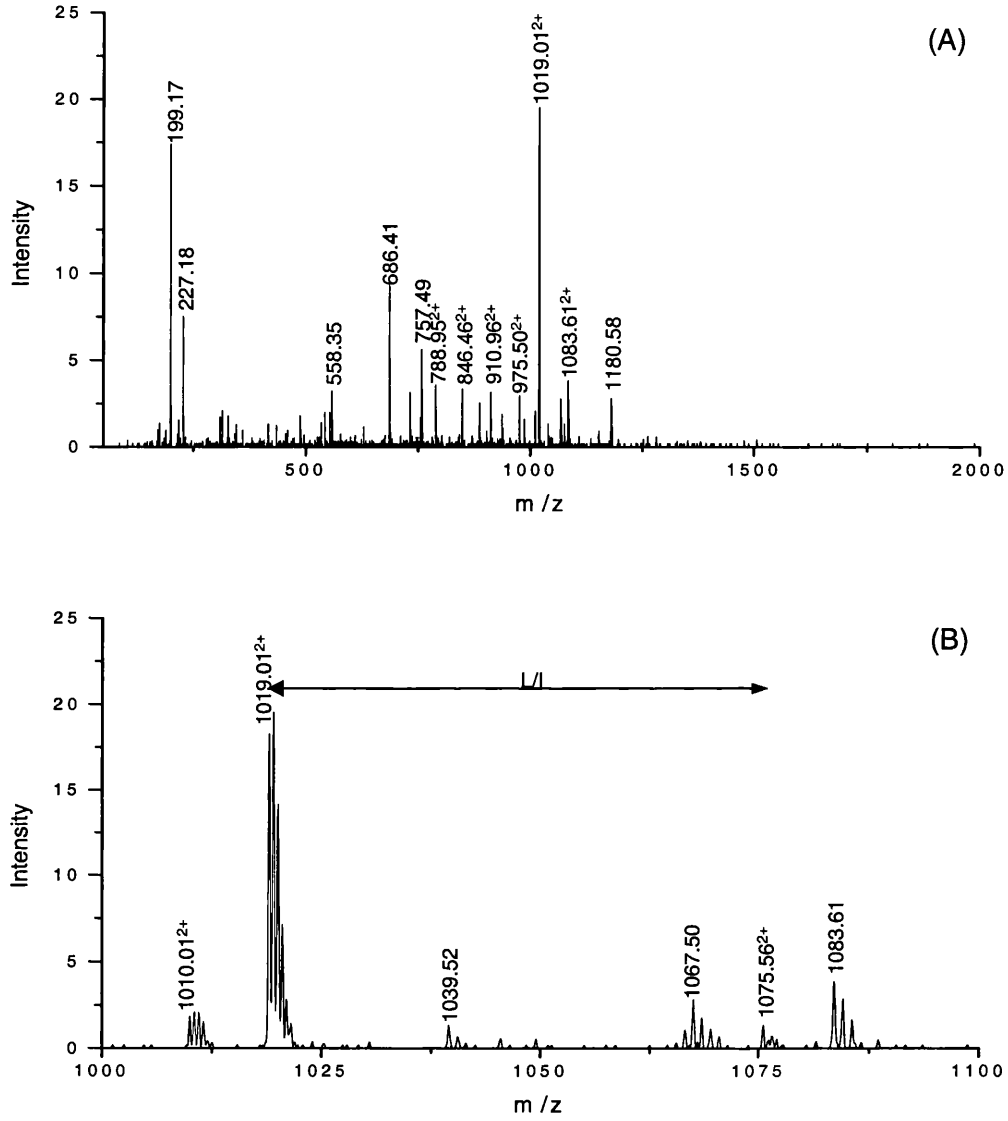


Figure 5.5: ESI-CID-MS-MS spectrum of [M + 2H + Na]$^{3+}$ m/z 755.06 from an LC-MS run of tryptic SRF.

The spectrum is of the same peptide as fragmented in Figure 5.4. (A) Whole spectrum shows large number of fragment ions. (B) Magnified region contains doubly-charged fragment ion at m/z 1075.56, which corresponds to the loss of [L/I] from the parent ion. Combined with the knowledge that the two most N-terminal residues form a b_2 at m/z 227.18 (Figure 5.4), this determines the two most N-terminal residues as [L/I] [L/I].

Two major peaks in the MALDI mass spectrum at m/z 2241.25 and m/z 2257.22 were not assigned to tryptic peptides from SRF. Their difference in mass (16 Da) suggests that the m/z 2257.22 peak is likely to be a methionine-oxidised version of the m/z 2241.25. These two peaks appeared as doubly and triply-charged moieties in the LC-MS run, and were selected for CID-MS-MS fragmentation. The CID spectrum of the non-oxidised $[M + 2H]^{2+}$ m/z 1121.06 is shown in Figure 5.4. This spectrum contains an almost complete set of 'b' and 'y' ions. Hence, the majority of the peptide sequence can be relatively easily interpreted. The C-terminal half of this peptide matches to the most N-terminal part of the SRF protein sequence (MLPTQAGAAAALGR). However, this peptide extends beyond the proposed start of the protein a further eight residues. Four of these extra residues can be interpreted as ED[L/I]N. There is then a dipeptide gap of mass 216.09 Da and the most N-terminal residues formed a b_2 ion at m/z 227.18. A gap of 216.09 Da can be formed by the amino acid combinations SE or DT. A weak ion at m/z 342.22 suggests the combination DT in this peptide. There are two dipeptide combinations that can produce a b_2 ion of nominal mass m/z 227. PE forms an ion at m/z 227.10, whereas [L/I][L/I] creates an ion of m/z 227.17. The mass accuracy of the spectrum suggests [L/I][L/I] as the probable two most N-terminal residues. This assignment is confirmed in the MS-MS spectrum of $[M + 2H + Na]^{3+}$ m/z 755.06 (Figure 5.5) which is of the same peptide. A fragment ion at $[M + 2H]^{2+}$ m/z 1075.56 was observed, which corresponds to a doubly-charged y_{21} ion formed by the loss of a leucine or isoleucine from the parent ion, and the difference between this peak and the y_{20} ion at $[M + 2H]^{2+}$ m/z 1019.01 is a further [L/I]. Thus, the complete sequence of this peptide is [L/I][L/I][DT/SE]ED[L/I]NMLPTQAGAAAALGR. It was subsequently found that the protein contained a myc tag to facilitate protein purification. Hence, the correct sequence is LISEEDLNMLPTQAGAAAALGR.

Two more peaks that were not assigned in the MASCOT output were $[M + 3H]^{3+}$ m/z 861.77 and $[M + 3H]^{3+}$ m/z 888.39. These correspond to singly charged peaks at mass of m/z 2583.32 and m/z 2663.16 respectively. These were both observed in the MALDI mass fingerprint, and were assigned to the peptide residues 213 - 235 containing an acrylamide modified cysteine residue, and a phosphorylated version of this peptide. A spectrum containing these two peaks, along with their extracted ion chromatograms is

given in Figure 5.6. These chromatograms show the potential phosphorylated peak eluting slightly after the unmodified peptide. Figure 5.7 shows a fully labelled spectrum of the phosphopeptide, and peak identities are given in Table 5.5. The crucial region that determines the site of phosphorylation is magnified in Figure 5.8. Both spectra contain a doubly-charged y_{11} ion at m/z 635.35. However, whereas the y_{12}^{2+} ion in the spectrum of the unmodified peptide appears at m/z 678.88, in the phosphorylated spectrum a doubly-charged peak at m/z 718.84 is observed instead. Also a doubly-charged peak at m/z 669.89 is observed, which corresponds to the loss of H_3PO_4 from the phosphorylated y_{12} ion. Hence, this fragmentation spectrum establishes serine 224 as a previously unreported site of phosphorylation of SRF.

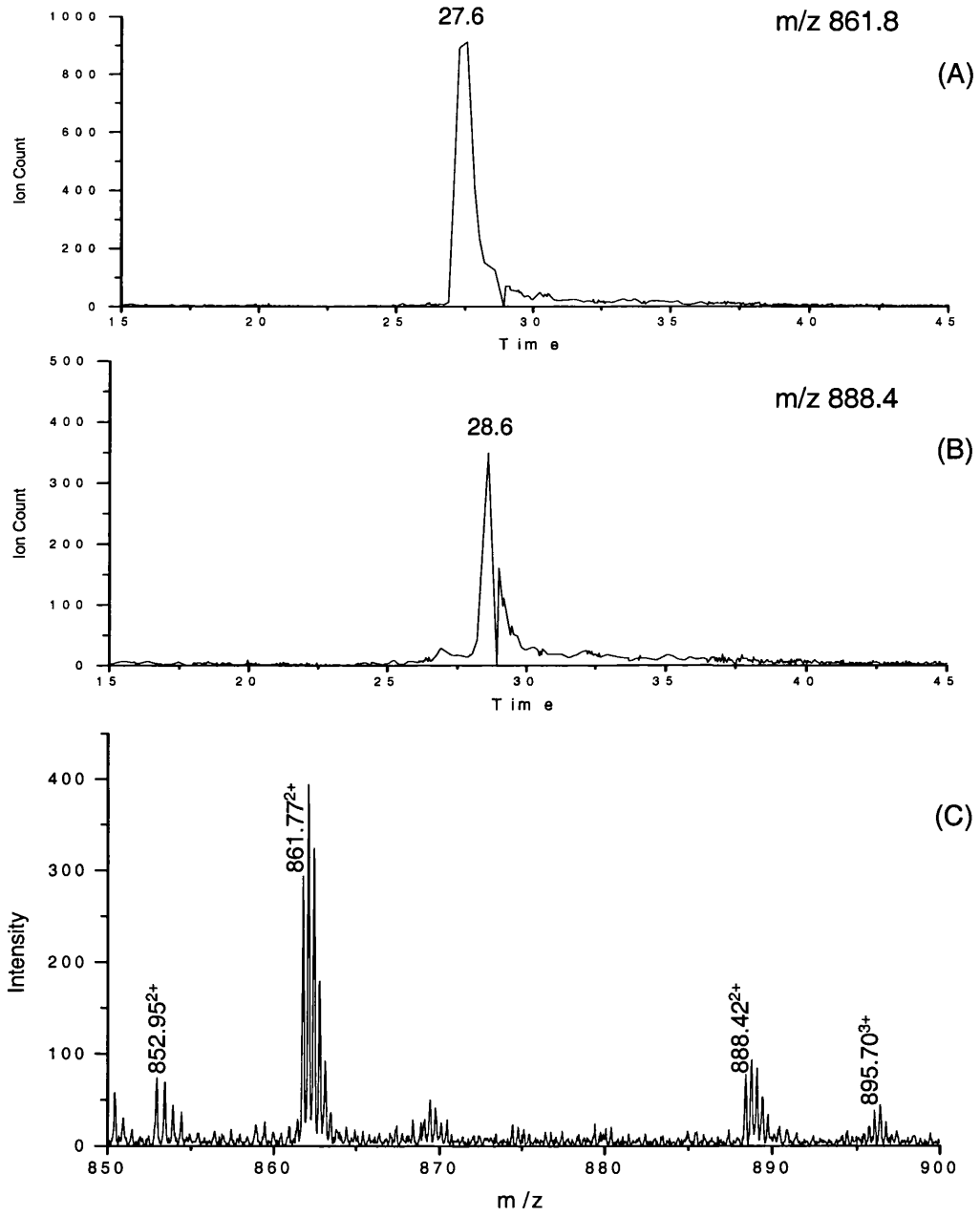


Figure 5.6: Extracted ion chromatograms of unmodified and phosphorylated versions of the tryptic peptide spanning residues 213 – 235 of SRF.

(A) Ion chromatogram of the unmodified peptide. (B) Ion chromatogram of the phosphorylated peptide. (C) The combined spectrum during the period in which the unmodified and phosphorylated peptides elute contains the unmodified peptide at $[M + 3H]^{3+}$ m/z 861.77 and the phosphorylated peptide at $[M + 3H]^{3+}$ m/z 888.42.

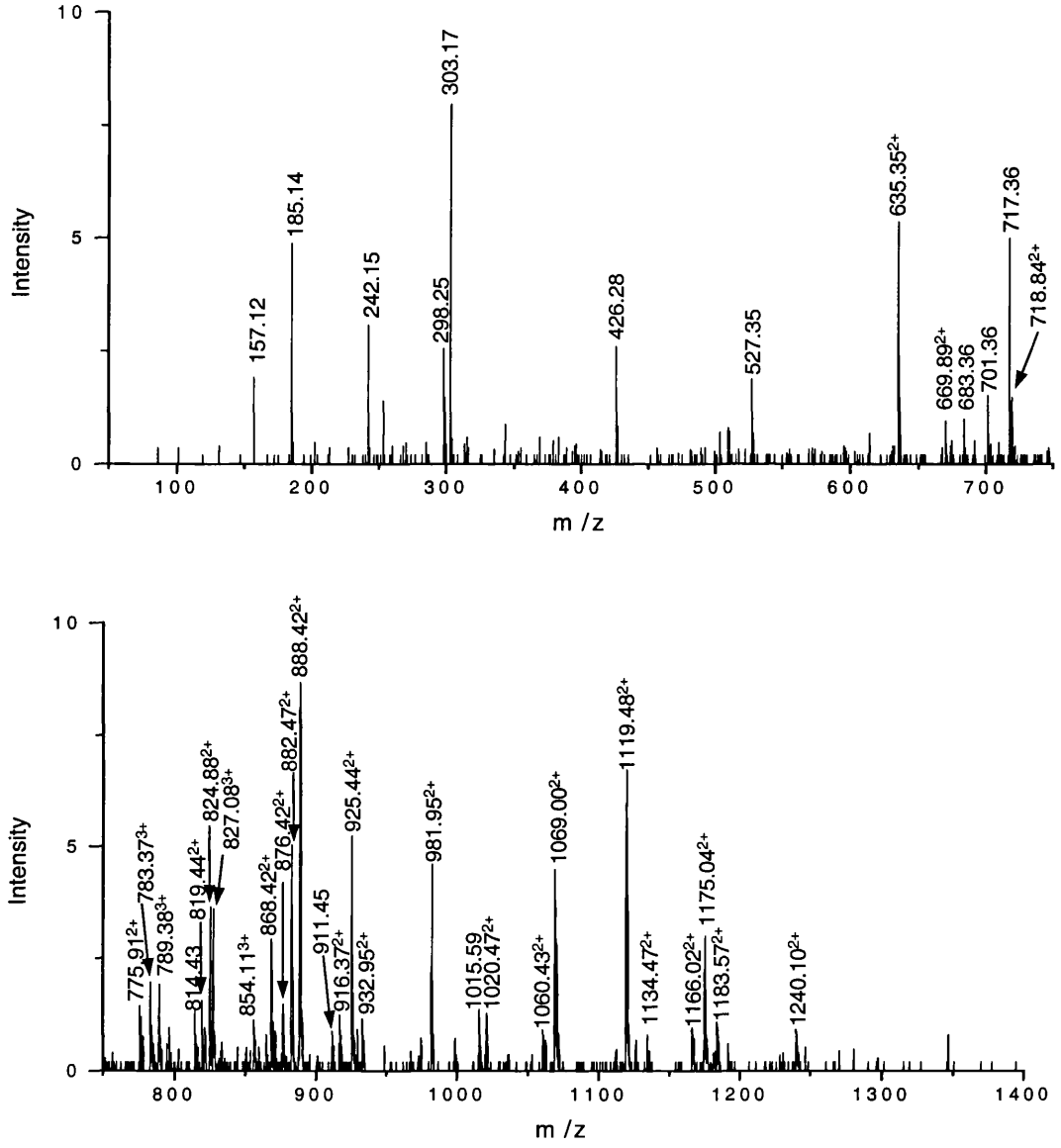


Figure 5.7: ESI-CID-MS-MS spectrum of the phosphorylated peptide [M + 3H]$^{3+}$ m/z 888.42 from a tryptic digest of SRF.

Identities of observed fragment ions are listed in Table 5.5.

Table 5.5: Identities of peaks observed in the ESI-CID-MS-MS spectrum of the phosphorylated peptide $[M + 3H]^{3+}$ m/z 888.42.

Peak	Match	Peak	Match	Peak	Match
157.12	a2	783.37^{3+}	y20-NH_3	932.95^{2+}	y17-H_3PO_4
185.14	b2	789.38^{3+}	y20	981.95^{2+}	y17
242.15	IQ	814.43	b7	1015.59	b9
298.25	b3	819.44^{2+}	y15-H_3PO_4	1020.47^{2+}	y18-H_3PO_4
303.17	y2	824.88^{2+}	y14	1060.43^{2+}	y18-NH_3
426.28	b4	827.08^{3+}	y21	1069.00^{2+}	y18
527.35	b5	854.11^{3+}	MH_3^{3+}-H_3PO_4	1119.48^{2+}	y19
635.35^{2+}	y11	868.42^{2+}	y15	1134.47^{2+}	b20-H_3PO_4
669.89^{2+}	y12-H_3PO_4	876.42^{3+}	MH_3^{3+}-$2H_2O$	1166.02^{2+}	y20-NH_3-H_2O
683.36	b6-H_2O	882.47^{3+}	MH_3^{3+}-H_2O	1175.04^{2+}	y20-NH_3
701.36	b6	888.42^{3+}	MH_3^{3+}	1183.57^{2+}	y20
717.36	y6	911.45	b8-NH_3	1240.10^{2+}	y21
718.84^{2+}	y12	916.37^{2+}	b16		
775.91^{2+}	y13	925.44^{2+}	y16		

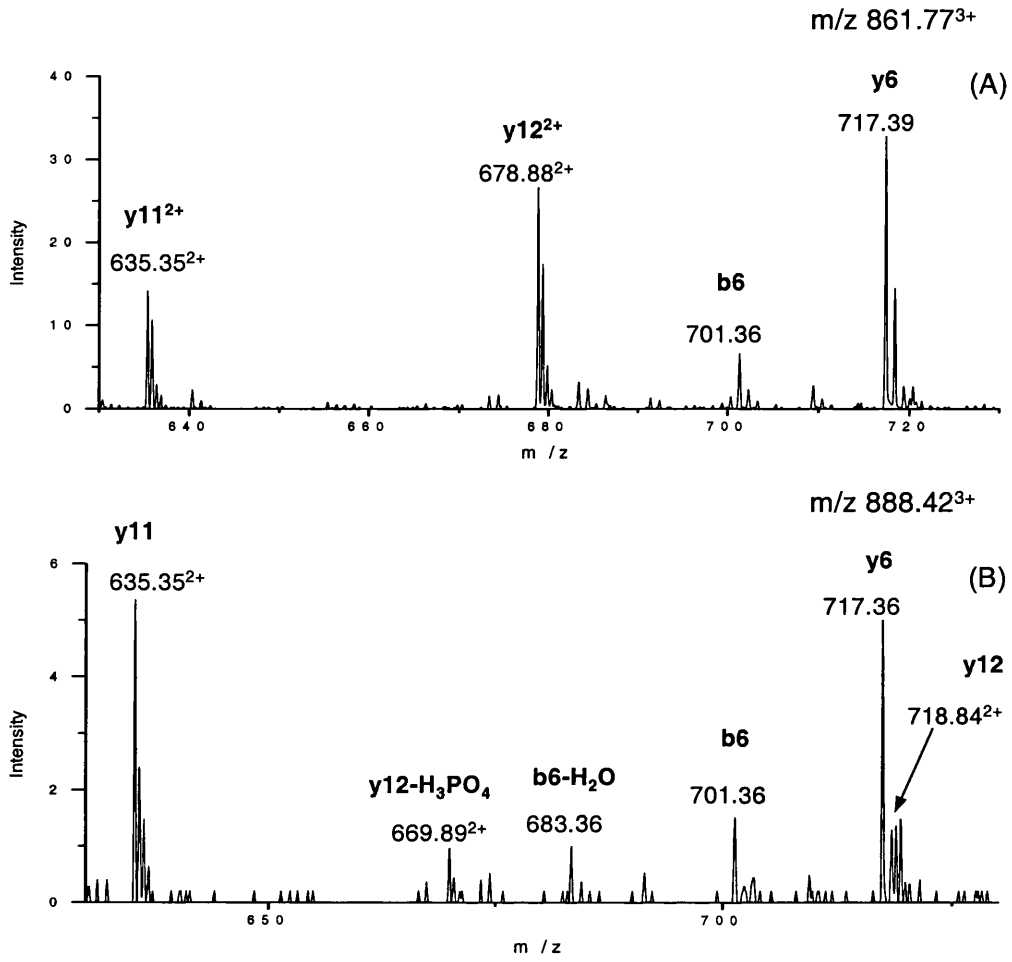


Figure 5.8: Magnified region of ESI-CID-MS-MS spectra of unmodified and phosphorylated versions of the peptide spanning residues 213 – 235 of SRF.

In both (A) unmodified and (B) phosphorylated spectra y_{11} ions appear at m/z 635.35, but the y_{12} ion in the phosphopeptide spectrum appears at m/z 718.84, instead of m/z 678.88 in the spectrum from the unmodified peptide, signifying serine 223 is phosphorylated.

Another peptide was observed in the LC-MS run at $[M + 3H]^{3+}$ m/z 1092.15, and was automatically selected for CID fragmentation (Figure 5.9). This peptide peak is a phosphorylated version of the peptide spanning residues 101 – 135 of SRF, and the identities of the fragment ions observed are in Table 5.6. This is not a predicted tryptic fragment, as although residue 135 is a lysine residue, residue 136 is a proline, and trypsin is not predicted to cleave after lysines that are followed by prolines [193]. The peptide contains a known phosphorylation site at residue 103 [185-187]. None of the 'y' ions are observed in a phosphorylated state, but all 'b' ions observed are either phosphorylated or have lost H_3PO_4 to produce an ion 18Da smaller than an unmodified fragment. The smallest phosphorylated 'b' ion is the b_6 ion at m/z 757.42. This defines the phosphorylation to be on either serine 101 or serine 103. b_1 and b_2 ions are not observed, so there is no data in this spectrum to differentiate between the two sites. However, serine 103 is more likely, as trypsin is unlikely to cleave after arginine 100 if serine 101 was phosphorylated, due to the steric hindrance of the bulky phosphate group. Figure 5.10 shows the extracted ion chromatograms for the phosphopeptide at m/z 1092.3 and its unmodified counterpart at m/z 1065.5. The combined spectrum over the period these peaks eluted shows the modified peak at $[M + 3H]^{3+}$ m/z 1092.15 is significantly more intense than the unmodified $[M + 3H]^{3+}$ m/z 1065.50, suggesting a high stoichiometry of modification at this site.

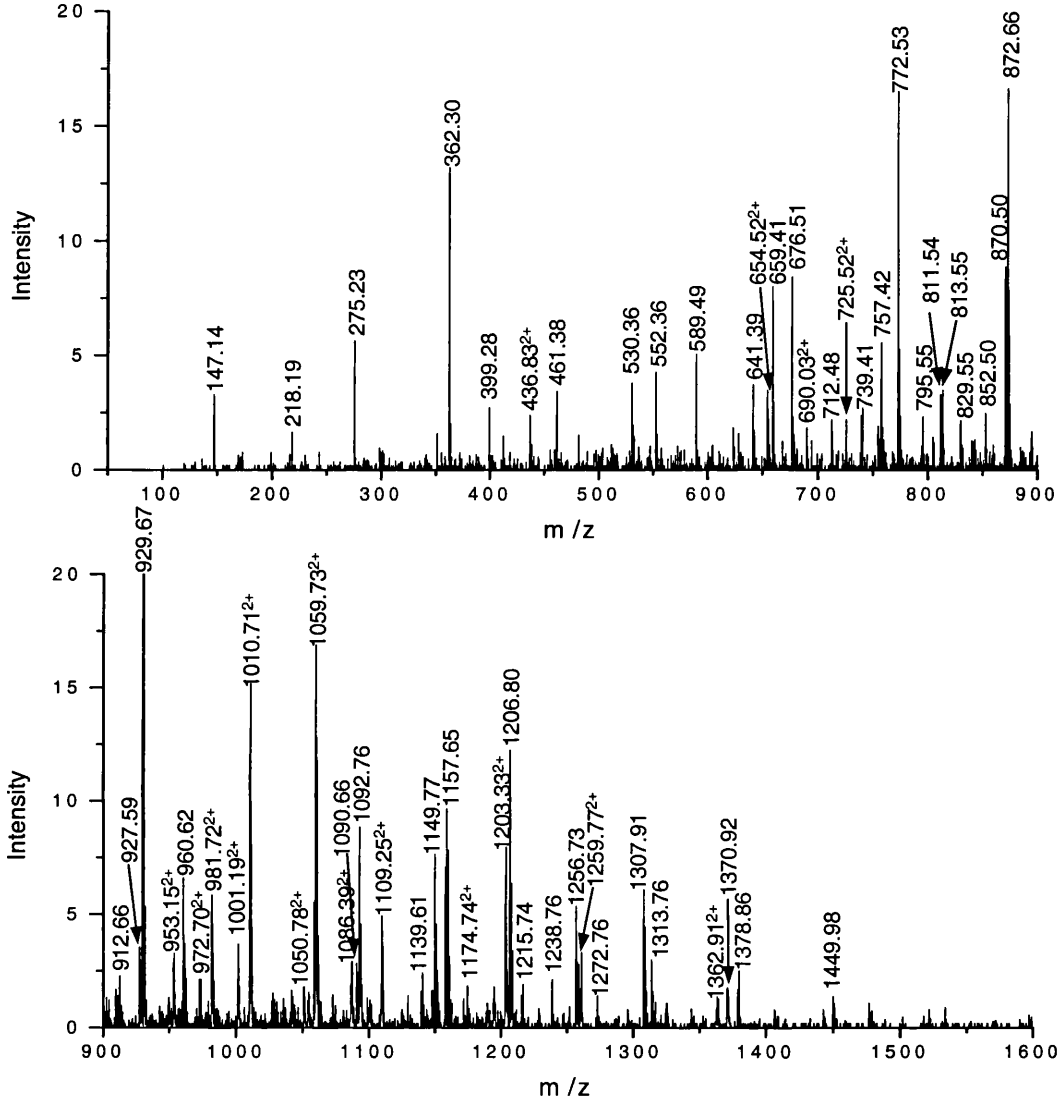


Figure 5.9: ESI-CID-MS-MS spectrum of [M + 3H]$^{3+}$ m/z 1092.37 from a tryptic digest of SRF.

This peptide is residues 101-135 of SRF, and the identities of peaks observed are given in Table 5.6.

Table 5.6: Identities of the peaks observed in the ESI-CID-MS-MS spectrum of the phosphorylated peptide [M + 3H]$^{3+}$ m/z 1092.3.

This fragmentation spectrum is shown in Figure 5.9, and is of the peptide spanning residues 101 – 135 of SRF.

Peak	Match	Peak	Match	Peak	Match
147.14	y1	795.55	GGPEASAAAT-H_2O	1092.76	y12
218.19	y2	811.54	VGGPEASAAA	1109.25^{2+}	y26
275.23	y3	813.55	GGPEASAAAT	1139.61	b10-H_2O
362.30	y4	829.55	b8-H_3PO_4	1149.77	y13
399.28	b4-H_3PO_4	852.50	b7- H_2O	1157.65	b10
436.83^{2+}	y10	870.50	b7	1174.74^{2+}	y27
461.38	y5	872.66	y10	1203.33^{2+}	y28
530.36	b5- H_3PO_4	912.66	y11-NH_3	1206.80	y14
552.36	GGPEASA-H_2O	927.59	b8	1215.74	b12-H_3PO_4
589.49	y7	929.67	y11	1238.76	b11-H_2O
641.39	b6-H_3PO_4-H_2O	953.15^{2+}	y22	1256.73	b11
654.52^{2+}	y15	960.62	b9-H_3PO_4	1259.77^{2+}	y29
659.41	b6-H_3PO_4	972.70^{2+}	y23-H_2O	1272.76	b13-H_3PO_4
676.51	y8	981.72^{2+}	y23	1307.91	y15
690.03^{2+}	y16	1001.19^{2+}	y24-H_2O	1313.76	b12
712.48	GGPEASAAA	1010.71^{2+}	y24	1362.91^{2+}	b29-H_2O
725.52^{2+}	y17	1050.78^{2+}	y25-H_2O	1370.92	b13
739.41	b6-H_2O	1059.73^{2+}	y25	1378.86	y16
757.42	b6	1086.39^{2+}	MEIGMVVGGPEASAA ATGGY	1449.98	y17
772.53	b7-H_3PO_4	1090.66	GGPEASAAATGGY		

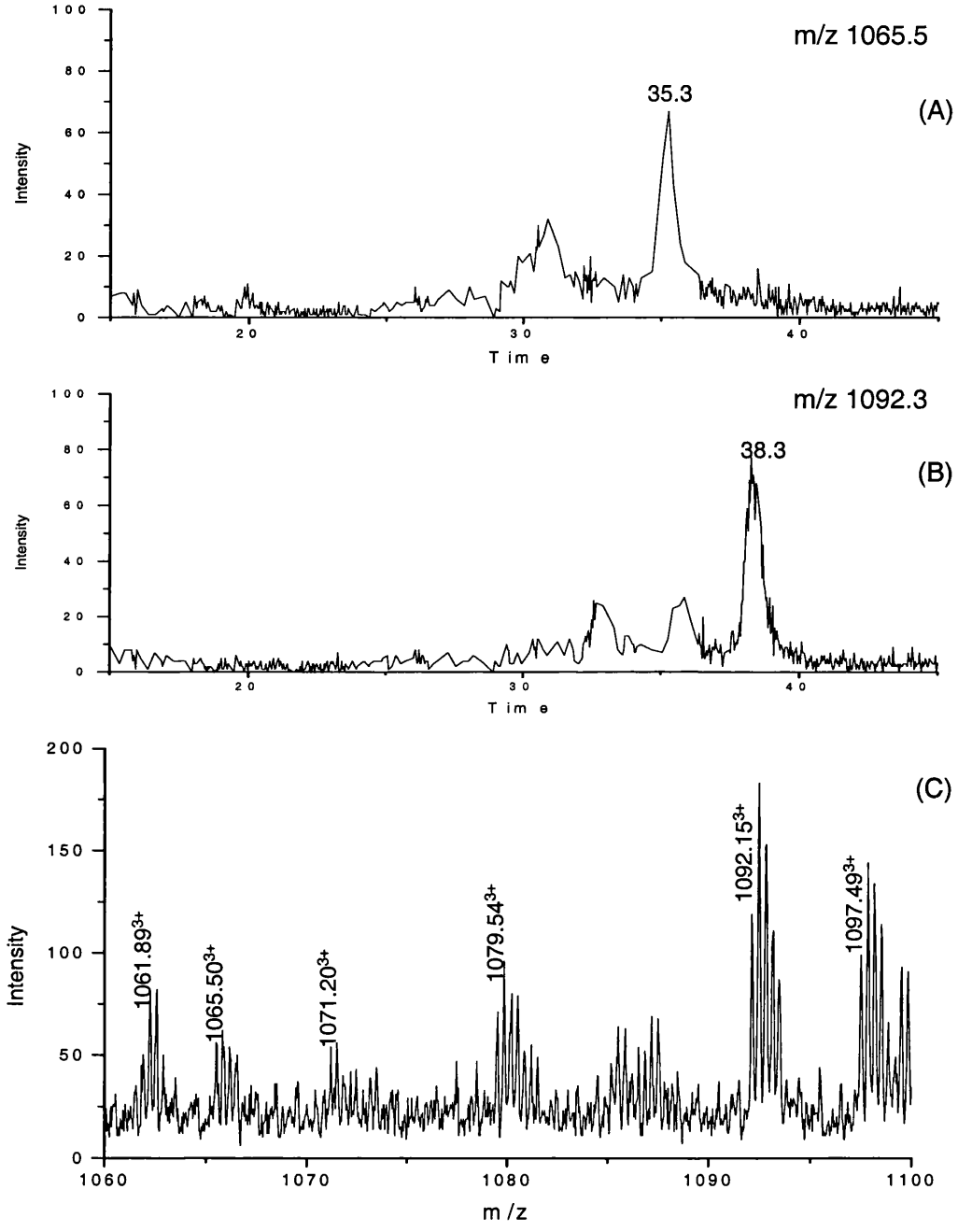
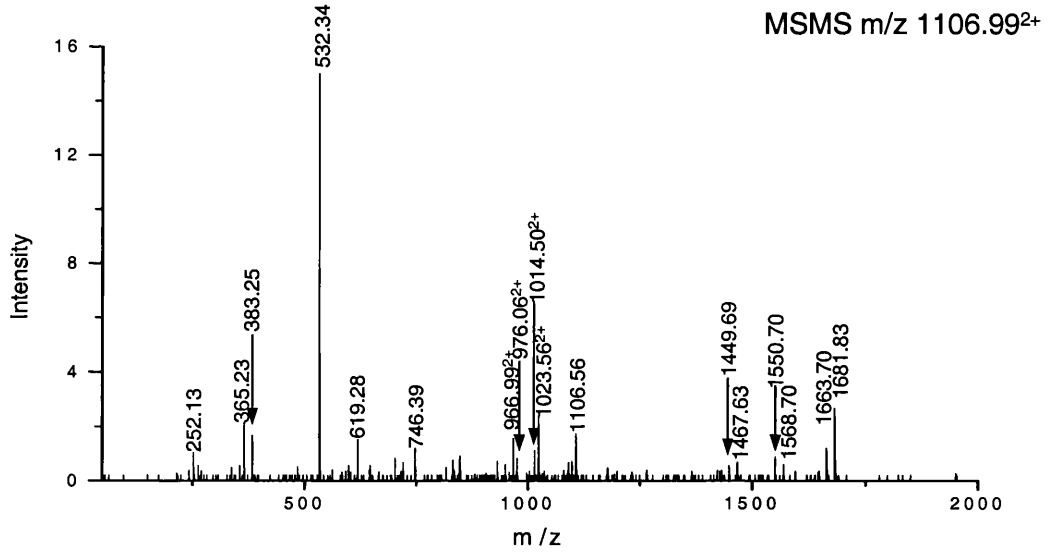


Figure 5.10: Extracted ion chromatograms of unmodified and phosphorylated versions of the tryptic peptide spanning residues 101 - 135 of SRF.

Extracted ion chromatograms of (A) the unmodified peptide and (B) the GlcNAc-modified peptide. (C) The combined mass spectrum of the period during which the unmodified and phosphorylated peptides elute, contains the unmodified peptide at $[M + 2H]^{2+}$ *m/z 1065.50 and the phosphorylated peptide at* $[M + 2H]^{2+}$ *m/z 1092.15.*



Rank	# Unmatched Ions	Sequence					MH⁺ Calculated (Da)	MH⁺ Error (Da)	Protein MW (Da)/pI	Species	SwissProt.7.7.2001 Accession #	MS-Digest Index #	Protein Name
1	2/17	(R)DSSTDLTQTSSSGTVTLPATIM(T)					2213.0495	-0.0773	51592.9 / 7.83	HUMAN	P11831	17568	SERUM RESPONSE FACTOR (SRF)

	Fragment-ion (m/z)	252.13	365.23	383.25	532.34	619.28	746.39	1106.56	1449.69	1467.63	1550.70	1568.70	1663.70	1681.83	966.99^{+2}	976.06^{+2}	1014.50^{+2}	1023.56^{+2}
	Fragment-ion (m/z)	252.13	365.23	383.25	532.34	619.28	746.39	1106.56	1449.69	1467.63	1550.70	1568.70	1663.70	1681.83	1932.97	1951.11	2027.99	2046.11
	Ion-type Delta ppm		PATI-H_2O 30.4	PATI 53.6	y_5 111.8	b_6 36.3	y_7 -29.8	b_{11}-NH_3 100.4	b_{15}-H_2O 38.3	b_{18} -10.3	b_{16}-H_2O 11.5	b_{16} 4.6	b_{17}-H_2O -39.8	b_{17} 31.6	b_{20}-H_2O^{+2} 35.4	b_{20}^{+2} 101.4		

Figure 5.11: ESI-CID-MS-MS spectrum of [M + 2H]$^{2+}$ m/z 1106.99 from tryptic SRF.

Fragment ions observed were entered into MS-TAG, and the search result presented identifies the peptide as residues 374 – 395 of SRF. This is a product of a non-specific cleavage at the C-terminus.

An peak unassigned by MASCOT at [M + 2H]$^{2+}$ m/z 1106.99 from the LC-MS run was selected for CID-MS-MS (Figure 5.11). The peaks observed were entered for a MS-TAG search using no enzyme specificity for peptide cleavage [44]. The result identifies the peptide as residues 374 – 395 of SRF (Figure 5.11). The N-terminal cleavage is trypsin specific, but the C-terminal cleavage is after a methionine. This peptide contains a published GlcNAc modification site at residue 383 [92]. Hence, the survey mass spectrum was examined to see if a peak corresponding to a GlcNAc-modified version of this peptide was detected. Figure 5.12 shows extracted ion chromatograms for the doubly-charged peak and the theoretical mass of a doubly-charged GlcNAc-modified

version of this peptide. Each ion chromatogram contains two peaks. The first peak in the extracted ion chromatogram of m/z 1107.5, at time 30.9 minutes, relates to a peak at $[M + 2H]^{2+}$ m/z 1106.05, which is a trypsin autolysis peak. The second peak in the chromatogram is at 37.3 minutes corresponds to the peptide spanning 374 – 395, and a peak is observed just before this (after 36.7 minutes) in the extracted ion chromatogram of the mass of the potential GlcNAc-modified peptide (m/z 1209.0). A sugar residue makes a peptide more hydrophilic, so would cause a glycosylated peptide to elute earlier. The peak in the chromatogram of m/z 1209.0 after 33.0 minutes is from an unrelated doubly-charged peptide of monoisotopic mass 1208.09.

The combined spectrum over this region is shown in Figure 5.12. The peak at $[M + H]^{2+}$ m/z 1208.54 is the correct mass for the GlcNAc-modified peptide. This peak was specifically selected for ESI-CID-MS-MS in a further LC-MS run. The resulting spectrum is shown in Figure 5.13, while peak identities are given in Table 5.7. A prominent peak at m/z 204.10 is observed, confirming this is a glycopeptide, and most ions are assigned to predicted fragments. Unfortunately, none of the fragment ions are glycosylated, thus providing no information on the site of modification.

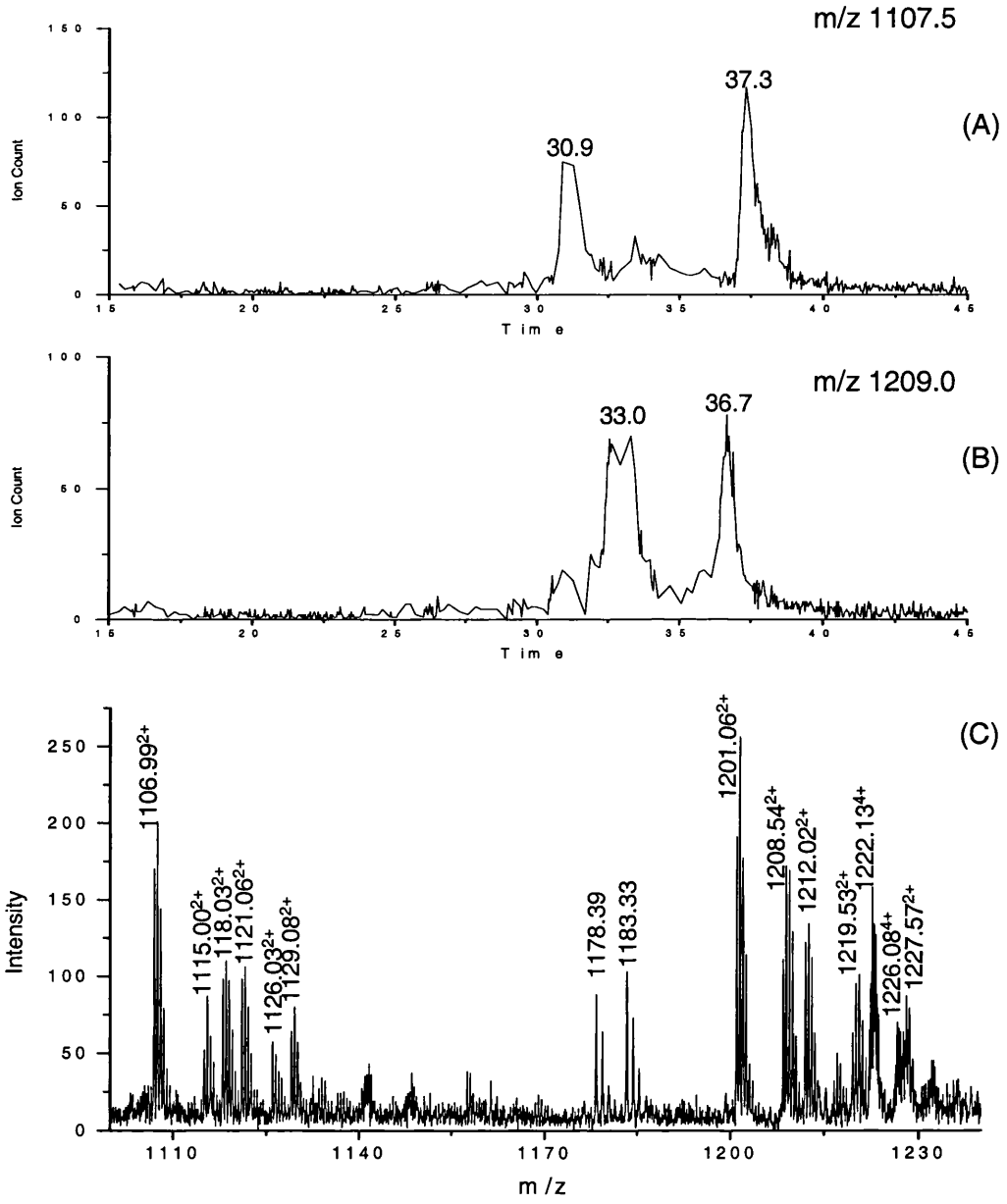


Figure 5.12: Extracted ion chromatograms of unmodified and GlcNAc-modified versions of the peptide spanning residues 374 – 395 of SRF.

Extracted ion chromatograms of (A) the unmodified peptide and (B) the GlcNAc-modified peptide. (C) The combined mass spectrum during the period in which the unmodified and GlcNAc-modified peptides elute contains the unmodified peptide at $[M + 2H]^{2+}$ m/z 1106.99 and the glycosylated peptide at $[M + 2H]^{2+}$ m/z 1208.54.

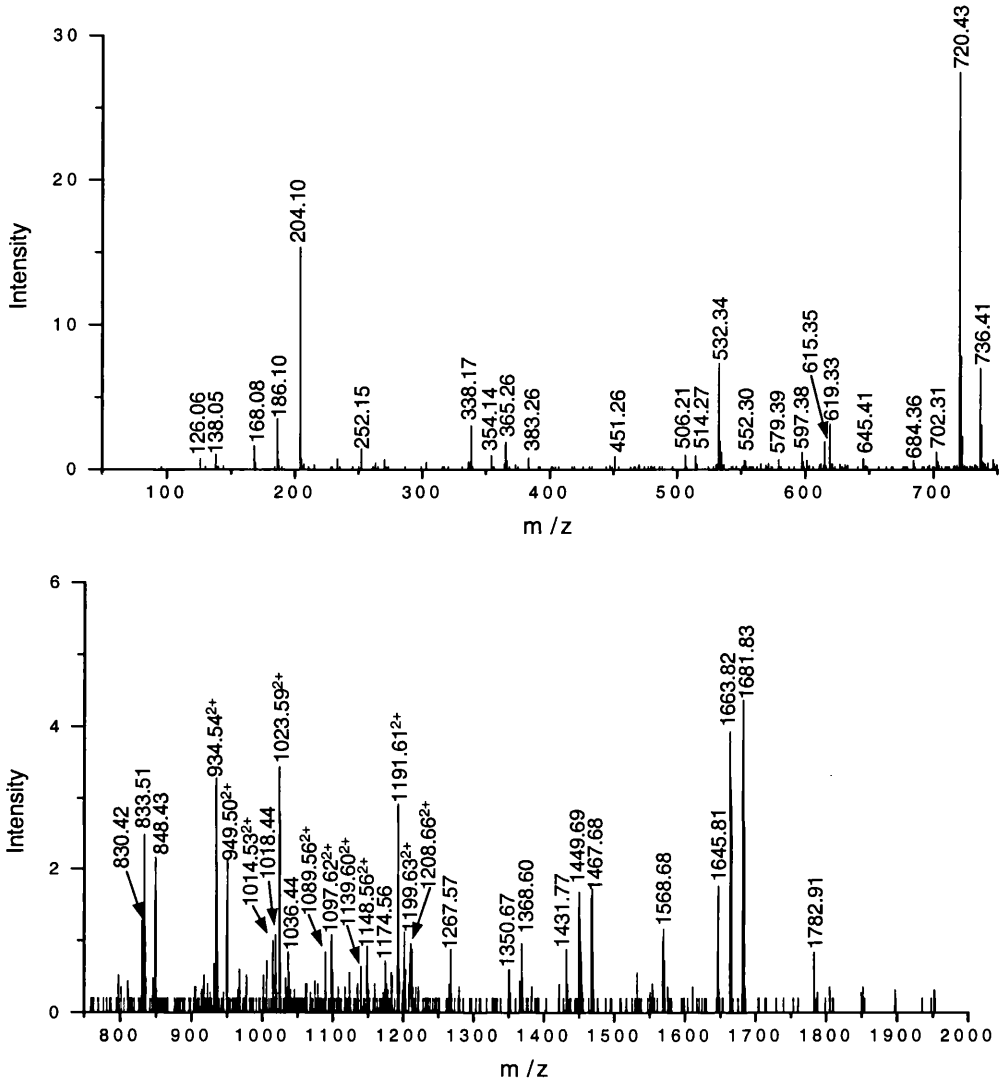


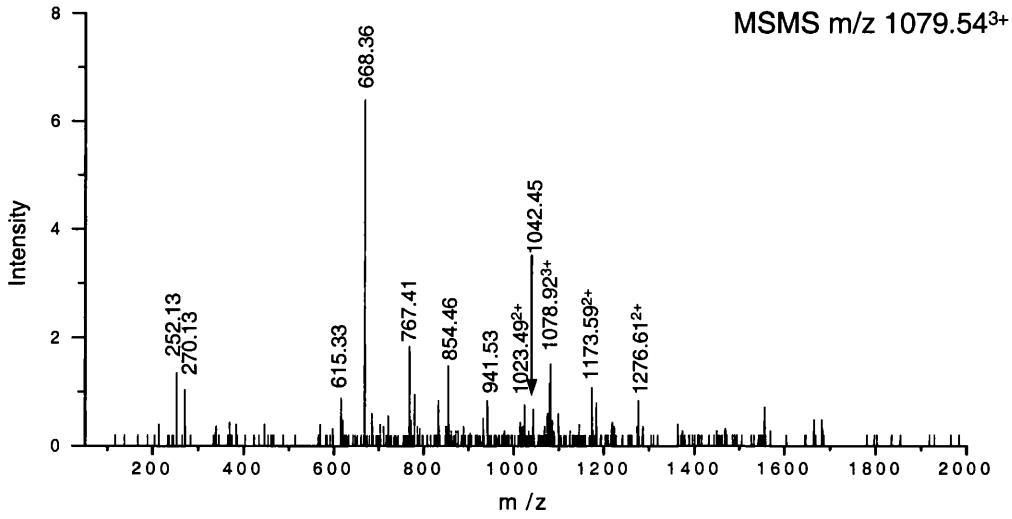
Figure 5.13: ESI-CID-MS-MS spectrum of the GlcNAc-modified peptide [M + 2H]$^{2+}$ m/z 1208.54 from a tryptic digest of SRF.

This peptide spans residues 374 - 395. Identities of observed peaks are given in Table 5.7.

Table 5.7: Identities of the peaks observed in the ESI-CID-MS-MS spectrum of the GlcNAc-modified peptide [M + 2H]$^{2+}$ m/z 1208.54.

The fragmentation spectrum is Figure 5.13. This peptide spans residues 374 – 395 of SRF.

Peak	Match	Peak	Match	Peak	Match
126.06	GlcNAc Fragment	645.41	y6	1191.61^{2+}	$M_GH_2^{2+}$-2H_2O
138.05	GlcNAc Fragment	684.36	b7-2H_2O	1199.63^{2+}	$M_GH_2^{2+}$-H_2O
168.08	GlcNAc-2H_2O	702.31	b7-H_2O	1208.66^{2+}	$M_GH_2^{2+}$
186.10	GlcNAc-2H_2O	720.43	b7	1267.57	b13
204.10	GlcNAc	736.41		1350.67	b14-H_2O
252.15	PAT	830.42	b8-H_2O	1368.60	b14
338.17		833.51		1431.77	b15-2H_2O
354.14		848.43	b8	1449.69	b15-H_2O
365.26	PATI-H_2O	934.54	QTSSSGTVTL-28	1467.68	b15
383.26	PATI	949.50	b9	1568.68	b16
451.26		1014.53^{2+}	b21-2H_2O	1645.81	b17-2H_2O
506.21	b5	1018.44	b10-H_2O	1663.82	b17-H_2O
514.43	y5-H_2O	1023.59^{2+}	b21-H_2O	1681.83	b17
532.34	y5	1036.44	b10	1782.91	
579.39	TLPATI-H_2O	1089.56^{2+}	MH_2^{2+}-2H_2O		
597.38	TLPATI	1097.62^{2+}	MH_2^{2+}-H_2O		
615.35		1148.56^{2+}			
619.33	b6	1174.56			



# Unmatched Ions	Sequence	MH$^+$ Calculated (Da)	MH$^+$ Error (Da)	Protein MW (Da)/pI	Species	PN.SRF Accession #	MS-Digest Index #	Protein Name
1/11	(R)DSSTDLTQTSSSGTVTLP ATIMTSSVPTTVGGH(M)	3236.5480	0.0563	51593.9 / 7.08	HUMAN	11831	1	Serum Response Factor

Fragment-ion (m/z)	252.13	270.13	615.33	668.36	767.41	854.46	941.53	1042.45	1023.49^{+2}	1173.59	1276.61^{+2}
Fragment-ion (m/1)	252.13	270.13	615.33	668.36	767.41	854.46	941.53	1042.45	2045.97	1173.59	2552.21
Ion-type Delta ppm	PAT-H_2O -19.1	y_3 36.2 PAT -56.9	PATIM T 20.1	y_7 34.8	y_8 6.3	y_9 26.7	y_{10} 64.5	y_{11} -64.2	b_{21}-H_2O^{+2} -7.7	y12 27.8	b_{26}-NH_3^{+2} 7.7

Figure 5.14: ESI-CID-MS-MS spectrum of [M + 3H]$^{3+}$ m/z 1079.54, from a tryptic digest of SRF.

Observed peaks were entered into MS-TAG and the search result displayed identifies this peptide as residues 374 – 406 of SRF.

Another unidentified peptide automatically selected for ESI-CID-MS-MS analysis was observed at [M + 3H]$^{3+}$ m/z 1079.54. Its spectrum is shown in Figure 5.14, along with the MS-TAG result using the peaks labelled in the spectrum. MS-TAG identifies the peptide as residues 374 – 406 of SRF. The N-terminus of this peptide is a predicted tryptic cleavage site, but the C-terminus is formed by a cleavage after a histidine, which is a non-specific cleavage. This peptide is a longer version of the [M + 2H]$^{2+}$ m/z 1106.99 peptide previously analysed in Figure 5.11. Hence, this peptide contains a GlcNAc modification site, and a glycosylated version of this peptide should also have been observed.

Figure 5.15 presents extracted ion chromatograms for the triply-charged unmodified peptide and for singly and doubly GlcNAc-modified versions. Peaks are observed in all these chromatograms, and the retention times become slightly earlier with increasing numbers of modifications, consistent with GlcNAc-modified peptides. The mass spectrum over the period these three peaks elute contains peaks at $[M + 3H]^{3+}$ m/z 1147.19 and $[M + 3H]^{3+}$ m/z 1214.93 corresponding in mass with singly and doubly GlcNAc-modified versions of the peptide (Figure 5.15D). Both modified peptides were specifically selected for ESI-CID-MS-MS analysis in a further LC-MS run. The fragmentation spectrum of $[M + 3H]^{3+}$ m/z 1147.19 is presented in Figure 5.16, whilst the peaks are identified in Table 5.8. The intense peaks in this fragmentation spectrum are the triply-charged glycosylated parent ion at m/z 1147.18, the doubly-charged 'deglycosylated' parent ion at m/z 1618.80 and the singly-charged GlcNAc oxonium ion at m/z 204.09. GlcNAc-modified versions of b26 fragment ions are observed, and a weak ion corresponding to $b21_G$ is seen at $[M + 2H]^{2+}$ m/z 1125.04. This fragment ion determines the glycosylation is not in the 12 most C-terminal residues, but cannot determine the exact site of modification. This data is consistent with the published site serine 383 being modified [92], and suggests that the modified residue in this peptide is the same as in the shorter peptide observed at $[M + 2H]^{2+}$ m/z 1208.54.

The ESI-CID-MS-MS spectrum of $[M + 3H]^{3+}$ m/z 1214.93 is shown in Figure 5.17 and the identities of the peaks in this spectrum are Table 5.9. The major fragment ions in this spectrum are doubly-charged peaks corresponding to the loss of one and two GlcNAc residues from the parent ion, at $[M + 2H]^{2+}$ m/z 1720.27 and $[M + 2H]^{2+}$ m/z 1618.80 respectively, and a prominent GlcNAc oxonium ion at m/z 204.09. Unfortunately, this spectrum is of low intensity, and there are no glycosylated fragment ions. Nevertheless, it does confirm this is a doubly GlcNAc-modified peptide.

Thus, there is a previously undetected site of GlcNAc modification in the longer peptide spanning residues 374 to 406, which is not present in the shorter peptide of residues 374 – 395, locating the undetected site of GlcNAcylation in the region spanning residues 396 – 406. This region has the sequence TSSVPTTVGGHM, and contains a number of potential residues that could bear the modification.

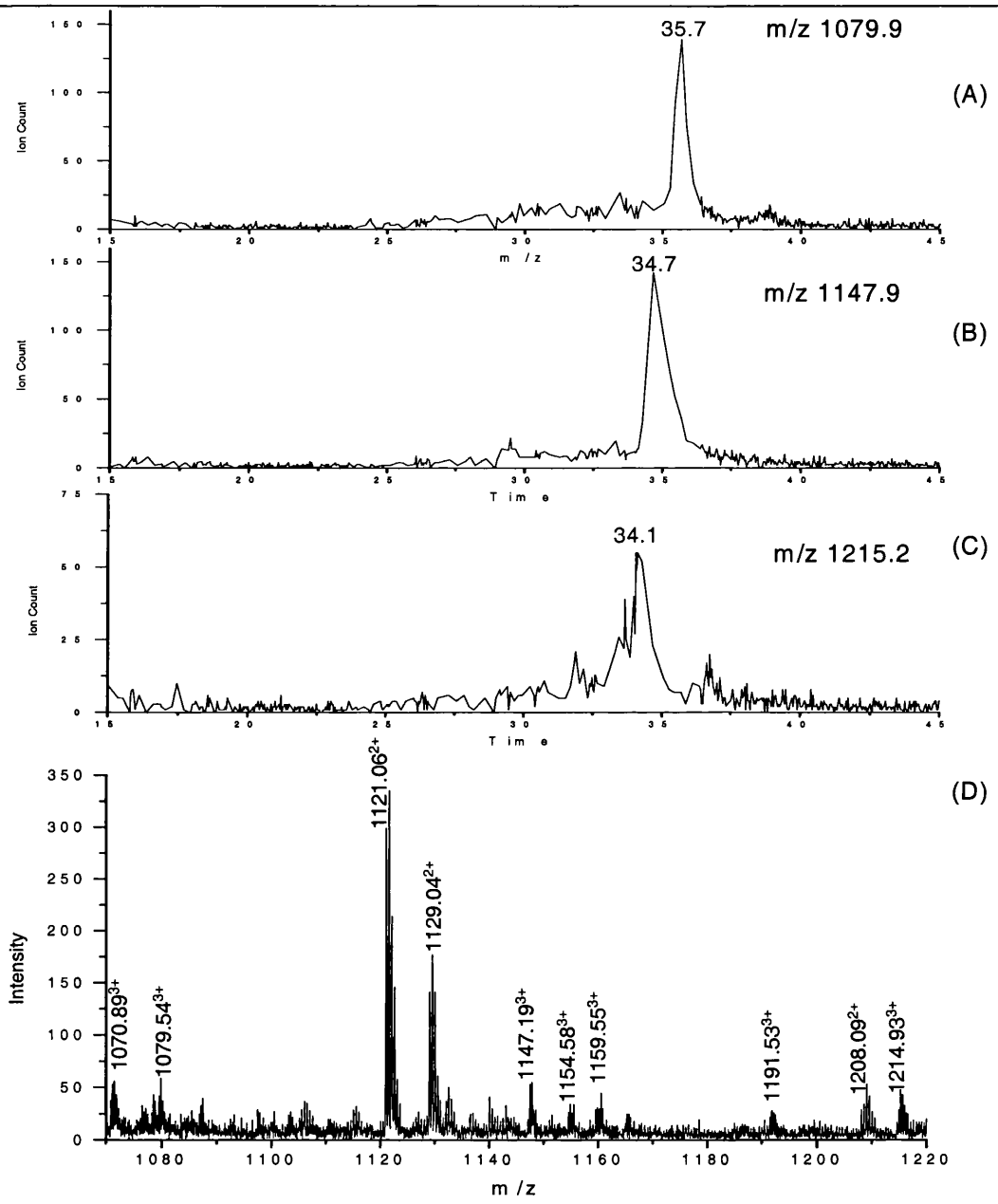


Figure 5.15: Extracted ion chromatograms of the unmodified, singly GlcNAc-modified and doubly GlcNAc-modified peptide spanning residues 374 - 406 of SRF.

Extracted ion chromatograms of (A) unmodified, (B) singly GlcNAc-modified and (C) doubly GlcNAc-modified peptide. (D) The combined mass spectrum during the period in which the these three peaks elute contains the unmodified peptide at $[M + 3H]^{3+}$ m/z 1079.54, the singly GlcNAc-modified peptide at $[M + 3H]^{3+}$ m/z 1147.19 and the doubly GlcNAc-modified peptide at $[M + 3H]^{3+}$ m/z 1214.93.

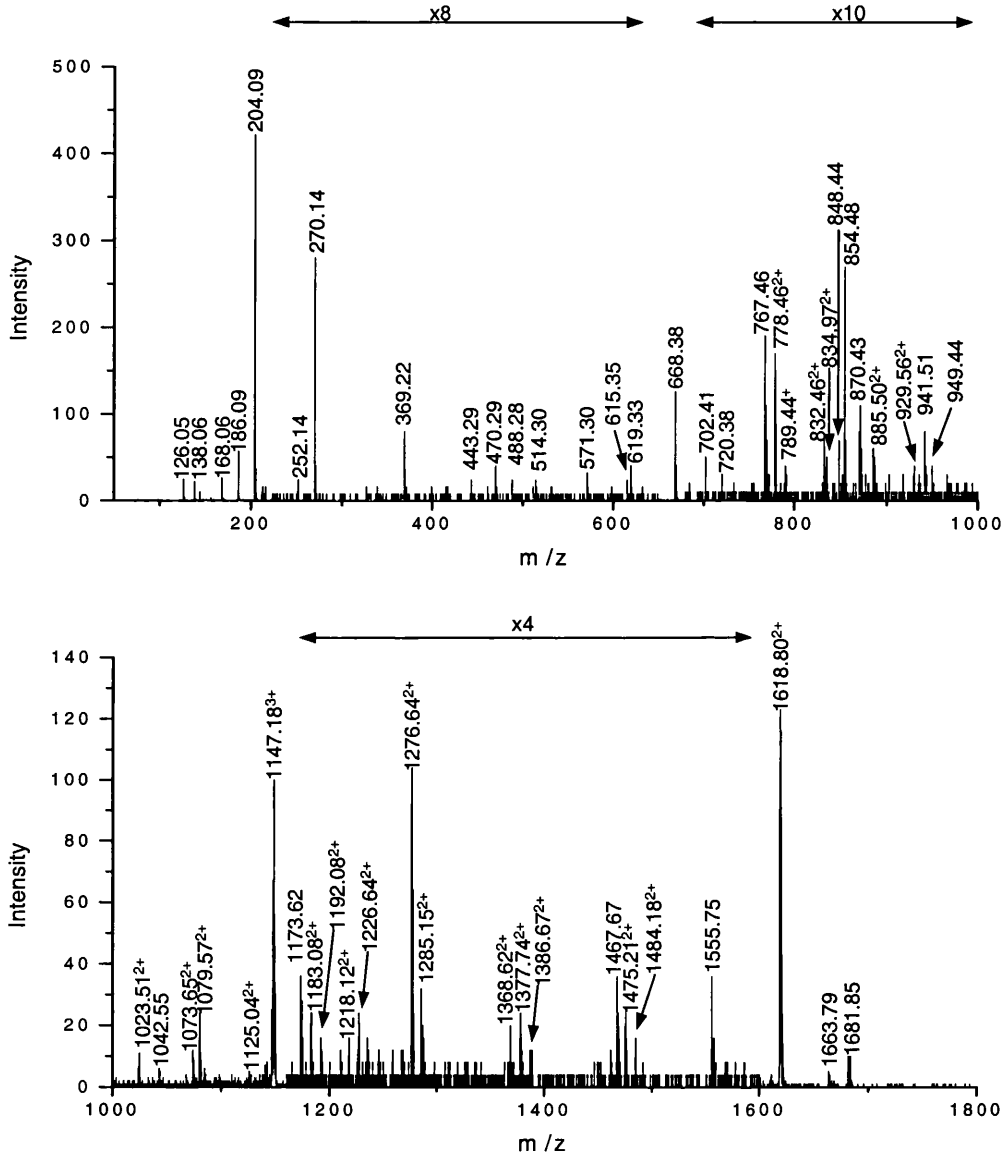
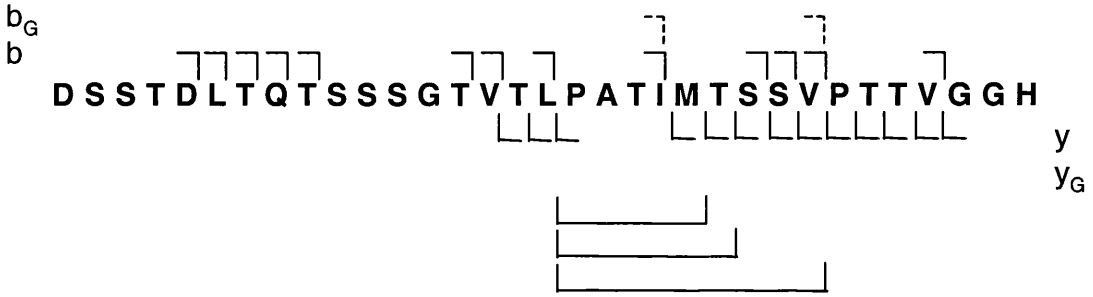


Figure 5.16: ESI-CID-MS-MS spectrum of the singly GlcNAc-modified peptide [M + 3H]$^{3+}$ m/z 1147.90.

This peptide corresponds to residues 374 – 406. The identities of peaks and a graphical representation of fragment ions observed are given in Table 5.8.

Table 5.8: Identities of peaks observed in the ESI-CID-MS-MS spectrum of the singly GlcNAc-modified peptide $[M + 3H]^{3+}$ m/z 1147.19.

The fragmentation spectrum is Figure 5.16, and is of residues 374 – 406 of SRF. A graphical representation of the fragment ions observed is also presented.



Peak	Match	Peak	Match	Peak	Match
126.05	GlcNAc fragment	767.46	y8	1183.08^{2+}	b24-H_2O
138.06	GlcNAc fragment	778.46^{2+}	y16	1192.08^{2+}	b24
168.06	GlcNAc fragment	789.44		1218.12^{2+}	b25-$2H_2O$
186.09	GlcNAc fragment	832.46^{2+}	b17-H_2O	1226.64^{2+}	b25-H_2O
204.09	GlcNAc	834.97^{2+}	y17	1276.64^{2+}	b26-H_2O
252.14	y3-H_2O	848.44	b8	1285.15^{2+}	b26
270.14	y3	854.48	y9	1368.62^{2+}	b14
369.22	y4	870.43	PATIMTSSV-H_2O	1377.74^{2+}	$b26_G$-H_2O
443.29		885.50^{2+}	y18	1386.67^{2+}	$b26_G$
470.29	y5	929.56^{2+}		1467.67	b15
488.28	b5-H_2O	941.51	y10	1475.21^{2+}	b30-H_2O
514.30	PATIM	949.44	b9	1484.18^{2+}	b30
571.30	y6	1023.51^{2+}	b21-H_2O	1555.75	y16
615.35	PATIMT	1042.55	y11	1618.80^{2+}	MH^{2+}
619.33	b6	1073.65^{2+}	MH^{3+}-H_2O	1663.79	b17-H_2O
668.38	y7	1079.57^{2+}	MH^{3+}	1681.85	b17
702.41	b7-H_2O	1125.04^{2+}	$b21_G$-H_2O		
720.38	b7	1173.62	y12		

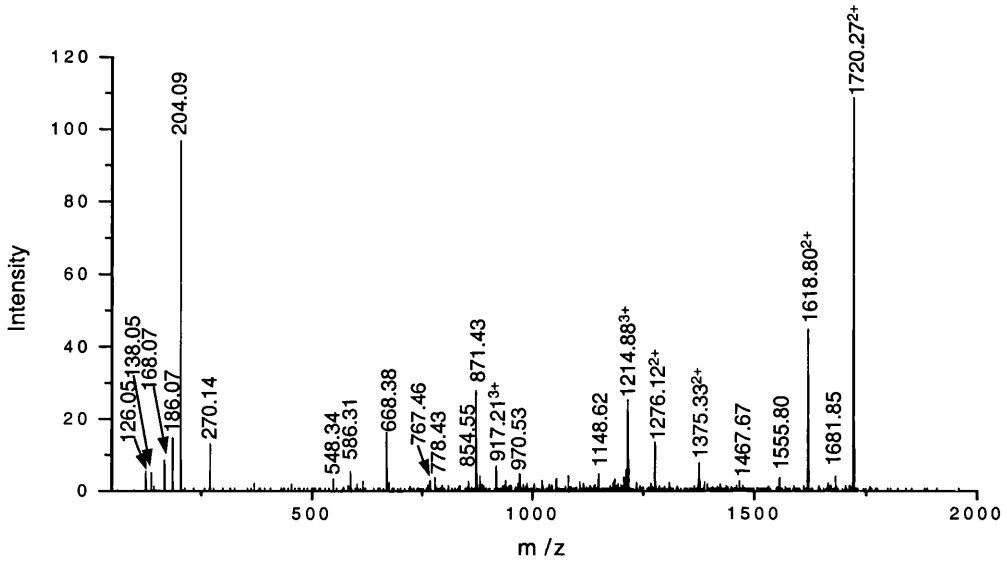


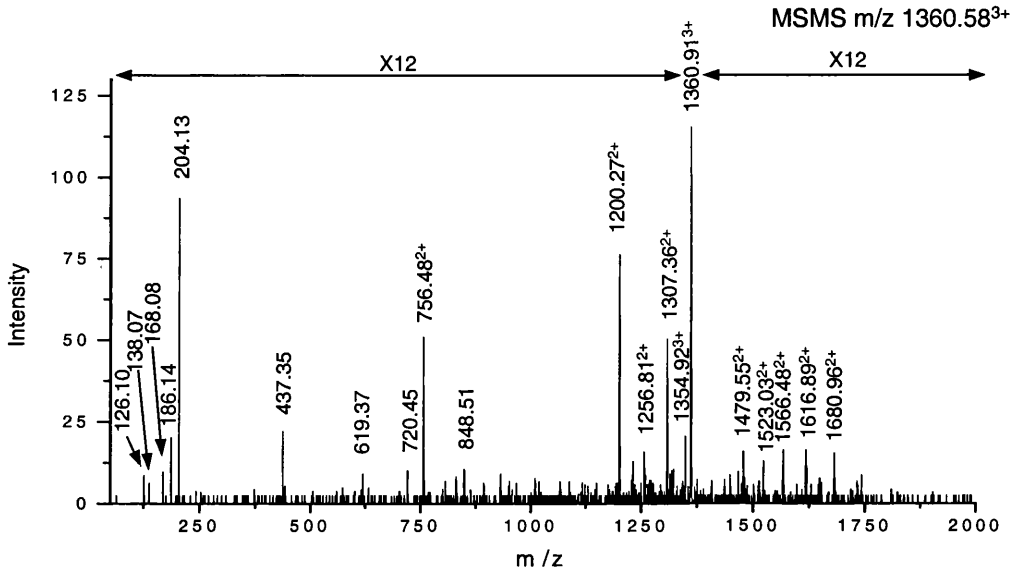
Figure 5.17: ESI-CID-MS-MS spectrum of the doubly GlcNAc-modified peptide $[M + 3H]^{3+}$ m/z 1214.93.

This peptide spans residues 374 – 406 of SRF. Identities of observed peaks are given in Table 5.9.

Table 5.9: Identities of peaks observed in the ESI-CID-MS-MS spectrum of the doubly GlcNAc-modified peptide $[M + 3H]^{3+}$ m/z 1214.93.

The fragmentation spectrum is shown in Figure 5.17, and is of residues 374 – 406 of SRF.

Peak	Match	Peak	Match	Peak	Match
126.05	GlcNAc fragment	668.38	y7	1276.12^{2+}	b26
138.05	GlcNAc fragment	767.46	y8	1375.33^{2+}	b28
168.07	GlcNAc fragment	778.43^{2+}	y16	1467.67	b16
186.07	GlcNAc fragment	854.55	y9	1555.80	y16
204.09	GlcNAc	871.43	PATIMTSSV-H_2O	1618.85^{2+}	MH^{2+}
270.14	y3	917.21^{3+}	b28	1681.85	b17
548.34		970.53		1720.27^{2+}	MH_G^{2+}
586.31		1148.62	MH_G^{3+}		



# Unmatched Ions	Sequence						MH+ Calculated (Da)	MH+ Error (Da)	Protein MW (Da)/pI	Species	PN.SRF Accession #	MS-Digest Index #	Protein Name	
1/14	(R)DSSTDLTQTSSSGTVTLPATIMTSSVPTT VGGHMMYPSPH(A)						4079.8888	-0.1644	51593.9 / 7.08	HUMAN	11831	1	Serum Response Factor	
Fragment-ion (m/z)	204.13	437.35	619.37	720.45	848.51	756.48^{+2}	1200.27^{+2}	1256.81^{+2}	1307.36^{+2}	1479.55^{+2}	1523.03^{+2}	1566.48^{+2}	1616.89^{+2}	1680.96^{+2}
Fragment-ion (m/z)	204.13	437.35	619.37	720.45	848.51	1511.95	2399.53	2512.61	2613.71	2958.09	3045.05	3131.95	3232.77	3360.91
Ion-type Delta ppm		y_4 309.1	b_6 181.7 IMTSSV 92.8	b_7 201.0	b_8 172.4	y_{14}^{+2} 181.7	y_{23}^{+2} 171.9	y_{24}^{+2} 162.5	y_{25}^{+2} 176.3	y_{29}^{+2} 226.9	y_{30}^{+2} 196.8	y_{31}^{+2} 149.1	y_{32}^{+2} 74.0	y_{33}^{+2} 95.4

Figure 5.18: ESI-CID-MS-MS spectrum of $[M + 3H]^{3+}$ m/z 1360.58.

Observed peaks were entered into MS-TAG and the search result displayed identifies the peptide as residues 374 – 413 of SRF.

Another fragmentation spectrum not identified by MASCOT was of a triply-charged peak observed at $[M + 3H]^{3+}$ m/z 1360.58 (Figure 5.18). Observed peaks were entered into MS-TAG and the search result assigns the spectrum to residues 374 – 413 of SRF. This is another peptide starting at residue 374 that has a non-specific cleavage site at its C-terminus, on this occasion after a histidine again. Extracted ion chromatograms of this peak, and peaks corresponding to singly and doubly GlcNAc-modified versions of this peptide (Figure 5.19), show a set of peaks with increasingly earlier retention times as the level of glycosylation increases. The combined spectrum of the region of the

chromatogram when these three peptides elute shows triply-charged peaks corresponding to the unmodified, singly GlcNAc-modified and doubly-glycosylated peptide, along with a methionine-oxidised version of each (Figure 5.19).

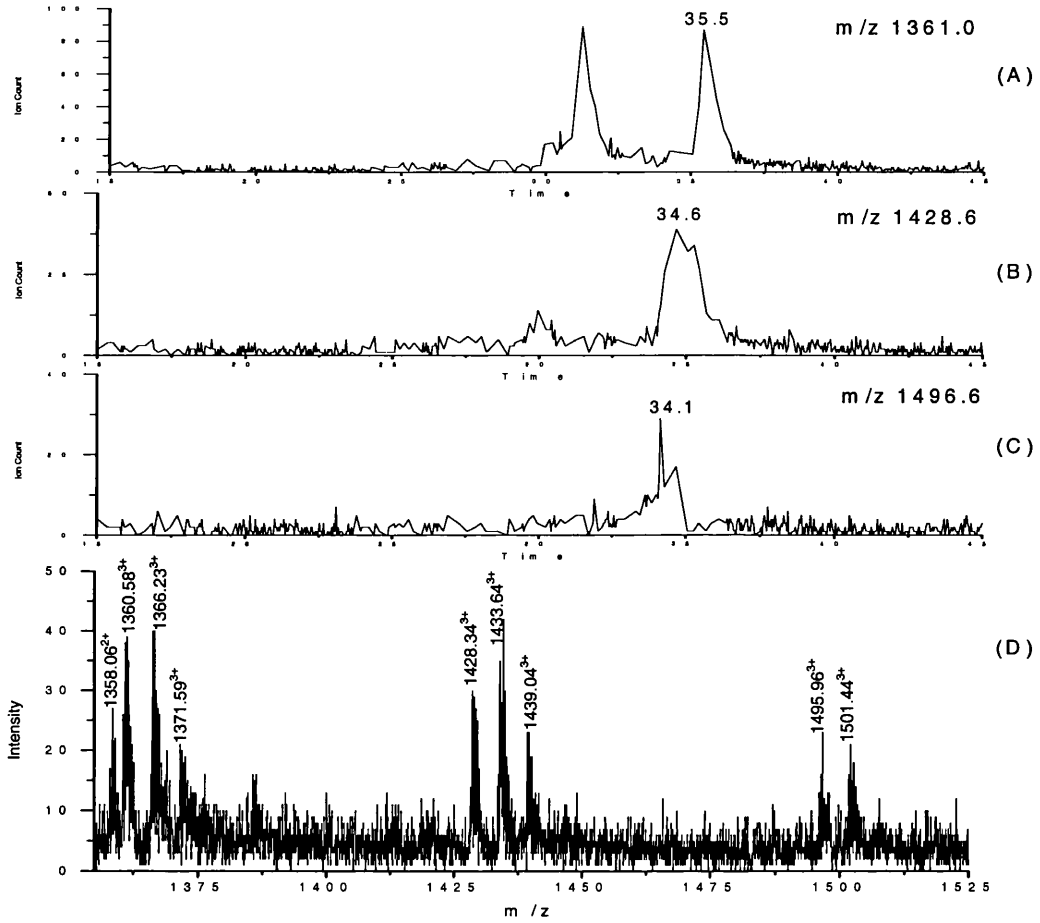


Figure 5.19: Extracted ion chromatograms of the unmodified, singly GlcNAc-modified and doubly GlcNAc-modified peptide spanning residues 374 - 413 of SRF.
Extracted ion chromatograms of (A) unmodified, (B) singly GlcNAc-modified and (C) doubly GlcNAc-modified peptide. (D) The combined mass spectrum during the period in which these three peaks elute contains the unmodified peptide at $[M + 3H]^{3+}$ m/z 1360.58, the singly glycosylated peptide at $[M + 3H]^{3+}$ m/z 1428.34 and the doubly GlcNAc-modified peptide at $[M + 3H]^{3+}$ m/z 1495.96. For each of these peptides a methionine-oxidised version is also observed.

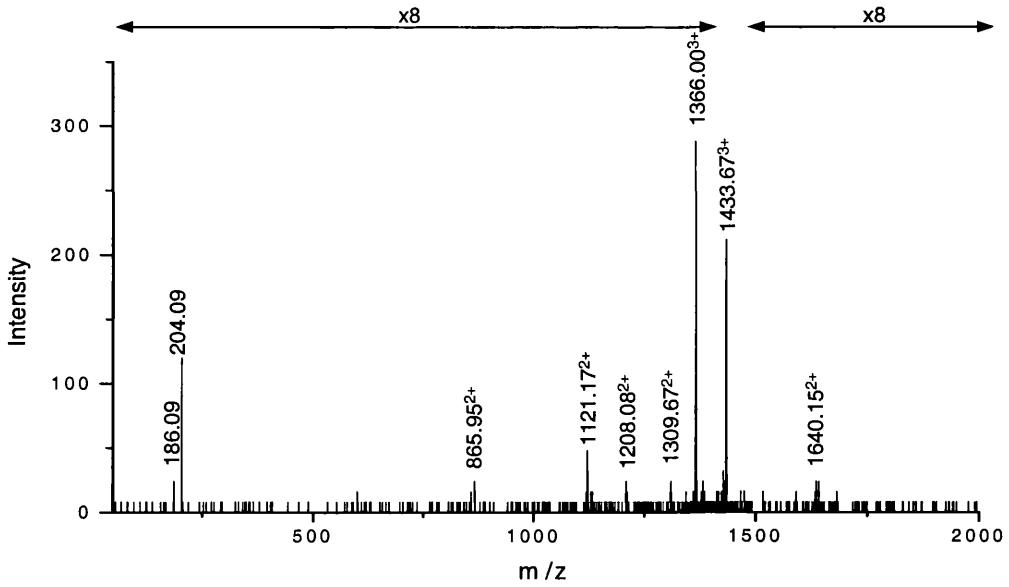


Figure 5.20: ESI-CID-MS-MS spectrum of the singly GlcNAc-modified, singly methionine-oxidised peptide $[M + 3H]^{3+}$ m/z 1433.67.

The identities of peaks observed are given in Table 5.10.

Table 5.10: Identities of the peaks observed in the ESI-CID-MS-MS spectrum of the GlcNAc-modified peptide $[M + 3H]^{3+}$ m/z 1433.67.

The fragmentation spectrum is shown in Figure 5.20, and is of a methionine-oxidised version of residues 374 – 413 of SRF.

Peak	Match	Peak	Match	Peak	Match
186.09	GlcNAc fragment	1121.17^{2+}		1433.67^{3+}	$M_GH_3^{3+}$
204.09	GlcNAc	1208.08^{2+}	y23	1640.15^{2+}	
600.34	y5	1309.67^{2+}	$y23_G$?	1681.75	b17
865.95^{2+}		1366.00^{3+}	MH_3^{3+}		

The singly glycosylated peptide and its methionine-oxidised companion were both selected in a further ESI-CID-MS-MS analysis. The oxidised species at $[M + 3H]^{3+}$ m/z 1433.64 produced the more intense fragment spectrum (Figure 5.20), and the peak

identities in this spectrum are given in Table 5.10. This spectrum identifies the peptide as being GlcNAc-modified due to the prominent loss of GlcNAc from the parent ion to form $[M + 3H]^{3+}$ m/z 1366.00 and the formation of the GlcNAc oxonium ion at m/z 204.09, but there is little other significant fragmentation. A doubly-charged ion is observed at m/z 1208.08, which equates to the y_{23} ion, and another doubly-charged ion at m/z 1309.67 could correspond to a GlcNAc-modified version of this fragment. However, these peaks are too weak to be confident of these assignments.

The doubly GlcNAc-modified version of this peptide was too weak to get any CID-MS-MS data from. Nevertheless, the presence of a doubly glycosylated peptide tallies with the presence of the doubly glycosylated peptide spanning residues 374 – 406.

In the MALDI mass spectrum (Figure 5.1), high mass peaks are observed at m/z 6383 and m/z 6586, corresponding to the unmodified and GlcNAc-modified peptide spanning residues 261 – 324. Searching the LC-MS run, these peaks were also observed and are presented in Figure 5.21. Peaks in this spectrum include unmodified, singly glycosylated and doubly glycosylated peptides containing no, one or two oxidised methionine residues.

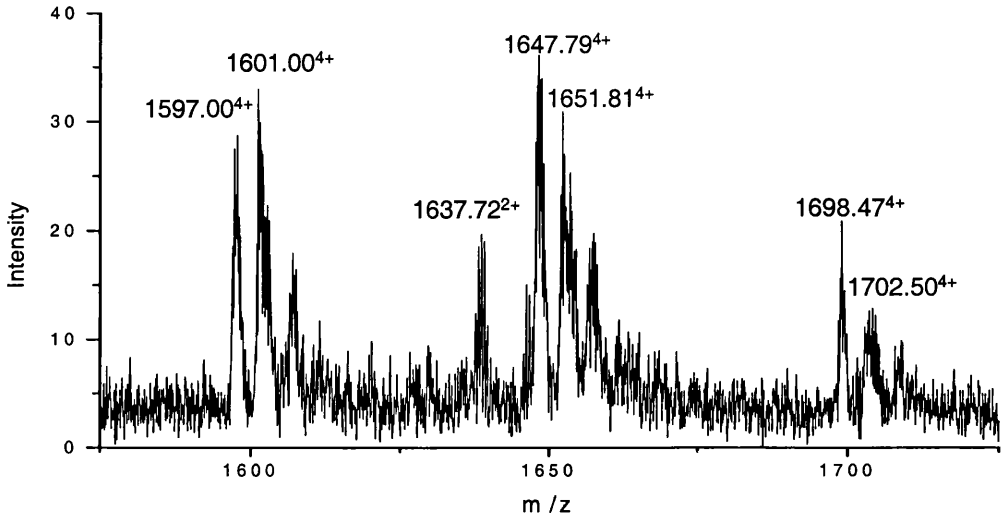


Figure 5.21: ESI-MS spectrum from an LC-MS run of tryptic SRF.

Peaks observed include those corresponding to the peptide of residues 261 – 324 in an unmodified state at $[M + 4H]^{4+}$ m/z 1597.00, singly GlcNAc-modified at $[M + 4H]^{4+}$ m/z 1647.79 and doubly GlcNAc-modified at $[M + 4H]^{4+}$ m/z 1698.47. Methionine-oxidised versions of each are also observed. Peaks are poorly resolved, so the monoisotopic peak is not necessarily detected and labelled.

These peptides are too large to obtain extensive sequence coverage using ESI-CID-MS-MS. However, a fragmentation spectrum of the singly glycosylated, singly methionine-oxidised peptide is shown in Figure 5.22 and peak identities are given in Table 5.11. The loss of the GlcNAc residue and formation of the GlcNAc oxonium ion is observed, plus a few fragment ions that confirm the identity of the peptide. This peptide required deposition of a large amount of internal energy to undergo unimolecular dissociation, and this caused gas-phase elimination of the O-GlcNAc moiety and no glycosylated fragment ions are seen. Sub-digestion of this peptide with another enzyme would be required to produce peptides of a manageable size.

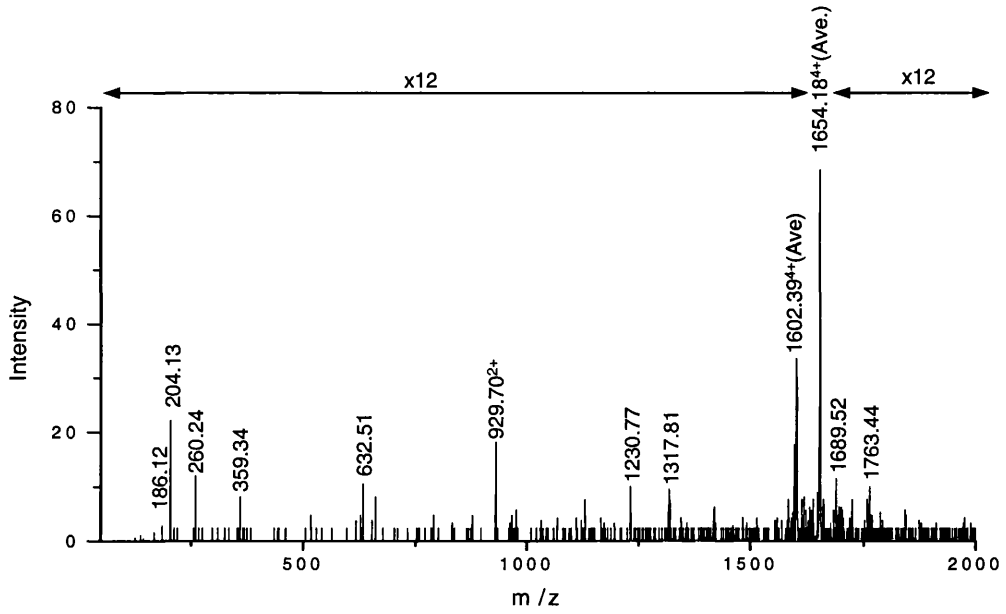


Figure 5.22: ESI-CID-MS-MS spectrum of the quadruply-charged peak at an average mass m/z 1654.18.

This peptide is a singly methionine-oxidised, singly GlcNAc-modified version of the peptide residues 261 – 324 of SRF. Peaks observed are defined in Table 5.11.

Table 5.11: Identities of the peaks observed in the CID-MS-MS spectrum of the quadruply-charged GlcNAc-modified peptide at average mass m/z 1654.18.

The fragmentation spectrum is shown in Figure 5.20, and is of a methionine-oxidised version of residues 261 – 324 of SRF.

Peak	Match	Peak	Match
168.11	GlcNAc fragment	929.70^{2+}	y20
186.12	GlcNAc fragment	1230.77	y13
204.13	GlcNAc	1317.81	y14
260.24	y2	1602.39^{4+} (Ave.)	MH^{4+}
359.34	y3	1689.52	
632.51		1763.44	PGGTSTTIQTAPSTSTTMQ -28

Extracted ion chromatograms of the fragment mass m/z 204.09 were examined of the two MS-MS functions from the LC-MS run (Figure 5.23). One peak was observed, at 16.3 minutes in the upper chromatogram. This peak was formed by the CID spectrum of residues 201 - 212 of SRF. This peptide has GK as its two most C-terminal residues. Hence, this peak arises from a y_2 fragment ion of the peptide, rather than a GlcNAc oxonium ion from a glycopeptide.

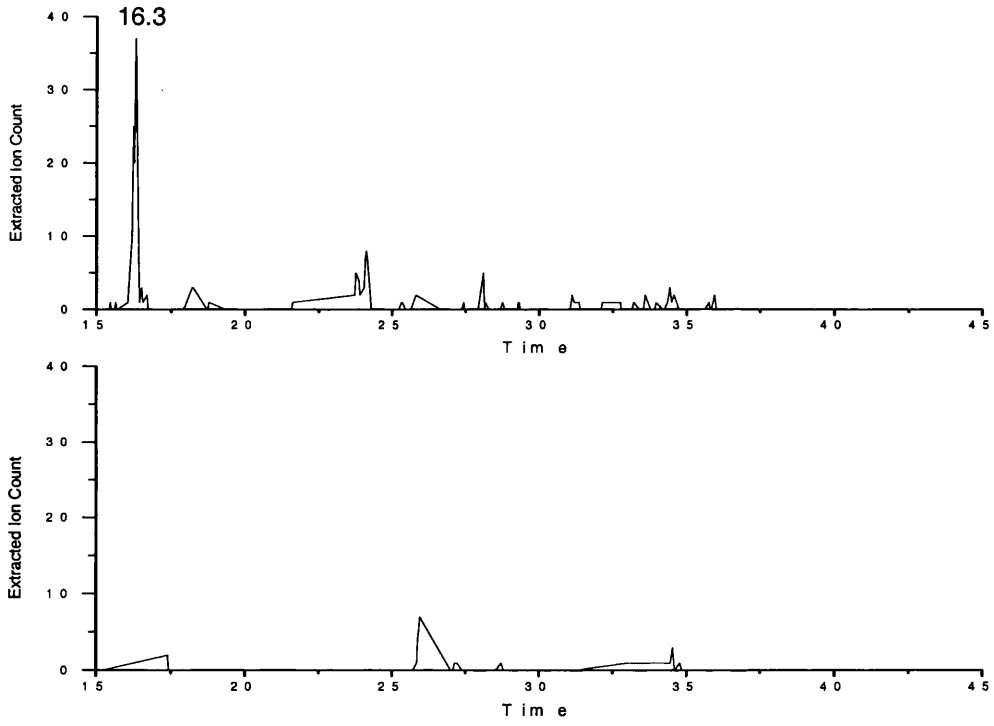


Figure 5.23: Extracted ion chromatograms of m/z 204.1 of CID-MS-MS functions of a tryptic digest of SRF.

5.3. Tryptic and Chymotryptic Digestion

The sequence of the C-terminal half of the SRF protein contains very few convenient enzymatic cleavage sites (see Appendix 3 for protein sequence). As well as a lack of lysine and arginine residues for tryptic cleavages, there are only two aspartic acid residues for Asp-N enzyme, and only two glutamic acid residues, which are cleavage sites for Glu-C enzyme. Thus, chymotrypsin was determined to be the only enzyme that

would specifically cleave this region of the protein into peptides of suitable sizes for fragmentation analysis. Chymotrypsin cleaves after phenylalanines, tyrosines and tryptophans. It also cleaves after selected leucine residues, and sometimes after other residues. Due to this low enzyme specificity, database searches generally have to allow for several 'missed cleavages', and some non-specific cleavage products will not be identified by peptide mass fingerprinting. To try to guarantee the observation of reasonably small peptides from the glycosylated regions of interest, a combined tryptic, then chymotryptic digest was carried out on SRF.

5.3.1. MALDI-MS

The resulting MALDI mass fingerprint is shown in Figure 5.24, and the MS-FIT match is given in Table 5.12. A total of 34 peaks are assigned to predicted peptides from a combined tryptic and chymotryptic digest of SRF with a myc tag. These peptides correspond to 75% sequence coverage of SRF, which is more than was assigned from the mass fingerprint of the digest using trypsin alone (Table 5.1). The majority of the unobserved sequence coverage corresponds to one long region spanning residues 386 – 487 (374 – 475 of the published sequence), which is a region that was also not covered in the tryptic digest of SRF (Table 5.1).

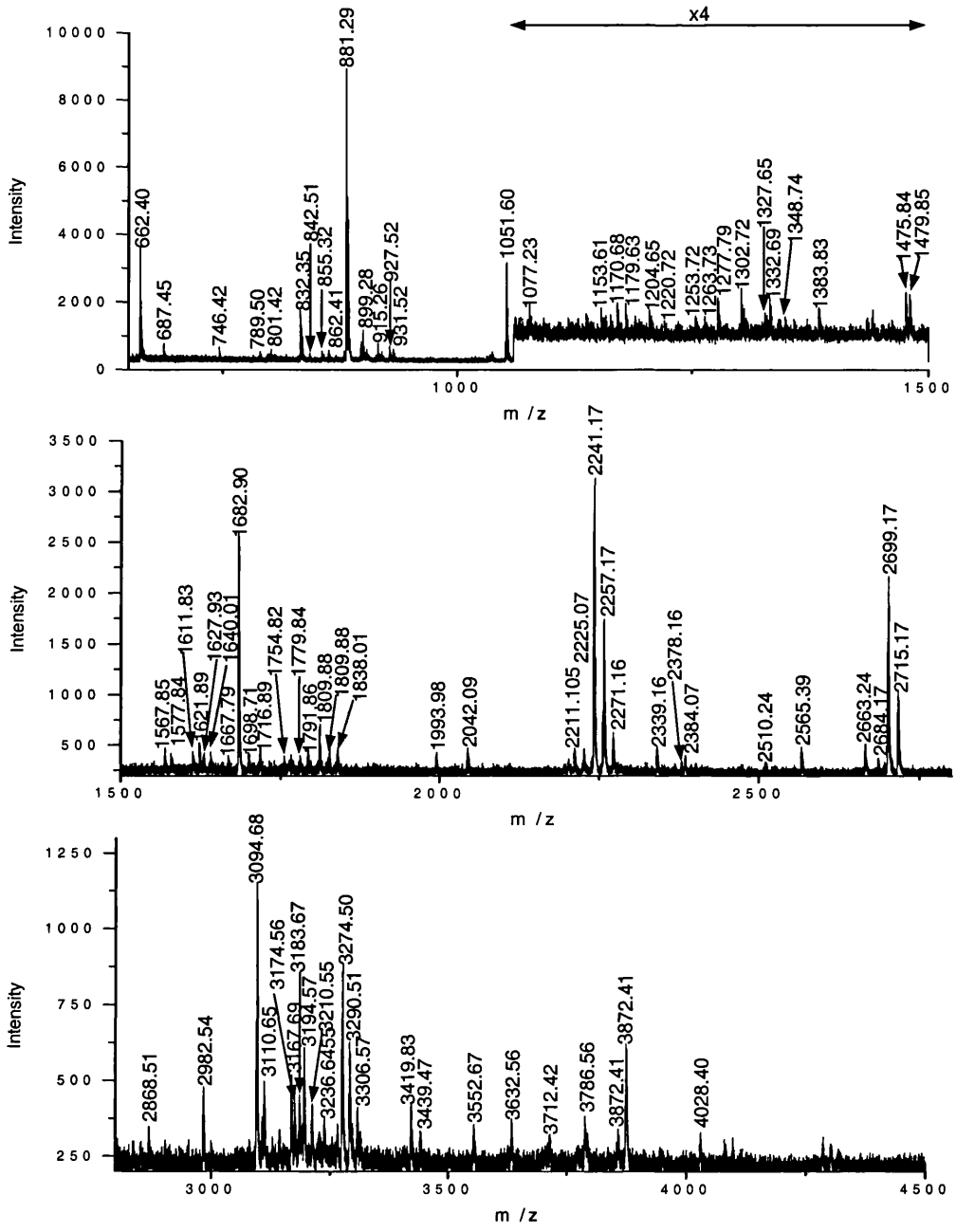


Figure 5.24: MALDI mass spectrum of the combined tryptic and chymotryptic in-gel digest of SRF.

Peak labels are all monoisotopic masses. The result of a MS-FIT search of this spectrum is given in Table 5.12.

Table 5.12: MS-FIT search result of the peaks observed in the MALDI mass fingerprint of a combined tryptic and chymotryptic digest of SRF.

This spectrum is shown in Figure 5.24. The sequence searched against includes the extra N-terminal residues of the myc tag.

•34/91 matches (36%). 51594.1 Da, pI = 7.08. Acc. # 11831. HUMAN. Serum Response Factor.

m/z submitted	MH+ Matched	Delta ppm	Start	End	Peptide Sequence	Mod
746.4206	746.3725	64.4072	166	171	(R)YTTFSK(R)	
927.524	927.4974	28.7362	174	181	(K)TGIMKKAY(E)	1Met-ox
931.5244	931.4961	30.3483	23	32	(R)GSALGGSLNR(T)	
1051.602	1051.5649	35.3481	52	63	(R)VPGNGAGLGPGR(L)	
1153.6124	1153.5927	17.0559	154	162	(K)IKMEFIDNK(L)	1Met-ox
1153.6124	1153.6291	-14.4847	174	183	(K)TGIMKKAYEL(S)	
1170.6749	1170.598	65.7038	33	45	(R)TPTGRPGGGGGTR(G)	
1204.6454	1204.6247	17.1167	210	220	(K)LQPMITSETGK(A)	
1220.7219	1220.6197	83.7391	210	220	(K)LQPMITSETGK(A)	1Met-ox
1332.6906	1332.7197	-21.8492	209	220	(R)KLQPMITSETGK(A)	
1348.7336	1348.7146	14.0938	209	220	(R)KLQPMITSETGK(A)	1Met-ox
1577.8385	1577.8838	-28.6627	150	162	(R)GRVKIKMEFIDNK(L)	
1611.8343	1611.8165	11.0368	221	235	(K)ALIQTCLNSPDSPPR(S)	
1621.8896	1621.9053	-9.6448	180	194	(K)AYELSTLTGTQVLLL(V)	
1627.934	1627.7862	90.7986	484	499	(L)HQMAVIGQQAGSSSNL(T)	
1682.904	1682.8435	35.9608	33	51	(R)TPTGRPGGGGGTRGANGGR(V)	
1682.904	1682.8536	29.9714	221	235	(K)ALIQTCLNSPDSPPR(S)	1Cys-am
1754.819	1754.901	-46.7226	27	45	(L)GGSLNRTPTGRPGGGGGTR(G)	
1779.8426	1779.9394	-54.3727	194	209	(L)LVASETGHVYTFATRK(L)	
1779.8426	1779.9394	-54.3727	195	210	(L)VASETGHVYTFATRKL(Q)	
1809.8797	1810.0411	-89.1747	133	151	(Y)GPVSGAVSGAKPGKKTRGR(V)	
1809.8797	1809.8693	5.7677	333	351	(K)STGSGPVSSGGLMQLPTSF(T)	
1825.883	1825.8642	10.2889	333	351	(K)STGSGPVSSGGLMQLPTSF(T)	1Met-ox
2042.0912	2042.1134	-10.837	311	332	(Y)LAPVSASVSPSAVSSANGTVLK(S)	
2241.1698	2241.1549	6.645	1	22	(-)LLDTEDLNMLPTQAGAAAALGR(G)	
2257.1735	2257.1498	10.48	1	22	(-)LLDTEDLNMLPTQAGAAAALGR(G)	1Met-ox
2472.3288	2472.2404	35.7615	2	26	(L)LDTEDLNMLPTQAGAAAALGRGSAL(G)	1Met-ox
2472.3288	2472.1387	76.9269	108	132	(K)RSLSEMEIGMVVGGPEASAAATGGY(G)	2Met-ox
2684.1713	2684.3388	-62.3744	211	235	(L)QPMITSETGKALIQTCLNSPDSPPR(S)	
2699.1731	2699.1518	7.9107	244	268	(R)MSATGFEETDLTYQVSESDSSGETK(D)	
2715.1642	2715.39	-83.1288	33	63	(R)TPTGRPGGGGGTRGANGGRVPGNGAGLGPGR(L)	
2715.1642	2715.1467	6.4617	244	268	(R)MSATGFEETDLTYQVSESDSSGETK(D)	1Met-ox
2868.5046	2868.4599	15.5813	210	235	(K)LQPMITSETGKALIQTCLNSPDSPPR(S)	1Cys-am
2868.5046	2868.4703	11.9678	487	514	(M)AVIGQQAGSSSNLTELQVVNLDTAHSTK(S)	
2982.5415	2982.4982	14.5385	269	297	(K)DTLKPAFTVTNLPGTTSTIQTAPSTSTTM(Q)	
3094.6829	3094.622	19.6704	352	381	(F)TLMPGGAVAQQVPVQAIQVHQAPQQASPSR(D)	
3110.6469	3110.6169	9.6215	352	381	(F)TLMPGGAVAQQVPVQAIQVHQAPQQASPSR(D)	1Met-ox
3274.5004	3274.6483	-45.1733	298	331	(M)QVSSGPSFPITNYLAPVSASVSPSAVSSANGTVL(K)	pyroGlu
3419.8248	3419.7698	16.0839	298	332	(M)QVSSGPSFPITNYLAPVSASVSPSAVSSANGTVLK(S)	
3552.6711	3552.5738	27.3882	67	103	(R)EAAAAAATTPAPTAGALYSGSEGDSESGEEEELGAER(R)	
3632.5554	3632.8053	-68.7713	108	146	(K)RSLSEMEIGMVVGGPEASAAATGGYGPVSGAVSGAKPGK(K)	

```
1          11         21         31         41         51         61         71
MEQKLISEED LNMLPTQAGA AAALGRGSAL GGSLNRTPTG RPGGGGGTRG ANGGRVPGNG AGLGPGRLER EAAAAAAITR
81         91         101        111        121        131        141        151
APTAGALYSG SEGDSESGEE EELGAEHRGL KRSLSEMEIG MVVGGPEASA AATGGYGPVS GAVSGAKPGK KTRGRVKIKM
161        171        181        191        201        211        221        231
EFIDNKLRRY ITESKRKTGI MKKAYELSTL TGTQVLLLVA SETGHVYTFA TRKLQPMITS ETGKALIQTC LNSPDSPPRS
241        251        261        271        281        291        301        311
DPTTDQRMSA TGFEETDLTY QVSESDSSGE TKDTLKPAFT VTNLPGTTST IQTAPSTSTT MQVSSGPSFP ITNYLAPVSA
321        331        341        351        361        371        381        391
SVSPSAVSSA NGTVLKSTGS GPVSSGGLMQ LPTSFTLMPG GAVAQQVPVQ AIQVHQAPQQ ASPSRDSSTD LTQTSSSGTV
401        411        421        431        441        451        461        471
TLPATIMTSS VPTTVGGHMM YPSPHAVMYA PTSGLGDGSL TVLNAESQAP STMQVSHSQV QEPGGVPQVE LTASSGTVQI
481        491        501        511
PVSAVQLHQM AVIGQQAGSS SNLTELQVVN LDTAHSTKSE
```

The matched peptides cover 75% (392/520 AA's) of the protein

5.3.2. LC-ESI-CID-MS-MS

The combined digest was analysed by LC-MS with automatic selection of peaks for MS-MS analysis. The entire set of fragmentation spectra were automatically analysed by MASCOT, and a summary of the search result is given in Table 5.13. A total of 23 CID-MS-MS spectra are assigned to 22 different peptides, corresponding to 40% sequence coverage. Unfortunately, this search could only be carried out against the published sequence in the database, so the N-terminus is not matched. In fact, the same N-terminal peptide as in the tryptic digest at $[M + 2H]^{2+}$ m/z 1121.10, $[M + 3H]^{3+}$ m/z 747.75 was observed, which spans residues 5 – 26 and contains extra sequence from the myc tag.

As can be seen from Table 5.13, all assigned peptides contain sites that should have been, but were not, cleaved under the conditions of this experiment. Most of the missed cleavages are after leucine residues. Included in the assigned peptides are spectra from two peptides at $[M + 2H]^{2+}$ m/z 810.39 and $[M + 2H]^{2+}$ m/z 835.97 that are from the region of residues 374 – 475, which was not covered in the MALDI mass fingerprint (Table 5.12). This increases the sequence coverage of this protein to over 80% when combining MALDI and ESI data.

Figure 5.25 shows extracted ion chromatograms of fragments of m/z 204.1 from the two MS-MS functions. The upper chromatogram contains three peaks, whilst in the bottom chromatogram, two are observed.

The first peak in the upper chromatogram, at time 21.4 minutes, is formed by the CID-MS-MS spectrum of residues 201 - 212 of SRF. This is the same peptide that was observed in the extracted ion chromatogram from the tryptic digest (Figure 5.23), where the C-terminus of residues GK formed the y_2 fragment ion of theoretical mass 204.13. The peak in the lower chromatogram at nominally the same time (21.6 minutes) is derived from the peptide spanning residues 202 – 212. Tthis is the same peptide as at 21.4 minutes, except it is missing the most N-terminal residue. Nevertheless, it forms the same y_2 ion.

Table 5.13: A summary of the MASCOT search result of CID-MS-MS spectra produced from an LC-MS analysis of a combined tryptic and chymotryptic digest of SRF.

1. gi|4507205 Mass: 51561 Total score: 572 Peptides matched: 23

serum response factor (c-fos serum response element-binding transcription factor) [Homo sapiens]

Observed	Mr(expt)	Mr(calc)	Delta	Miss	Peptide	Mod
466.25	930.48	930.49	-0.01	2	GSALGGSLNR	
538.27	1074.52	1074.53	-0.01	1	LVASETGHVY	
577.29	1152.56	1152.58	-0.03	2	IKMEFIDNK	1Met-Ox
594.81	1187.6	1187.62	-0.02	2	LLVASETGHVY	
602.81	1203.61	1203.62	-0.01	1	LQPMITSETGK	
610.8	1219.59	1219.61	-0.02	1	LQPMITSETGK	1Met-Ox
644.34	1286.67	1286.67	0	2	APTSGLGDGSLTVL	
666.84	1331.67	1331.71	-0.04	2	KLQPMITSETGK	
670.38	1338.75	1338.76	-0.02	1	GPVSGAVSGAKPGKK	
674.84	1347.66	1347.71	-0.04	2	KLQPMITSETGK	1Met-Ox
698.37	1394.73	1394.73	0	3	AYELSTLTGTQVL	
732.8	1463.59	1463.61	-0.02	2	MSATGFEETDLTY	
740.8	1479.59	1479.61	-0.02	2	MSATGFEETDLTY	1Met-Ox
756.35	1510.68	1510.73	-0.04	2	QVVNLDTAHSTKSE	1Pyro-glu
510.26	1527.76	1527.75	0	2	QVVNLDTAHSTKSE	
809.39	1616.76	1616.8	-0.05	1	EAAAAAATTPAPTAGALY	
810.39	1618.76	1618.82	-0.06	3	APTSGLGDGSLTVLNAF	
835.97	1669.93	1669.92	0	1	LTASSGTVQIPVSAVQL	
590.6	1768.77	1768.8	-0.03	1	NSPDSPPRSDPTTDQR	
913.41	1824.8	1824.86	-0.05	1	STGSGPVSSGGLMQLPTSF	1Met-Ox
1021.52	2041.02	2041.11	-0.08	2	LAPVSASVSPSAVSSANGTVLK	
681.37	2041.1	2041.11	-0.01	2	LAPVSASVSPSAVSSANGTVLK	

1 MLPTQAGAAA ALGRGSALGG SLNRTPTGRP GGGGGTRGAN GGRVPGNGAG
51 LGPGRLEREA AAAAATTPAP TAGALYSGSE GDSESGEEEE LGAERRGLKR
101 SLSEMEIGMV VGGPEASAAA TGGYGPVSGA VSGAKPGKKT RGRVKIKMEF
151 IDNKLRRYTT FSKRKTGIMK KAYELSTLTG TQVLLLVASE TGHVYTFATR
201 KLQPMITSET GKALIQTCLN SPDSPPRSDP TTDQRMSATG FEETDLTYQV
251 SESDSSGETK DTLKPAFTVT NLPGTTSTIQ TAPSTSTTMQ VSSGPSFPIT
301 NYLAPVSASV SPSAVSSANG TVLKSTGSGP VSSGGLMQLP TSFTLMPGGA
351 VAQQVPVQAI QVHQAPQQAS PSRDSSTDLT QTSSSGTVTL PATIMTSSVP
401 TTVGGHMMYP SPHAVMYAPT SGLGDGSLTV LNAFSQAPST MQVSHSQVQE
451 PGGVPQVFLT ASSGTVQIPV SAVQLHQMAV IGQQAGSSSN LTELQVVNLD
501 TAHSTKSE

Sequence Coverage: 40%

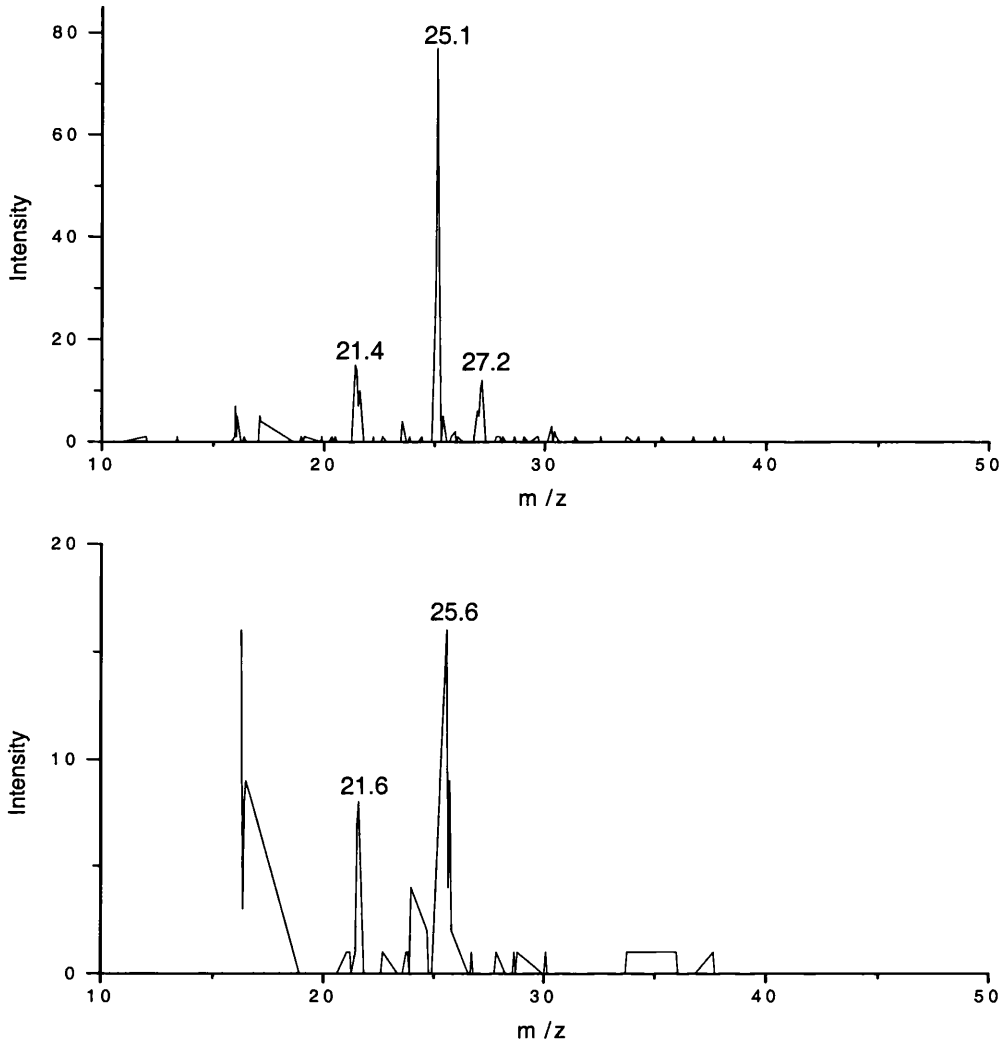


Figure 5.25: Extracted ion chromatograms of m/z 204.1 from ion chromatograms of CID-MS-MS functions of a combined tryptic/chymotryptic digest of SRF.
Peaks are labelled with retention times.

The peak at time 27.2 minutes in the upper chromatogram was formed by the CID-MS-MS spectrum of a GlcNAc-modified peptide of mass $[M + 2H]^{2+}$ m/z 1123.01. An unmodified version of this peptide at $[M + 2H]^{2+}$ m/z 1021.52 was selected for automated CID-MS-MS analysis, and is identified as the peptide residues 303 - 324 in Table 5.13. Figure 5.26 contains the extracted ion chromatograms for the unmodified and GlcNAc-

modified peptide, along with the mass spectrum over the time during which these peaks eluted. As predicted, the GlcNAc-modified peptide eluted slightly earlier than the unmodified. This peptide contains three published GlcNAc modification sites at residues serine 307, serine 309 and serine 316[92]. The monoistopic peak intensity of the modified peptide is roughly 10% of the intensity of the unmodified peptide. If GlcNAc modification halves the peak intensity [59] this suggests roughly a sixth of this peptide was singly GlcNAc-modified.

An LC-ESI-CID-MS-MS spectrum of this glycosylated peptide is shown in Figure 5.27, and the peak identities from this spectrum are provided in Table 5.14. The three major peaks in the spectrum are the parent ion at m/z 1123.17, the deglycosylated parent ion at m/z 1021.61 and the GlcNAc oxonium ion at m/z 204.09. Glycosylated fragment ions y_{13G} - y_{15G} are also observed. These restricts the site of modification to one of the serines 313, 316 or 317 or threonine 321. Hence, this result rules out modification of two of the sites previously reported [92].

This same peptide was also analysed by nanospray-MS. Although no peak was visible in the spectrum, m/z 1123.0 ± 1 Th was selected for CID-MS-MS fragmentation and data was acquired for several hours at this mass. The resulting spectrum is given in Figure 5.28, and peaks are explained in Table 5.15. This spectrum is very similar to that produced from the LC-MS run (Figure 5.27). However it contains significantly more chemical noise, as well as several peaks that were not fragment ions from the doubly-charged glycopeptide. The major contaminating peak is at m/z 365.19, but there are also a number of non-specific ions in the region m/z 1050 - 1120 formed by losses from the parent ion mass. Indeed, an intense singly charged ion is observed at m/z 1122.79, and there is a significant peak at m/z 1104.75 that corresponds to a loss of water from this singly charged parent ion. In terms of glycosylated fragment ions, the same three fragments are seen relating to y_{13G} - y_{15G}. Thus, this spectrum provides no further information on the site of modification, and a number of unassigned ions are observed as shown in Table 5.15, which one must be careful not to misinterpret. Therefore, the LC-ESI-CID-MS-MS spectrum is more reliable, although structurally not more informative than the nanospray-MS spectrum.

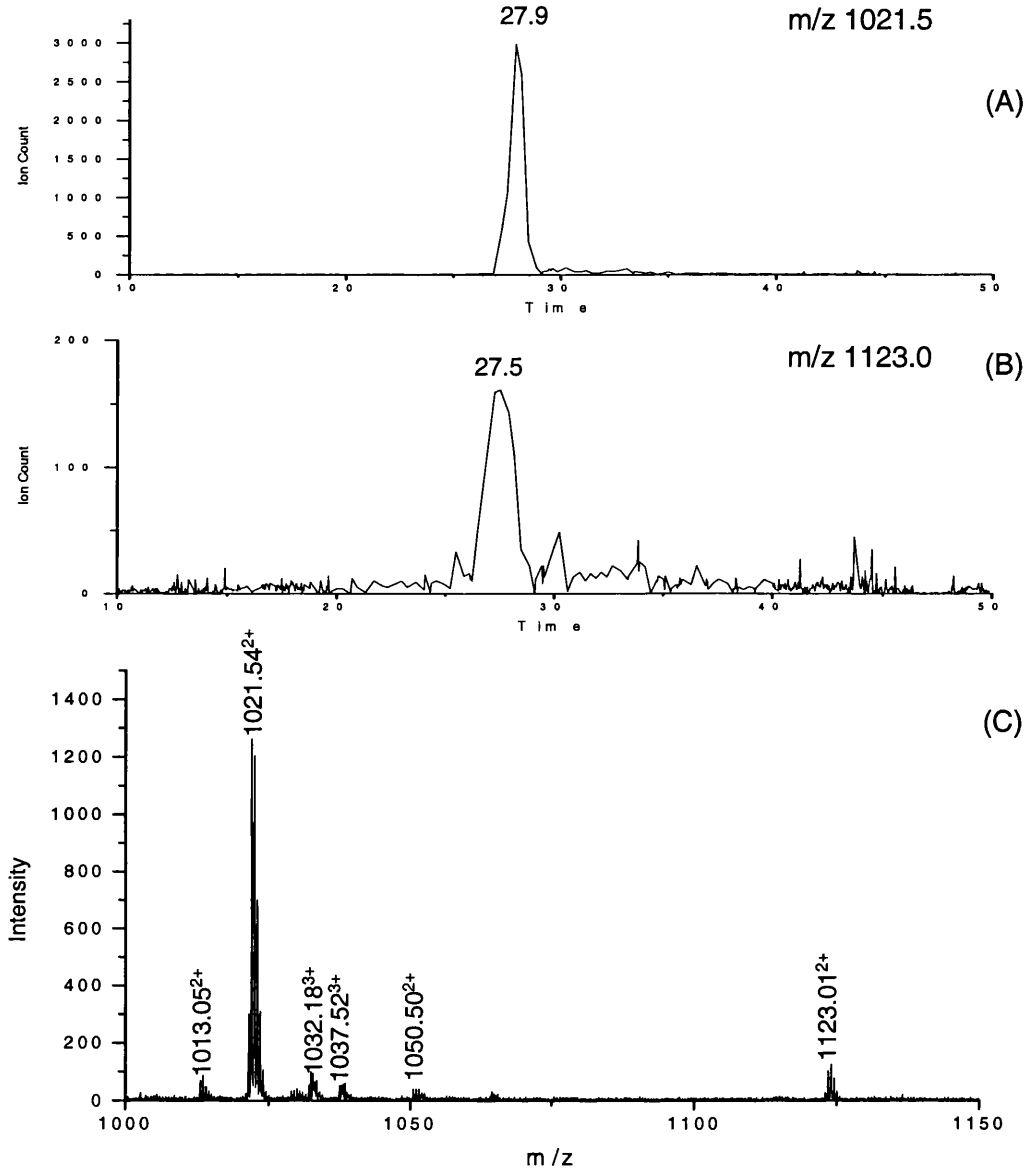


Figure 5.26: Extracted ion chromatograms of unmodified and GlcNAc-modified versions of the peptide spanning residues 303 – 324 of SRF.

Ion chromatograms of (A) unmodified and (B) GlcNAc-modified peptide. (C) The combined mass spectrum during the period, in which the unmodified and GlcNAc-modified peptides elute, contains the unmodified peptide at $[M + 2H]^{2+}$ m/z 1021.54 and the glycosylated peptide at $[M + 2H]^{2+}$ m/z 1123.01.

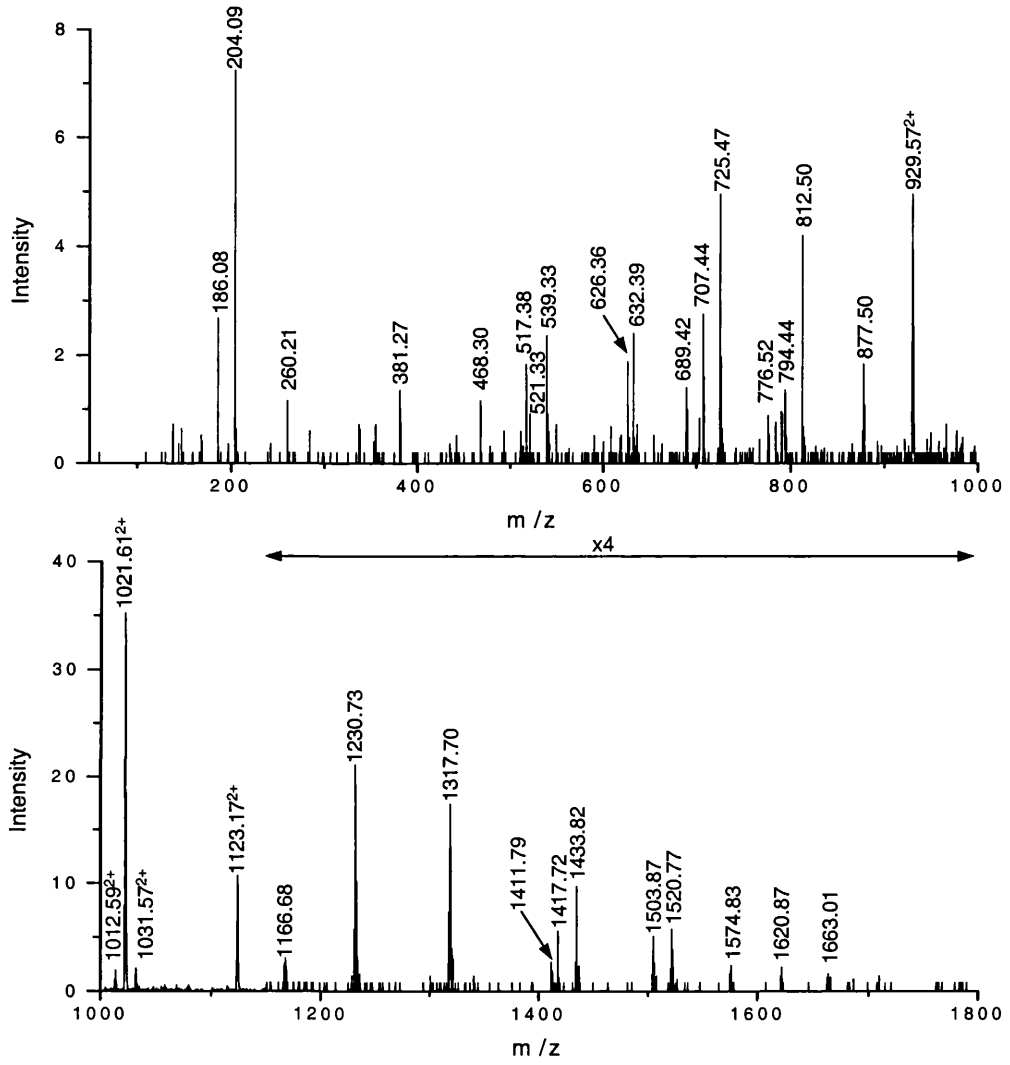
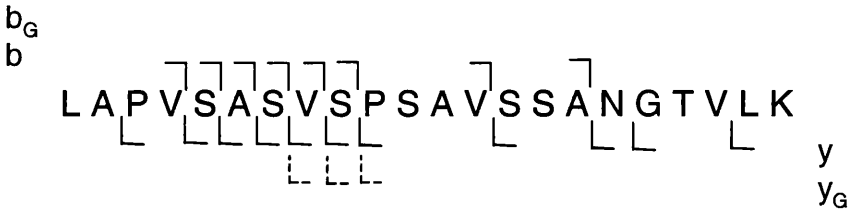


Figure 5.27: ESI-CID-MS-MS spectrum of the GlcNAc-modified peptide [M + 2H]$^{2+}$ m/z 1123.01.

This spectrum was acquired from an LC-MS run of a combined trypsin/chymotrypsin digest of SRF. The peptide is residues 303 – 324 of SRF, and peaks observed are identified in Table 5.14.

Table 5.14: Identities of the peaks observed in the ESI-CID-MS-MS spectrum of the GlcNAc-modified peptide $[M + 2H]^{2+}$ m/z 1123.01.

The fragmentation spectrum is shown in Figure 5.27. This peptide spans residues 303 – 324 of SRF. A graphical representation of the fragment ions observed is also presented.



Peak	Match	Peak	Match	Peak	Match
186.08	GlcNAc fragment	707.44	b8-H_2O	1166.68	b13
204.09	GlcNAc	725.47	b8	1230.73	y13
260.21	y2	776.52	b9-2H_2O	1317.70	y14
381.27	b4	794.44	b9-H_2O	1411.79	b16
468.30	b5	812.50	b9	1417.72	y15
517.38	y5	877.50	y9	1433.82	$y13_G$
521.33	b6-H_2O	929.57^{2+}	y20	1503.87	y16
539.33	b6	1012.59^{2+}	MH_2^{2+}-H_2O	1520.77	$y14_G$
626.36	b7	1021.61^{2+}	MH_2^{2+}	1574.83	y17
632.39	y6	1031.57^{2+}		1620.87	$y15_G$
689.42	b8-2H_2O	1123.17^{2+}	$M_GH_2^{2+}$	1663.01	y18

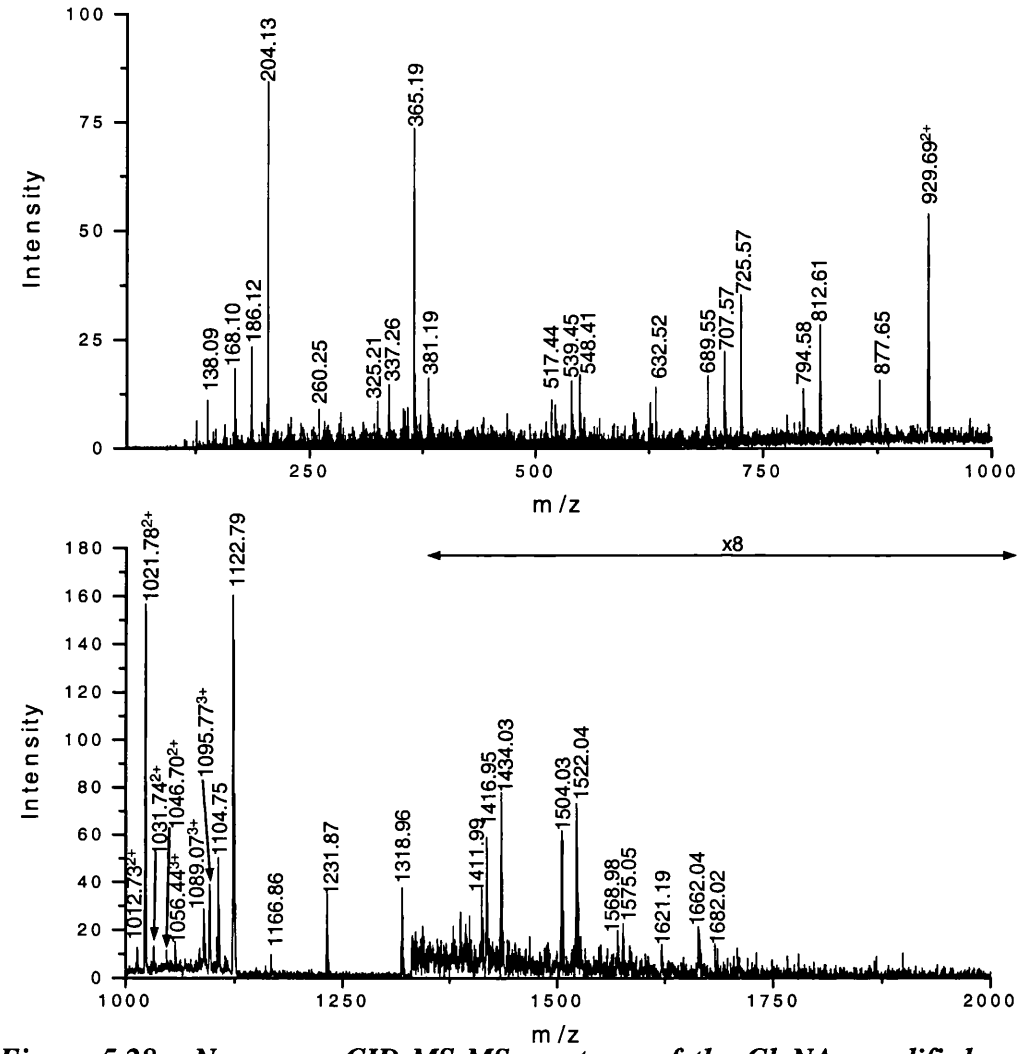


Figure 5.28: Nanospray-CID-MS-MS spectrum of the GlcNAc-modified peptide $[M + 2H]^{2+}$ m/z 1123.29.

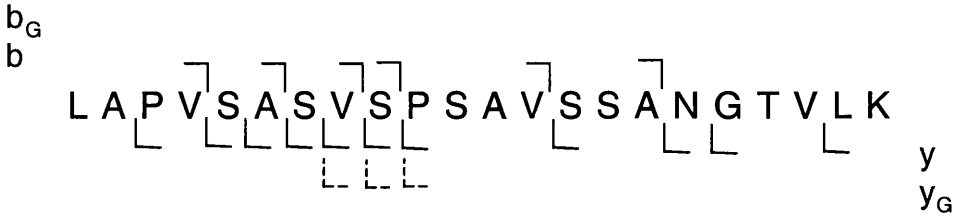
This peptide is residues 303 – 324 of SRF. Peaks observed in this spectrum are identified in Table 5.15.

In Figure 5.25, the major peak in the upper extracted ion chromatogram of m/z 204.1 appears at 25.1 minutes and is formed by the CID-MS-MS spectrum of a peak $[M + 3H]^{3+}$ m/z 844.02 (Figure 5.29). As well as the GlcNAc oxonium ion at m/z 204.09 and a number of GlcNAc fragment ions, the major peak is $[M + 2H]^{2+}$ m/z 1164.06. This represents the loss of GlcNAc from the parent ion. Hence, this peptide is clearly GlcNAc-modified. Indeed, extracted ion chromatograms of this m/z and that of an

unmodified triply-charged peptide (Figure 5.30) show an unmodified version of this peptide was present and eluted just after the GlcNAc-modified. The few low intensity fragment ions in the CID-MS-MS spectrum of the glycopeptide were entered into a MS-TAG search to identify the peptide. The search was carried out using the mass of the deglycosylated parent ion at m/z 1164.06 as the parent ion, because MS-TAG cannot allow for GlcNAc-modified peptides in its searches. The result of this search identifies this peptide as residues 249 - 270 from SRF (Figure 5.29). This peptide is formed by a chymotryptic cleavage at the N-terminus, and a non-specific cleavage (after a threonine) at the C-terminus. It does not contain any published GlcNAc modification site. Unfortunately, this fragmentation spectrum is too weak to observe GlcNAc-modified fragment ions, and in a further tryptic/chymotryptic digest of SRF this peptide was not present. Consequently, no information could be obtained on which residue in this peptide is modified.

Table 5.15: Identities of the peaks observed in the nanospray-CID-MS-MS spectrum of the GlcNAc-modified peptide $[M + 2H]^{2+}$ m/z 1123.01.

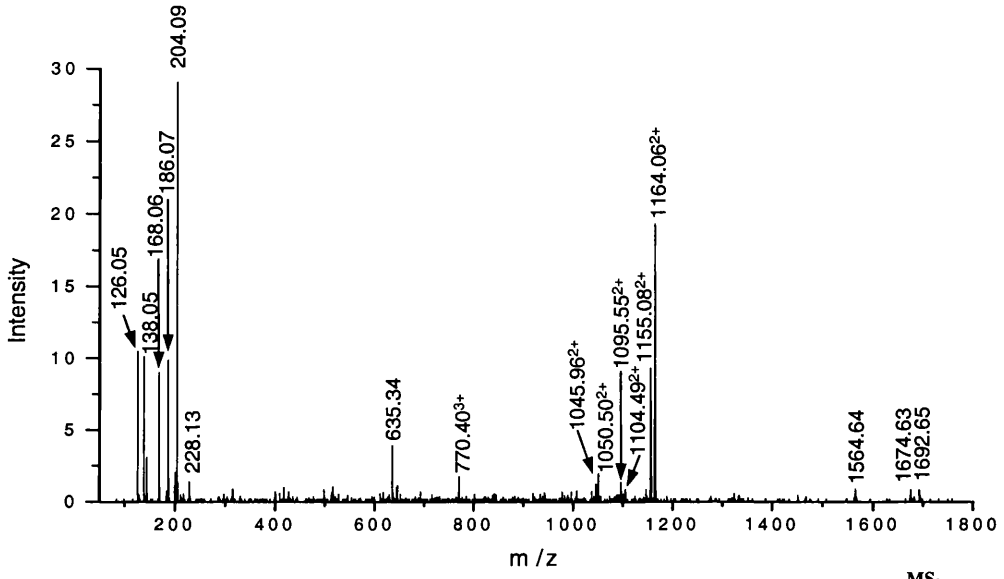
The fragmentation spectrum (Figure 5.28) is of residues 303 – 324 of SRF. A graphical representation of the fragment ions observed is also presented.



Peak	Match	Peak	Match	Peak	Match
138.09	GlcNAc Fragment	707.57	b8-H_2O	1122.79	
168.10	GlcNAc Fragment	725.57	b8	1166.86	b13
186.12	GlcNAc Fragment	794.58	b9-H_2O	1231.87	y13
204.13	GlcNAc	812.61	b9	1318.96	y14
260.25	y2	877.65	y9	1411.99	b16
325.21		929.69^{2+}	y20	1416.95	y15
337.26		1013.73^{2+}	MH_2^{2+}-H_2O	1434.03	$y13_G$
365.19		1021.78^{2+}	MH_2^{2+}	1504.03	y16
381.19	b4	1031.74^{2+}		1522.04	$y14_G$
517.44	y5	1046.70^{2+}		1568.98	
539.45	b6	1056.44^{3+}		1575.05	y17
548.41		1089.07^{3+}		1621.19	$y15_G$
632.52	y6	1095.77^{3+}		1662.04	y18
689.55	b8-$2H_2O$	1104.75		1682.02	

The largest peak in the lower chromatogram of Figure 5.25 is formed by the CID-MS-MS spectrum of a peak $[M + 3H]^{3+}$ m/z 862.08, and this spectrum is shown in Figure 5.31. The spectrum contains the GlcNAc oxonium ion and related fragment ions, as well as a deglycosylated parent ion at $[M + 2H]^{2+}$ m/z 1191.50. Hence, this is another spectrum of a glycopeptide. However, it is weak in intensity, and a MS-TAG search

using the peaks in this spectrum fails to identify the peptide. There is no predicted peptide from SRF at this mass that could be formed by a combined tryptic and chymotryptic digest, and in a second tryptic/chymotryptic combined digest this peptide was not observed. Consequently, the identity of this peptide remains a mystery.



Rank	# Unmatched Ions	Sequence	MH$^+$ Calculated (Da)	MH$^+$ Error (ppm)	Protein MW (Da)/pI	Species	PN.srf Accession #	MS-Digest Index #	Protein Name
1	0/11	(Y) QVSESDSSGETKDTLKPAFTVT(N)	2327.1254	-5.7	51594.1 / 7.08	HUMAN	11831	1	Serum Response Factor

Fragment-ion (m/z)	228.13	635.34	1564.64	1674.63	1692.65	1045.96^{+2}	1050.50^{+2}	1095.55^{+2}	1104.49^{+2}	1155.08^{+2}	770.40^{+3}
Fragment-ion (m/1)	228.13	635.34	1564.64	1674.63	1692.65	2090.91	2099.99	2190.09	2207.97	2309.15	2309.18
Ion-type Delta ppm	b_2 -21.1	y_6 -0.7	b_{15} -37.0	b_{16}-H_2O -90.9	b_{16} -84.4	b_{20}-H_2O^{+2} -36.4	y_{20}^{+2} -3.0	b_{21}-H_2O^{+2} 16.2	b_{21}^{+2} -43.0	b_{22}^{+2} 16.2 y_{22}-H_2O^{+2}	b_{22}^{+3} 30.1 y_{22}-H_2O^{+3}

Figure 5.29: ESI-CID-MS-MS spectrum of [M + 3H]$^{3+}$ m/z 844.02 from a combined tryptic/chymotryptic digest of SRF.

Observed peaks were entered into MS-TAG and the search result identifies the peptide as residues 249 – 270 of SRF.

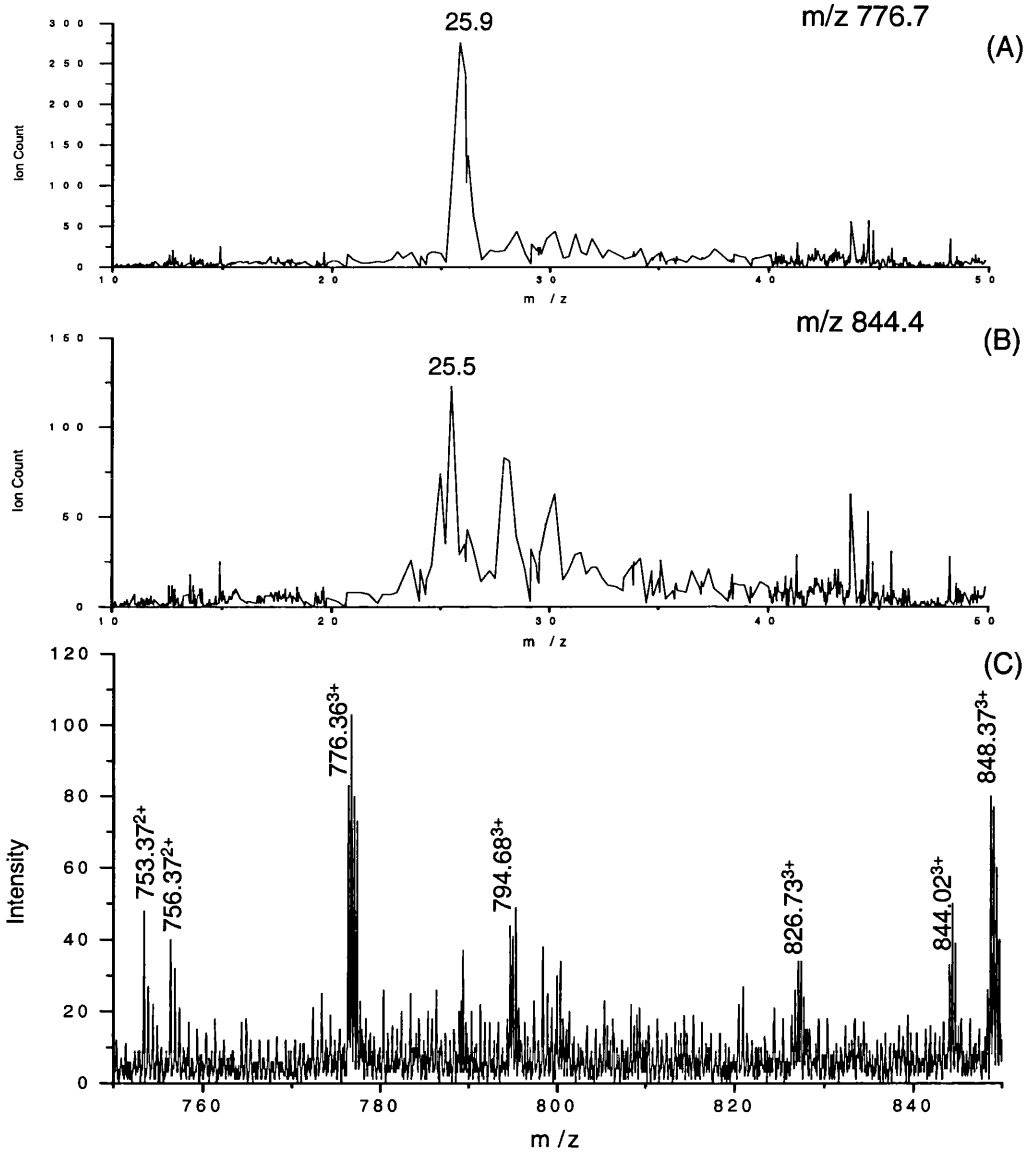


Figure 5.30: Extracted ion chromatograms of the unmodified and GlcNAc-modified peptide spanning residues 249 – 270 of SRF.

Extracted ion chromatograms of (A) unmodified and (B) GlcNAc-modified peptides. (C) The combined mass spectrum during the period in which the unmodified and GlcNAc-modified peptides elute, contains the unmodified peptide at $[M + 3H]^{3+}$ m/z 776.36 and the glycosylated peptide at $[M + 3H]^{3+}$ m/z 844.02.

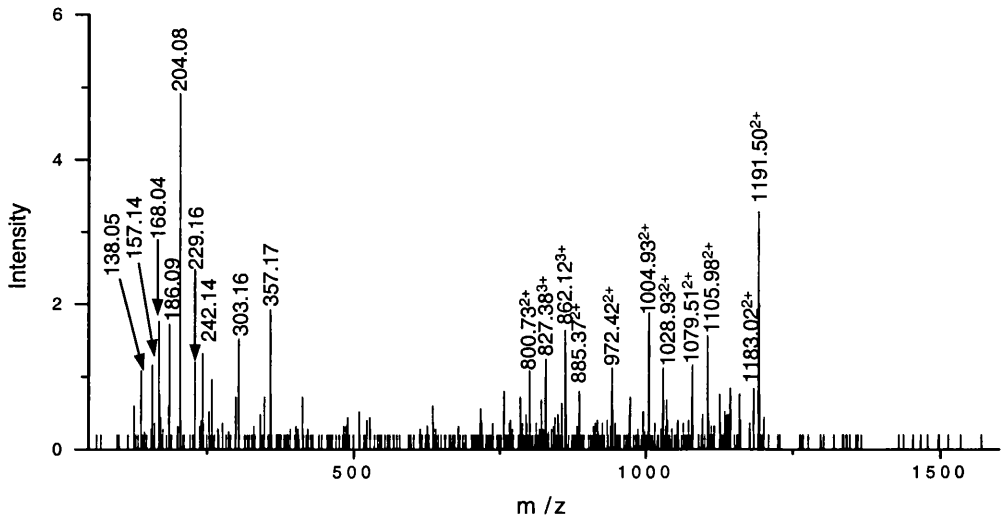


Figure 5.31: ESI-CID-MS-MS spectrum of the peptide at $[M + 3H]^{3+}$ m/z 862.08 produced by a combined tryptic/chymotryptic digest of SRF.

A MS-TAG search using the peaks observed in the mass spectrum failed to identify the peptide, although it does appear to be GlcNAc-modified due to the presence of the GlcNAc oxonium ion at m/z 204.08 and a prominent loss of GlcNAc from the parent ion to form $[M + 2H]^{2+}$ m/z 1191.50.

5.4. Chymotryptic Digestion

A digest using only chymotrypsin was also carried out. Although this gives larger peptides than the combined tryptic/chymotryptic digest, which is not desirable for fragmentation analysis, it does not produce peptides with basic C-termini. Hence, there is a better chance of producing overlapping 'b' and 'y' ion series, which may be required in order to identify a site of modification[107].

5.4.1. MALDI-MS

A MALDI mass fingerprint of the digest was acquired, which is presented in Figure 5.32. The peak list was submitted for a MS-FIT search and the result is given in Table 5.16. 35 peaks are assigned to the sequence of SRF with a myc tag extended N-terminus, comprising 63% sequence coverage of the protein.

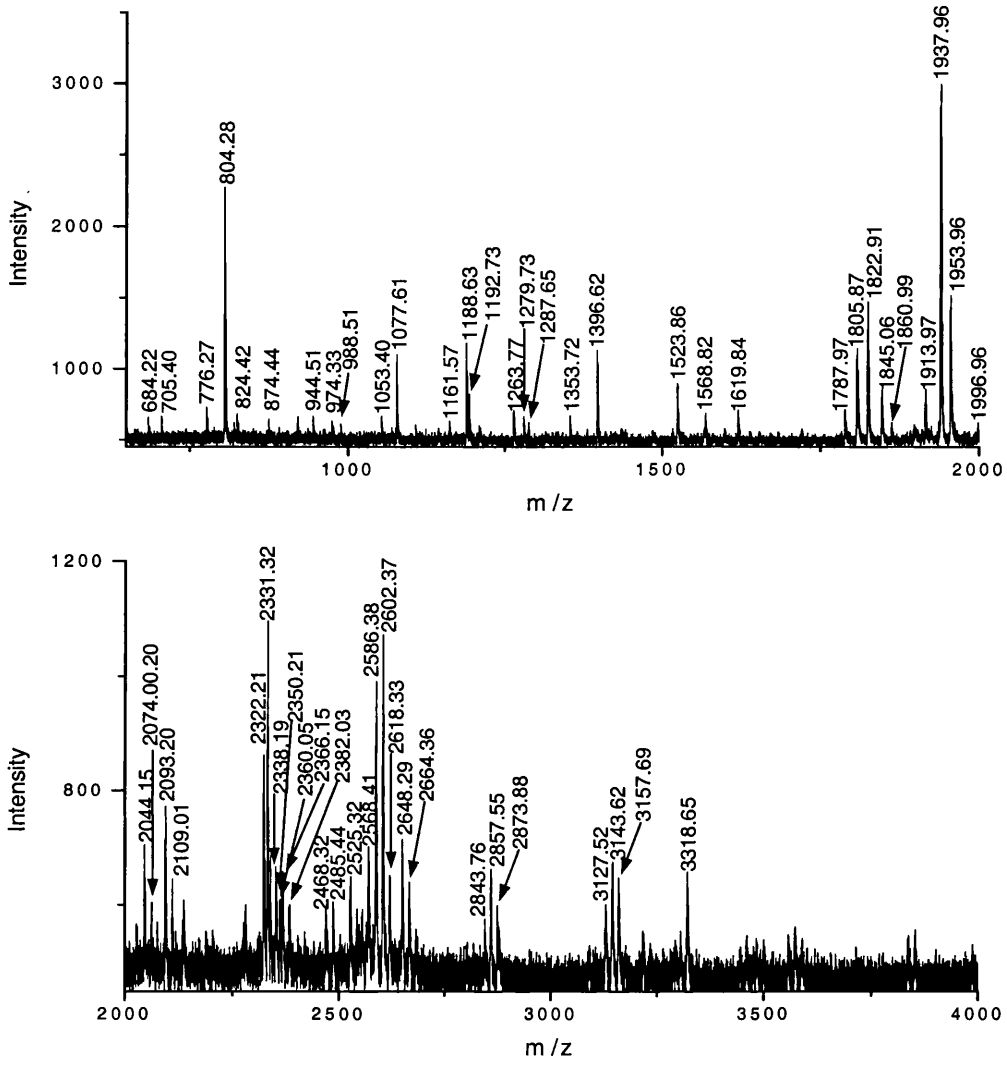


Figure 5.32: MALDI mass spectrum of a chymotryptic in-gel digest of SRF.

Peak labels are all monoisotopic masses. The MS-FIT search of this spectrum is Table 5.16.

Table 5.16: MS-FIT search result of peaks observed in a MALDI mass fingerprint of the chymotryptic digest of SRF.

This search is of the peptides observed in Figure 5.32. The sequence searched against includes the myc tag.

1. 35/59 matches (54%). 51594.1 Da, pI = 7.08. Acc. # 11831. HUMAN. Serum Response Factor.

m/z submitted	MH+ matched	Delta ppm	Start	End	Peptide Sequence	Mod
705.4	705.3307	98.2868	6	11	(L)LSEEDL(N)	
874.44	874.4747	-39.6286	25	34	(L)GRGSALGGSL(N)	
944.51	944.5351	-26.6226	210	217	(F)ATRKLQPM(I)	
1053.4	1053.4675	-64.093	444	453	(L)NAFSQAPSTM(Q)	
1077.61	1077.6169	-6.4046	163	170	(F)IDNKLRRY(T)	
1161.57	1161.6367	-57.4136	186	196	(Y)ELSTLTGTQVL(L)	
1188.63	1188.6265	2.9754	197	207	(L)LLVASETGHVY(T)	
1192.73	1192.6512	66.0391	208	217	(Y)TFATRKLQPM(I)	
1287.65	1287.6796	-23.0035	430	443	(Y)APTSGLGDGSLTVL(N)	
1353.72	1353.7279	-5.8422	161	170	(M)EFIDNKLRRY(T)	
1396.62	1396.6749	-39.2863	302	314	(M)QVSSGPSFPITNY(L)	
1619.84	1619.8281	7.3611	430	446	(Y)APTSGLGDGSLTVLNAF(S)	
1787.97	1787.9543	8.7589	191	207	(L)TGTQVLLLVASETGHVY(T)	
1805.87	1805.8822	-6.7763	454	470	(M)QVSHSQVQEPGGVPQVF(L)	pyroGlu
1822.91	1822.9088	0.6658	454	470	(M)QVSHSQVQEPGGVPQVF(L)	
1845.06	1845.0268	17.9982	210	226	(F)ATRKLQPMITSETGKAL(I)	
1860.99	1861.0217	-17.0377	210	226	(F)ATRKLQPMITSETGKAL(I)	1Met-ox
1913.97	1914.0184	-25.2852	315	335	(Y)LAPVSASVSPSAVSSANGTVL(K)	
1937.96	1937.9643	-2.1942	336	355	(L)KSTGSGPVSSGGLMQLPTSF(T)	
1953.96	1954.1159	-79.8019	174	190	(F)SKRKTGIMKKAYELSTL(T)	
1953.96	1953.9592	0.4264	336	355	(L)KSTGSGPVSSGGLMQLPTSF(T)	1Met-ox
1953.96	1954.0068	-23.9435	472	490	(L)TASSGTVQIPVSAVQLHQM(A)	
2044.15	2044.0272	60.0548	5	24	(-)LLSEEDLNMLPTQAGAAAAL(G)	1Met-ox
2093.2	2093.1429	27.2866	208	226	(Y)TFATRKLQPMITSETGKAL(I)	
2109.01	2109.1378	-60.5934	208	226	(Y)TFATRKLQPMITSETGKAL(I)	1Met-ox
2322.21	2322.1288	34.9889	430	453	(Y)APTSGLGDGSLTVLNAFSQAPSTM(Q)	
2331.32	2331.2448	32.2714	186	207	(Y)ELSTLTGTQVLLLVASETGHVY(T)	
2338.19	2338.1237	28.3706	430	453	(Y)APTSGLGDGSLTVLNAFSQAPSTM(Q)	1Met-ox
2350.21	2350.067	60.841	408	429	(M)TSSVPTTVGGHMMYPSPHAVMY(A)	
2366.15	2366.0619	37.2204	408	429	(M)TSSVPTTVGGHMMYPSPHAVMY(A)	1Met-ox
2382.03	2382.0568	-11.2712	408	429	(M)TSSVPTTVGGHMMYPSPHAVMY(A)	2Met-ox
2485.44	2485.1921	99.7589	429	453	(M)YAPTSGLGDGSLTVLNAFSQAPSTM(Q)	
2525.32	2525.2095	43.7741	447	470	(F)SQAPSTMQVSHSQVQEPGGVPQVF(L)	
2568.41	2568.2438	64.7143	111	136	(L)KRSLSEMEIGMVVGGPEASAAATGGY(G)	
2618.33	2618.3372	-2.7549	35	63	(L)NRTPTGRPGGGGGTRGANGGRVPGNGAGL(G)	
2857.55	2857.3579	67.2242	444	470	(L)NAFSQAPSTMQVSHSQVQEPGGVPQVF(L)	
3127.52	3127.4086	35.6309	89	117	(Y)SGSEGDSESGEEEELGAERRGLKRSLSEM(E)	1Met-ox

1	11	21	31	41	51	61	71
MEQKLLDTED	LNMLPTQAGA	AAALGRGSAL	GGSLNRTPTG	RPGGGGGTRG	ANGGRVPGNG	AGLGPGRLER	EAAAAAATTP
81	91	101	111	121	131	141	151
APTAGALYSG	SEGDSESGEE	EELGAERRGL	KRSLSEMEIG	MVVGGPEASA	AATGGYGPVS	GAVSGAKPGK	KTRGRVKIKM
161	171	181	191	201	211	221	231
EFIDNKLRRY	TTFSKRKTGI	MKKAYELSTL	TGTQVLLLVA	SETGHVYTEA	TRKLQPMITS	ETGKALIQTC	LNSPDSPPRS
241	251	261	271	281	291	301	311
DPTTDQRMSA	TGEEETDLTY	QVSESDSSGE	TKDTLKPAET	VTNLPGTTST	IQTAPSTSTT	MQVSSGPSFP	ITNYLAPVSA
321	331	341	351	361	371	381	391
SVSPSAVSSA	NGTVLKSTGS	GPVSSGGLMQ	LPTSETLMPG	GAVAQQVPVQ	AIQVHQAPQQ	ASPSRDSSTD	LTQTSSSGTV
401	411	421	431	441	451	461	471
TIMTSSVPTT	VGGHMMYPSP	HAVMYAPTSG	LGDGSLTVLN	AFSQAPSTMQ	VSHSQVQEPG	GVPQVELTAS	SGTVQIPVSA
481	491	501	511				
TLPAVQLHQM	AVIGQQAGSS	SNLTELQVVN	LDTAHSTKSE				

The matched peptides cover 63% (327/520 AA's) of the protein

5.4.2. LC-ESI-CID-MS-MS

The sample was then analysed by LC-MS with automatic precursor ion selection for CID-MS-MS analysis. The resulting data were submitted to a MASCOT search in a 'no enzyme specificity mode'. Table 5.17 shows that MASCOT assigned 32 MS-MS spectra to 29 different peptides.

Figure 5.33 shows extracted ion chromatograms of m/z 204.1 from the two MS-MS functions, in which two peaks are observed in the upper chromatogram. The peak at 32.3 minutes is formed by the CID-MS-MS spectrum of $[M + 2H]^{2+}$ m/z 1059.01. This peak is a glycosylated peptide of residues 303 – 323, so is similar to the glycosylated peptide of residues 303 – 324 observed at $[M + 2H]^{2+}$ m/z 1123.01 in the combined tryptic/chymotryptic digest (Figure 5.26). However, it does not contain the final C-terminal lysine residue. Extracted ion chromatograms of the glycosylated and unmodified peptide (Figure 5.34), show the glycosylated peptide eluted slightly before the unmodified peptide. An LC-ESI-CID-MS-MS spectrum of this glycopeptide is shown in Figure 5.35, and the peak identities are listed in Table 5.18. Comprehensive fragmentation of this peptide is observed. There are 11 'y' ions in the spectrum, as there were for the spectrum of the peptide spanning residues 303 – 324 (Table 5.14), but 14 'b' ions are present in this spectrum, compared with eight from the peptide observed in the chymotryptic/tryptic combined digest. This shows the benefit of not having a basic C-terminal residue for increasing the number of N-terminally derived fragment ions observed. There are four glycosylated 'y' ions in this spectrum, confirming that serines 307, 309 and 311 are not modified residues, agreeing with the results of the fragmentation spectrum of residues 303 - 324. However, a glycosylated b_{13G} ion is present in this spectrum at m/z 1369.80. The combination of b_{13G} and y_{13G} ions define the site of glycosylation among the middle four residues PSAV. Of these, only serine 313 can bear the modification. This establishes a new site of modification on serine 313.

Table 5.17: Summary of the MASCOT search results of ESI-CID-MS-MS spectra produced from an LC-MS analysis of a chymotryptic digest of SRF.

32 CID spectra were assigned to SRF, corresponding to 39% sequence coverage.

1. gi|4507205 **Mass:** 51561 **Total score:** 539 **Peptides matched:** 32
serum response factor (c-fos serum response element-binding transcription factor) [Homo sapiens]

Observed	Mr(expt)	Mr(calc)	Delta	Miss	Peptide	Mod
437.74	873.46	873.47	0	1	GRGSALGGSL	
494.76	987.51	987.51	0	2	GRGSALGGSLN	
527.23	1052.44	1052.46	-0.02	1	NAFSQAPSTM	
538.27	1074.53	1074.53	0	0	LVASETGHVY	
539.32	1076.63	1076.61	0.02	1	IDNKLRRY	
542.29	1082.57	1082.57	0	0	GPVSGAVSGAKPG	
594.81	1187.59	1187.62	-0.02	2	LLVASETGHVY	
598.26	1194.5	1194.52	-0.02	1	MYPSPHAVMY	
644.32	1286.63	1286.67	-0.05	2	APTSGLGDGSLTVL	
644.33	1286.65	1286.67	-0.02	2	APTSGLGDGSLTVL	
690.31	1378.6	1378.64	-0.04	1	QVSSGPSFPITNY	1Pyro-glu
698.83	1395.64	1395.67	-0.03	1	QVSSGPSFPITNY	
725.85	1449.69	1449.74	-0.05	3	YAPTSGLGDGSLTVL	
506.26	1515.77	1515.83	-0.06	1	KLQPMITSETGKAL	
810.39	1618.76	1618.82	-0.06	3	APTSGLGDGSLTVLNAF	
835.94	1669.86	1669.92	-0.07	1	LTASSGTVQIPVSAVQL	
894.46	1786.91	1786.95	-0.04	3	TGTQVLLLVASETGHVY	
904.47	1806.93	1806.98	-0.06	2	LTASSGTVQIPVSAVQLH	
911.93	1821.84	1821.9	-0.06	0	QVSHSQVQEPGGVPQVF	
615.67	1843.97	1844.02	-0.05	1	ATRKLQPMITSETGKAL	
621	1859.99	1860.01	-0.03	1	ATRKLQPMITSETGKAL	1Met-Ox
969.46	1936.91	1936.96	-0.05	2	KSTGSGPVSSGGLMQLPTSF	
977.45	1952.89	1952.95	-0.06	2	KSTGSGPVSSGGLMQLPTSF	1Met-Ox
977.47	1952.92	1952.95	-0.03	2	KSTGSGPVSSGGLMQLPTSF	1Met-Ox
977.47	1952.93	1952.95	-0.02	2	KSTGSGPVSSGGLMQLPTSF	1Met-Ox
1022.53	2043.04	2043.09	-0.05	1	TLMPGGAVAQQVPVQAIQVH	
698.36	2092.07	2092.14	-0.07	2	TFATRKLQPMITSETGKAL	
703.69	2108.05	2108.13	-0.08	2	TFATRKLQPMITSETGKAL	1Met-Ox
789.33	2364.97	2365.05	-0.09	1	TSSVPTTVGGHMMYPSPHAVMY	1Met-Ox
842.38	2524.11	2524.2	-0.09	0	SQAPSTMQVSHSQVQEPGGVPQVF	
953.1	2856.26	2856.35	-0.09	1	NAFSQAPSTMQVSHSQVQEPGGVPQVF	
958.42	2872.24	2872.34	-0.11	1	NAFSQAPSTMQVSHSQVQEPGGVPQVF	1Met-Ox

```
  1 MLPTQAGAAA ALGRGSALGG SLNRTPTGRP GGGGGTRGAN GGRVPGNGAG
 51 LGPGRLEREA AAAAATTPAP TAGALYSGSE GDSESGEEEE LGAERRGLKR
101 SLSEMEIGMV VGGPEASAAA TGGYGPVSGA VSGAKPGKKT RGRVKIKMEF
151 IDNKLRRYTT FSKRKTGIMK KAYELSTLTG TQVLLLVASE TGHVYTFATR
201 KLQPMITSET GKALIQTCLN SPDSPPRSDP TTDQRMSATG FEETDLTYQV
251 SESDSSGETK DTLKPAFTVT NLPGTTSTIQ TAPSTSTTMQ VSSGPSFPIT
301 NYLAPVSASV SPSAVSSANG TVLKSTGSGP VSSGGLMQLP TSFTLMPGGA
351 VAQQVPVQAI QVHQAPQQAS PSRDSSTDLT QTSSSGTVTL PATIMTSSVP
401 TTVGGHMMYP SPHAVMYAPT SGLGDGSLTV LNAFSQAPST MQVSHSQVQE
451 PGGVPQVFLT ASSGTVQIPV SAVQLHQMAV IGQQAGSSSN LTELQVVNLD
501 TAHSTKSE
```

Sequence Coverage: 39%

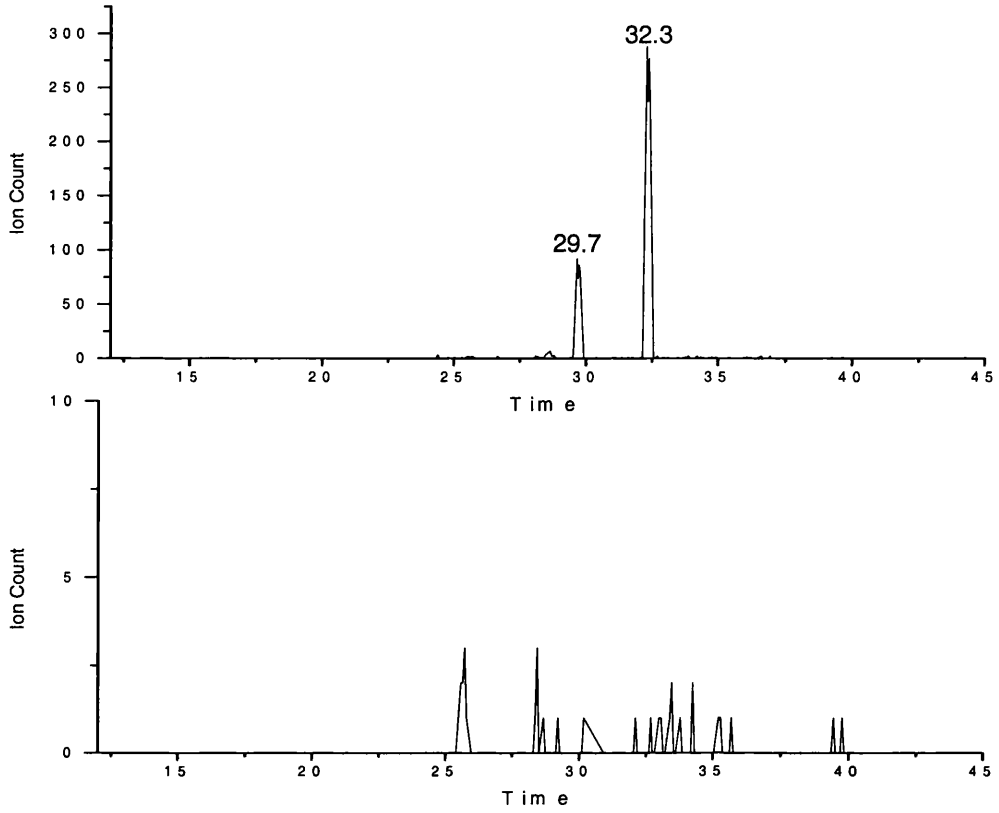


Figure 5.33: Extracted ion chromatograms of m/z 204.1 in the two CID-MS-MS functions of a chymotryptic digest of SRF.

Two peaks are observed in the upper chromatogram at retention times of 29.7 minutes and 32.3 minutes.

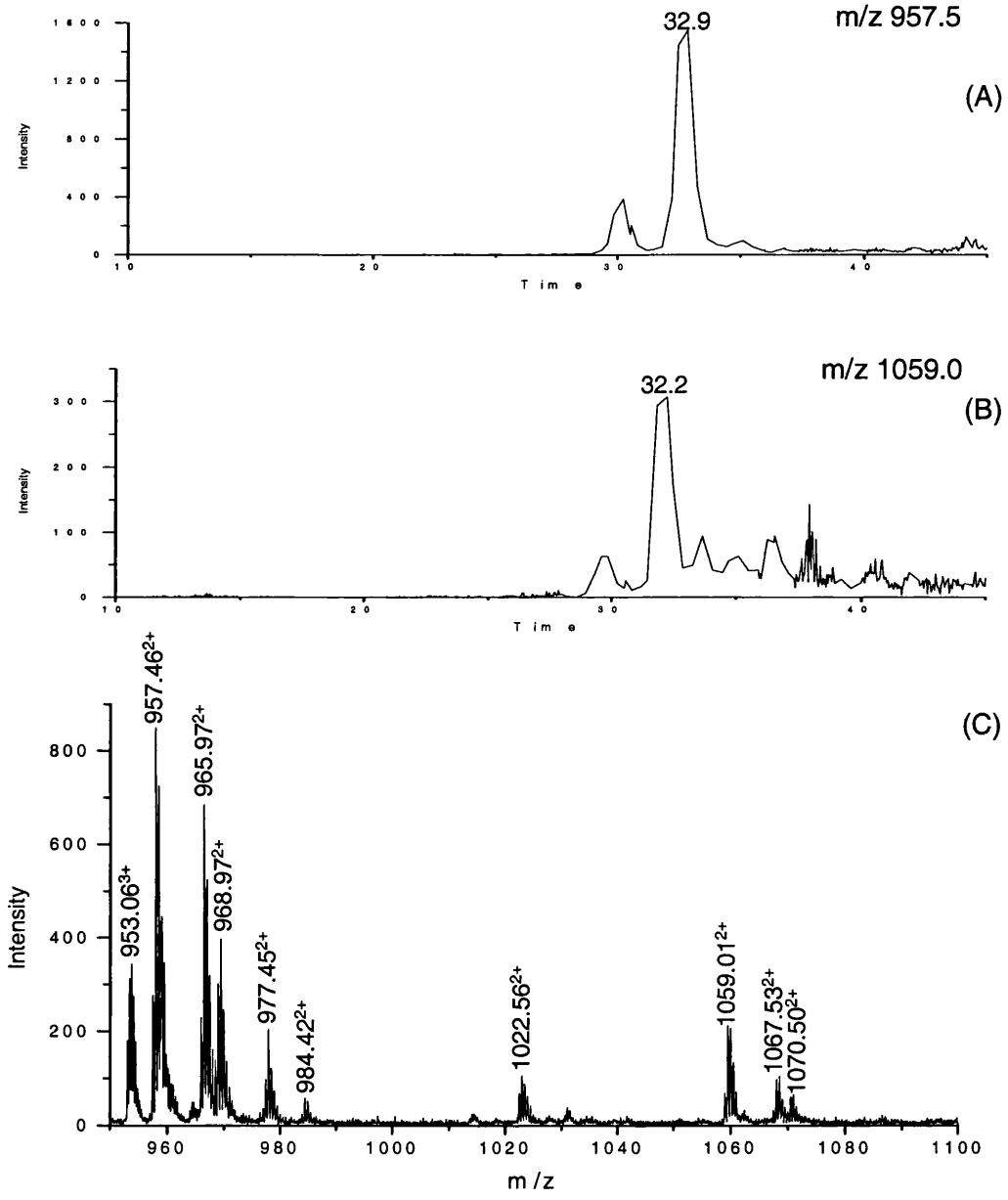


Figure 5.34: Extracted ion chromatograms of unmodified and GlcNAc-modified versions of the peptide spanning residues 303 - 323 of SRF.

Extracted ion chromatograms of the (A) unmodified and (B) GlcNAc-modified peptide. (C) The combined mass spectrum during the period in which the unmodified and GlcNAc-modified peptides elute contains the unmodified peptide at $[M + 2H]^{2+}$ m/z 957.46 and the glycosylated peptide at $[M + 2H]^{2+}$ m/z 1059.01.

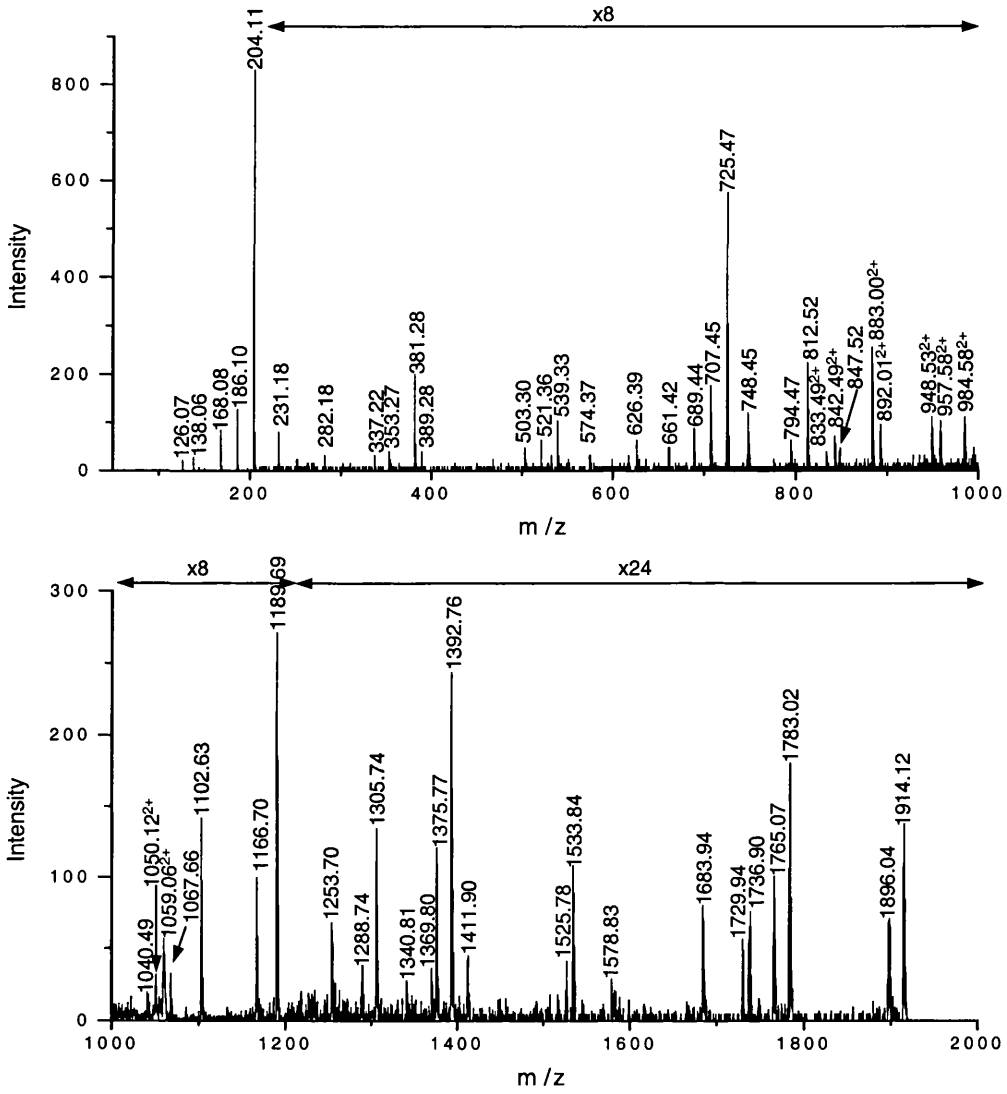
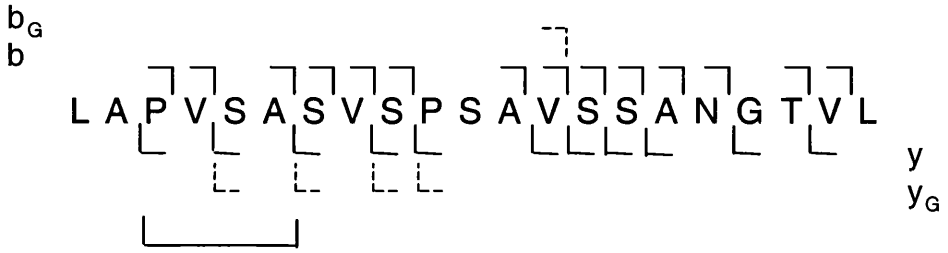


Figure 5.35: ESI-CID-MS-MS spectrum of the GlcNAc-modified peptide at [M + 2H]$^{2+}$ m/z 1059.01, corresponding to residues 303 – 323 of SRF.

This peptide corresponds to residues 303 – 323 of SRF, and peaks observed are identified in Table 5.18. The ions at m/z 1369.80 ($b13_G$) and m/z 1392.76 ($y13_G$) determine the modified residue as serine 313.

Table 5.18: Identities of the peaks observed in the ESI-CID-MS-MS spectrum of the GlcNAc-modified peptide [M + 2H]$^{2+}$ m/z 1059.01.

This fragmentation spectrum (Figure 5.35) is of residues 303 – 323 of SRF. A graphical representation of the fragment ions observed is also given. The ions at m/z 1369.80 (b_{13G}) and m/z 1392.76 (y_{13G}) identify the modified residue as serine 313.



Peak	Match	Peak	Match	Peak	Match
126.07	GlcNAc Fragment	725.47	b8	1253.70	b14
138.06	GlcNAc Fragment	748.45	y8	1288.74	$y12_G$-NH_3
168.08	GlcNAc Fragment	794.47	b9-H_2O	1305.74	$y12_G$
186.10	GlcNAc Fragment	812.52	b9	1340.81	b15
204.11	GlcNAc	833.49^{2+}	y19-H_2O	1369.80	$b13_G$
231.18	y2	842.49^{2+}	y19	1375.77	$y13_G$-NH_3
282.18	b3	847.52	y9	1392.76	$y13_G$
337.27	PVSA-H_2O	883.00^{2+}	b20-H_2O	1411.90	b16
353.27	a4	892.01^{2+}	b20	1525.78	b17
381.28	b4	948.53^{2+}	MH_2^{2+}-H_2O	1553.84	y17
389.28	y4	957.58^{2+}	MH_2^{2+}	1578.83	$y15_G$
503.30	b6-2H_2O	984.58^{2+}	$b20_G$-H_2O	1683.94	b19
521.36	b6-H_2O	1040.49		1729.94	y19
539.33	b6	1050.12^{2+}	$M_GH_2^{2+}$-H_2O	1736.90	$y17_G$
574.37	y6	1059.06^{2+}	$M_GH_2^{2+}$	1765.07	b20-H_2O
626.39	b7	1067.66	b12	1783.02	b20
661.42	y7	1102.63	y12	1896.04	MH^+-H_2O
689.44	b8-2H_2O	1166.70	b13	1914.12	MH^+
707.45	b8-H_2O	1189.69	y13		

The peak in the extracted ion chromatogram of m/z 204.1 (Figure 5.33) at 29.7 minutes is formed by the CID-MS-MS spectrum of $[M + 3H]^{3+}$ m/z 851.601. This is a GlcNAc-modified peptide of residues 396 – 417. Glycosylated peptides were observed from this region of the protein in the tryptic digest, and a previously unidentified site of GlcNAc modification was detected between residues 396 and 406. The GlcNAc modification site in this peptide will be the same as that observed in the tryptic digest. Figure 5.36 shows extracted ion chromatograms for the modified and unmodified peptide, as well as a mass spectrum containing both species. The GlcNAc-modified peptide eluted before the unmodified peptide, as is expected.

The CID-MS-MS spectrum of the modified peptide at $[M + 3H]^{3+}$ m/z 851.60 is shown in Figure 5.37, and the fragment ions are identified in Table 5.19. The spectrum contains a huge number and variety of fragment ions. There are 13 'b' ions, but, unfortunately, none of them are observed in a glycosylated state. There are ten 'y' ions, and GlcNAc-modified $y18_G$ - $y20_G$ are detected. There are also a large number of internal fragment ions, whose formation was favoured by the presence of multiple proline residues in the sequence, and three of these internal ions are observed in a glycosylated state. These glycosylated ions discount the four most N-terminal residues as sites of modification, and restrict the GlcNAc-modification site to one of threonine 401, threonine 402 or serine 411. The previous findings from the tryptic digest (Section 5.2.2) dismissed the serine, so the newly detected site of modification is one of the threonine residues.

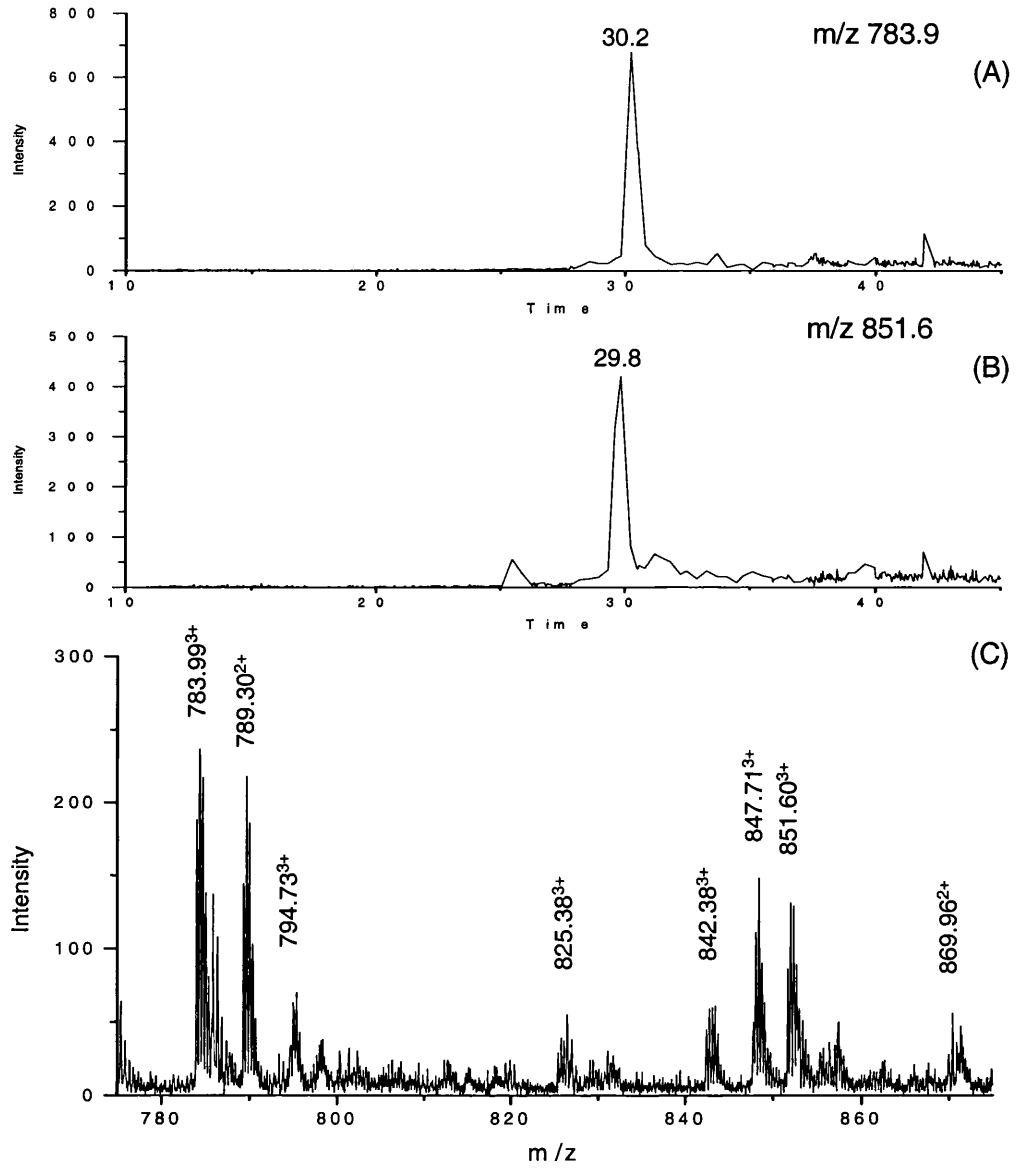


Figure 5.36: Extracted ion chromatograms of unmodified and GlcNAc-modified versions of the peptide spanning residues 396 - 417 of SRF.

Extracted ion chromatograms of (A) unmodified and (B) GlcNAc-modified peptide. (C) The combined mass spectrum during the period in which the unmodified and GlcNAc-modified peptides elute contains the unmodified peptide at $[M + 3H]^{3+}$ m/z 783.99 and the glycosylated peptide at $[M + 3H]^{3+}$ m/z 851.60.

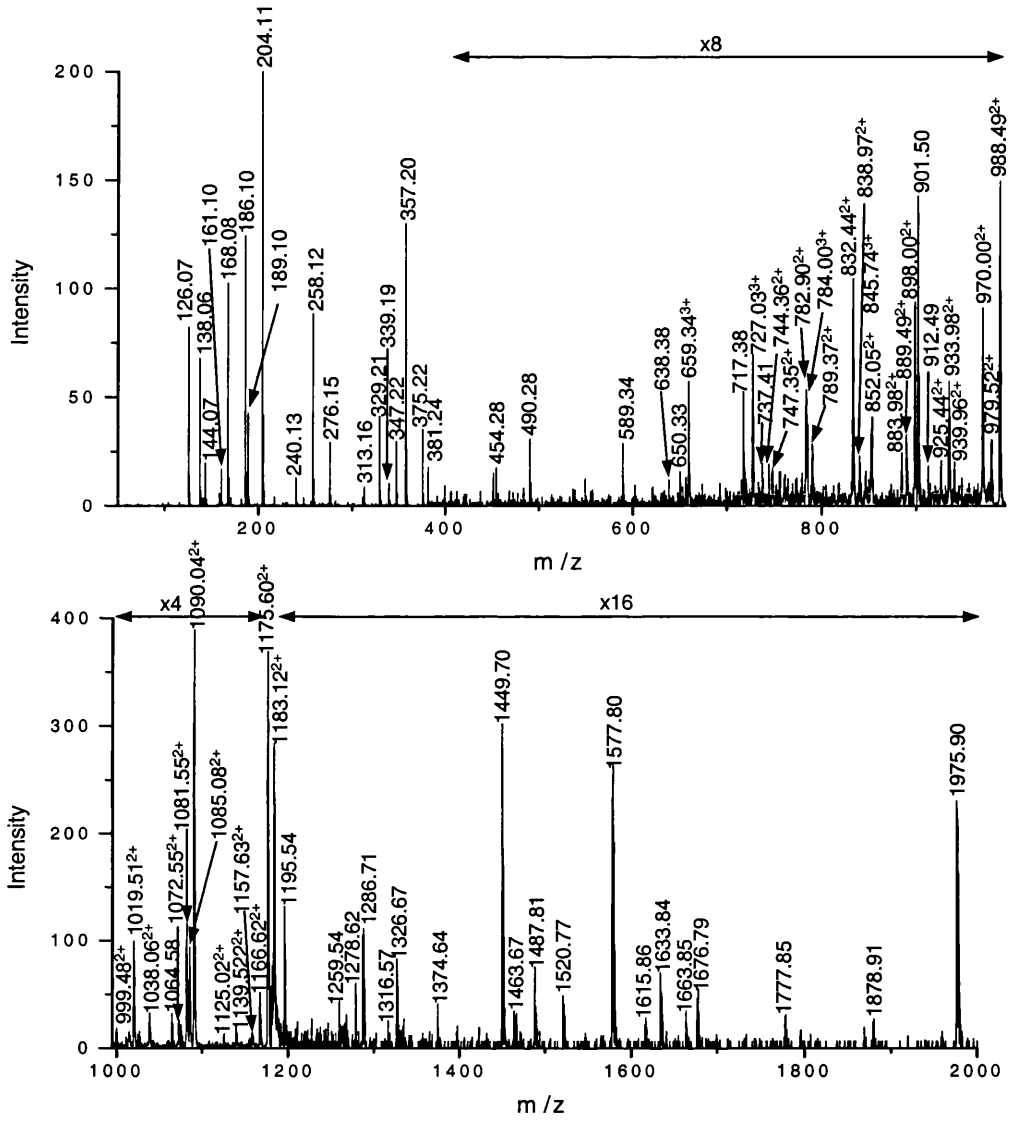
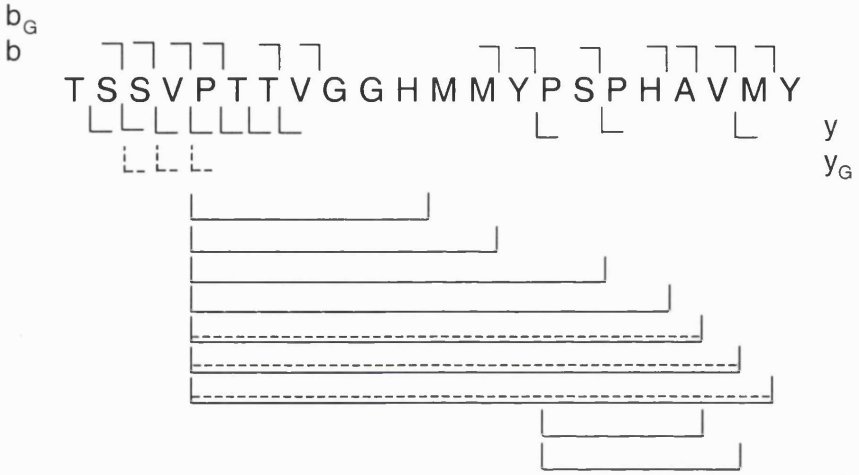


Figure 5.37: ESI-CID-MS-MS spectrum of the GlcNAc-modified peptide [M + 3H]$^{3+}$ m/z 851.66, corresponding to residues 396 - 417 of SRF.

Peaks observed are identified in Table 5.19.

Table 5.19: Identities of the peaks observed in the ESI-CID-MS-MS spectrum of the GlcNAc-modified peptide $[M + 3H]^{3+}$ m/z 851.66.

The fragmentation spectrum (Figure 5.37) is of the peptide spanning residues 396 – 417 of SRF. A graphical representation of the fragment ions observed is also given.



Peak	Match	Peak	Match	Peak	Match
126.07	GlcNAc Fragment	744.36^{2+}		1090.04^{2+}	$y18_G$
138.06	GlcNAc Fragment	747.35^{2+}	PTTVGGHMMYPSPH	1125.02^{2+}	y21
144.07	GlcNAc Fragment	782.90^{2+}	PTTVGGHMMYPSPHA	1139.52^{2+}	$y19_G$
161.10	a2	784.00^{3+}	TTVGGHMMYPSPHAV	1157.63^{2+}	MH_2^{2+}-$2H_2O$
168.08	GlcNAc Fragment	789.37^{2+}	y14	1166.62^{2+}	MH_2^{2+}-H_2O
186.10	GlcNAc Fragment	832.44^{2+}	PTTVGGHMMYPSPHAV	1175.60^{2+}	MH_2^{2+}
189.10	b2	838.97^{2+}	y15	1183.12^{2+}	$y20_G$
204.11	GlcNAc	845.74^{3+}	MH_G^{3+}-H_2O	1195.54	y10
240.13	b3-$2H_2O$	852.05^{2+}	MH_G^{3+}	1259.54	PTTVGGHMMYPS
258.12	b3-H_2O	883.98^{2+}	PTTVGGHMMYPSPHA$_G$	1278.62	
276.15	b3	889.49^{2+}	y16	1286.71	b13
313.16	y2	898.00^{2+}	PTTVGGHMMYPSPHAVM	1316.57	
329.21	a4-H_2O	901.50	y8	1326.67	y11
339.19	b4-$2H_2O$	912.49	PTTVGGHMM	1374.64	
347.22	a4	925.44^{2+}	b18-H_2O	1449.70	b14
357.20	b4-H_2O	933.98^{2+}	PTTVGGHMMYPSPHAV$_G$	1463.67	y12
375.22	b4	939.96^{2+}	y17	1487.81	
381.24	VPTT-H_2O	970.00^{2+}	b19	1520.77	y13
454.28	b5-H_2O	979.52^{2+}	y18-H_2O	1577.80	y14
490.28	PSPHA	988.49^{2+}	y18	1615.86	b16-H_2O
589.34	PSPHAV	999.48^{2+}	PTTVGGHMMYPSPHAVM$_G$	1633.84	b16
638.38	b7-$2H_2O$	1019.51^{2+}	b20	1676.79	y15
650.33	PTTVGGH	1038.06^{2+}	y19	1777.85	y16
659.34^{3+}	y18	1064.58	y9	1878.91	y17
717.38	y6	1072.55^{2+}	$y18_G$-$2H_2O$	1975.90	y18
727.03^{3+}	$y18_G$	1081.55^{2+}	$y18_G$-H_2O		
737.41	b8-$2H_2O$	1085.08^{2+}	b21		

5.4.3. Pro-C Digestion of Chymotryptic HPLC Fraction

To determine which of the two threonine residues is modified, a second chymotryptic digest of SRF was carried out. This digest was separated by HPLC, and fractions were collected. Each fraction was then screened by MALDI-MS to locate the fraction that contained the GlcNAc-modified peptide corresponding to residues 396 - 417. The mass spectrum of fraction nine is shown in Figure 5.38. This fraction contains several peptides, but ions at m/z 2350.24, m/z 2366.20, m/z 2382.20 and m/z 2398.13 are the non-glycosylated peptide 396 - 417 containing none, one, two and three oxidised methionine residues respectively. Also observed are weak ions at m/z 2569.38 and m/z 2585.42 which are the GlcNAc-modified versions of this peptide containing one and two oxidised methionine residues.

This fraction was then sub-digested with pro-C, which cleaves peptides C-terminal to proline residues, then analysed by LC-MS. If this enzyme worked efficiently, it should have digested the GlcNAc-modified peptide to produce the peptide TTVGGHMMYP. Figure 5.39 shows the extracted ion chromatograms of m/z corresponding to doubly-charged peaks for this peptide in non-glycosylated and GlcNAc-modified states containing one and two oxidised methionine residues. A peak is observed in the chromatogram of the non-glycosylated, singly oxidised peptide, and a weaker peak is observed for the doubly oxidised version of this peptide. However, there are no peaks for GlcNAc-modified versions of this peptide; the peak after 28.5 minutes in the chromatogram of m/z 656.8 relates to a singly charged peak at m/z 657.38. Figure 5.40 confirms that the two peaks in the ion chromatograms of the unmodified species relate to doubly-charged peptides of the correct masses. Thus, the enzyme had successfully produced unmodified peptides of this region, but there was no evidence of GlcNAc-modified peptides.

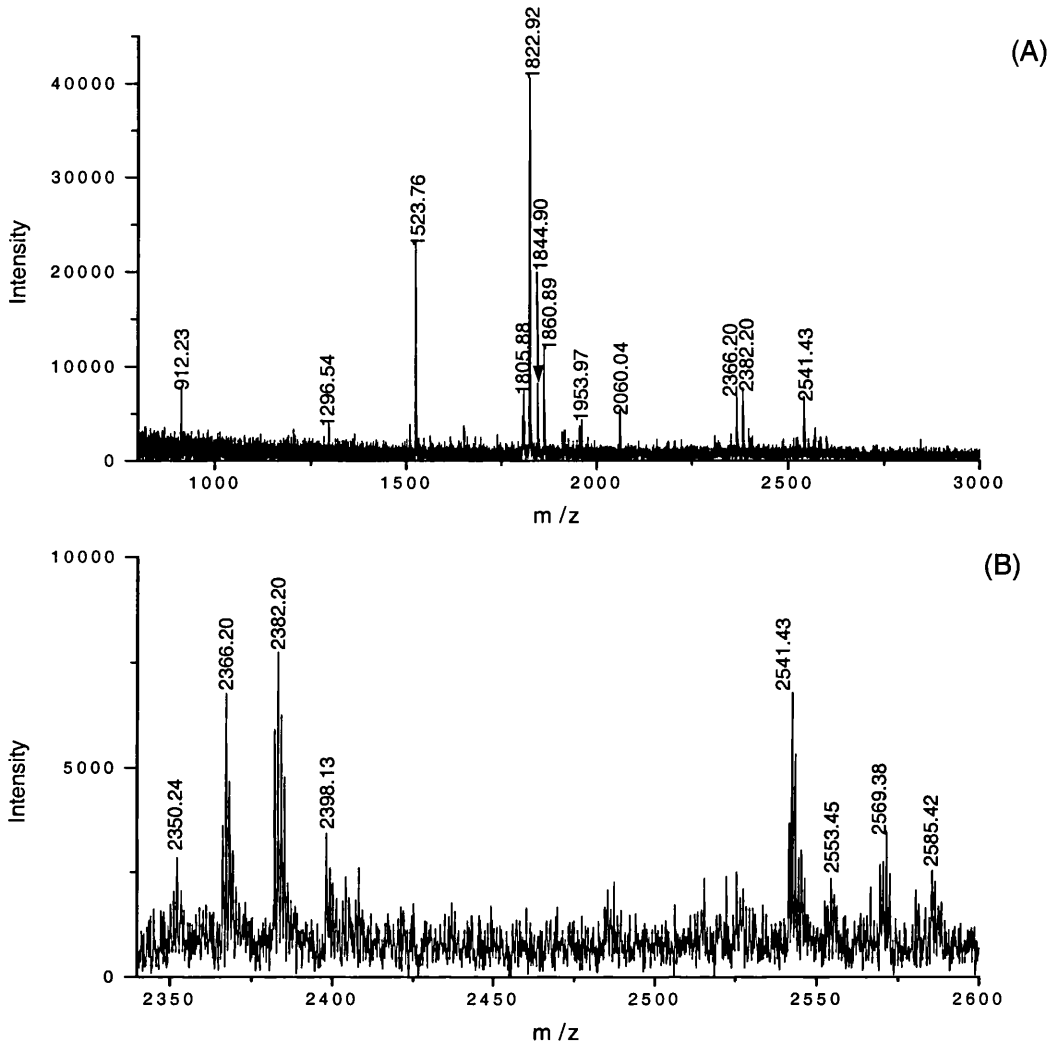


Figure 5.38: MALDI mass spectrum of HPLC fraction nine from a chymotryptic digest of SRF.

(A) The whole spectrum illustrates this fraction contains several peptides. (B) Included in these are the unmodified and GlcNAc-modified version of the peptide spanning residues 396 – 417. The unmodified peptide is observed with no, one, two and three oxidised methionine residues at m/z 2350.24, m/z 2366.20, m/z 2382.20 and m/z 2398.13. The glycosylated peptide is detected with one and two oxidised methionine residues at m/z 2569.38 and m/z 2585.42.

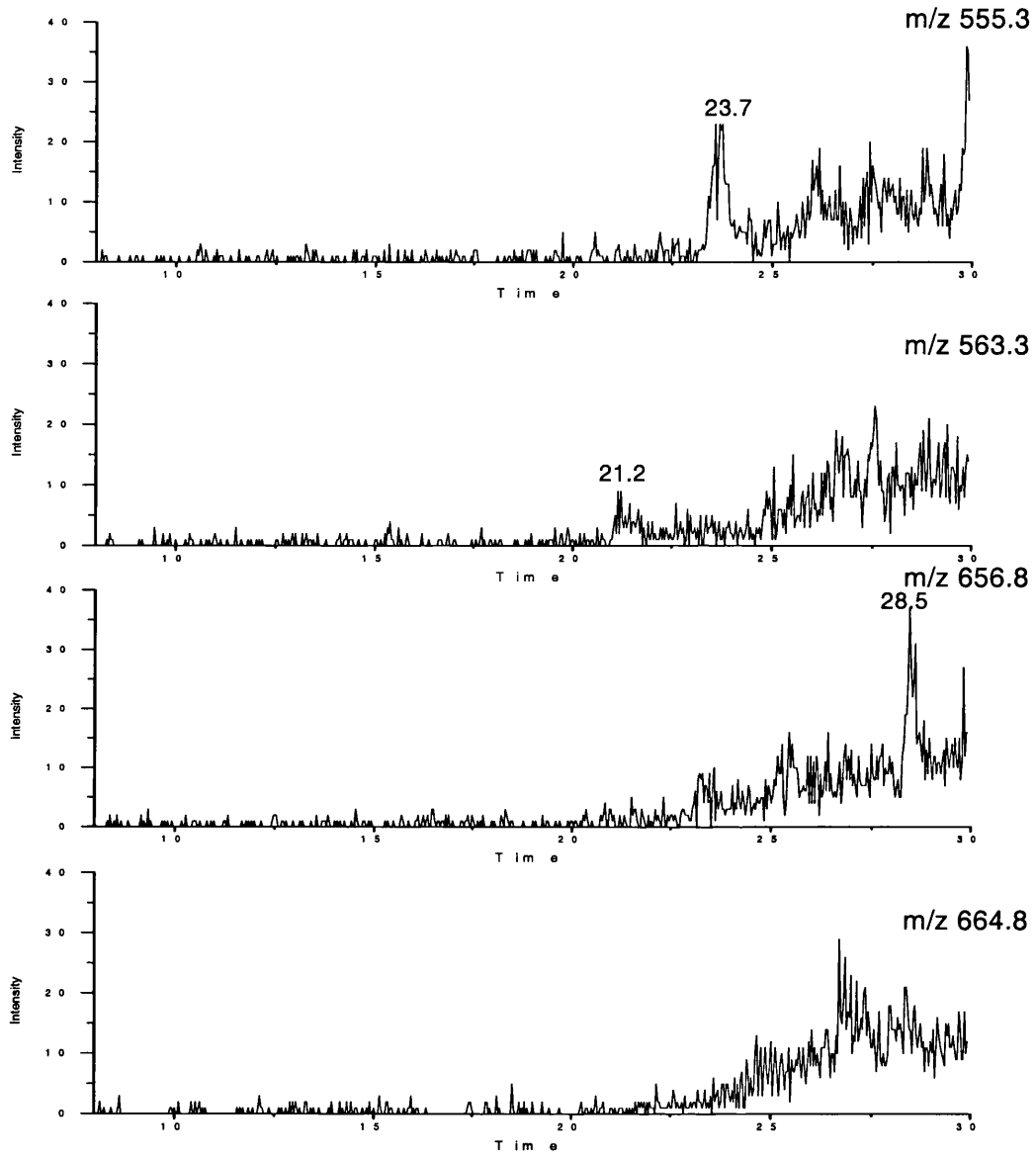


Figure 5.39: Extracted ion chromatograms of the doubly-charged versions of the peptide TTVGGHMMYP from an LC-MS analysis of a pro-C digest of HPLC fraction nine from a chymotryptic digest of SRF.

A peak is observed after 23.7 minutes in the chromatogram for the non-glycosylated singly methionine-oxidised peptide (m/z 555.3), and a weak peak is observed after 21.2 minutes in the chromatogram of the non-glycosylated doubly methionine-oxidised peptide (m/z 563.3), but there was no evidence for GlcNAc-modified versions of this peptide at either oxidation state. The peak after 28.5 minutes in the chromatogram of m/z 656.8 related to a singly charged peak at m/z 657.37.

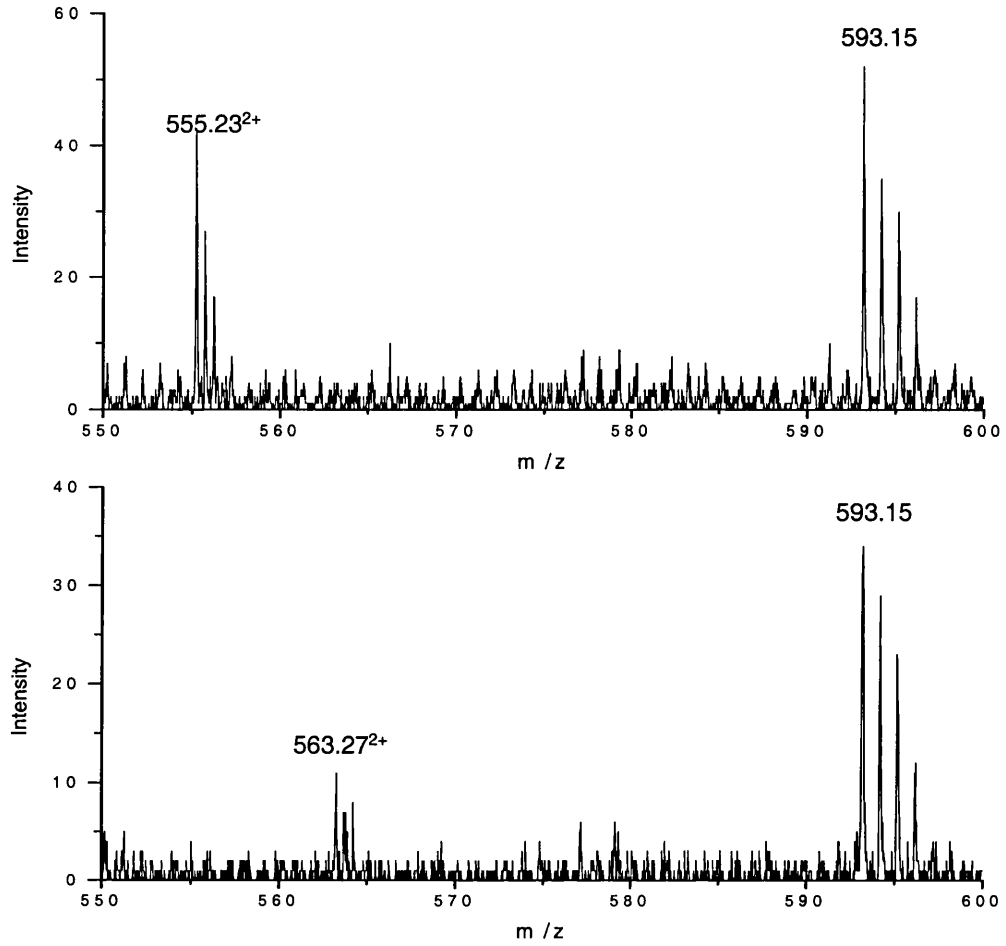


Figure 5.40: ESI-MS spectra from an LC-MS run of a pro-C digest of HPLC fraction nine of a chymotryptic digest of SRF.

(A) The peak observed at $[M + 2H]^{2+}$ m/z 555.23 is the predicted mass for the singly methionine-oxidised non-glycosylated peptide TTVGGHMMYP and (B) the peak at $[M + 2H]^{2+}$ m/z 563.27 is the doubly methionine oxised version of the same peptide.

Figure 5.41 shows extracted ion chromatograms of non-glycosylated and GlcNAc-modified peptides with one and two oxidised methionines for the peptide spanning residues 396 to 410 (TSSVPTTVGGHMMYP), which is the product of one missed cleavage. These chromatograms contain no peak for a singly oxidised non-glycosylated peptide, and the peak in the doubly oxidised ion chromatogram is formed by a singly charged peak at m/z 799.39, so is not the peak of interest. However, there are peaks in the chromatograms for GlcNAc-modified singly and doubly methionine-oxidised species,

and the spectra of the chromatogram at the times these peaks appear in the extracted ion chromatograms (Figure 5.42) show these to be doubly-charged peptides of the correct mass. Hence, this peptide is observed only in a GlcNAc-modified state.

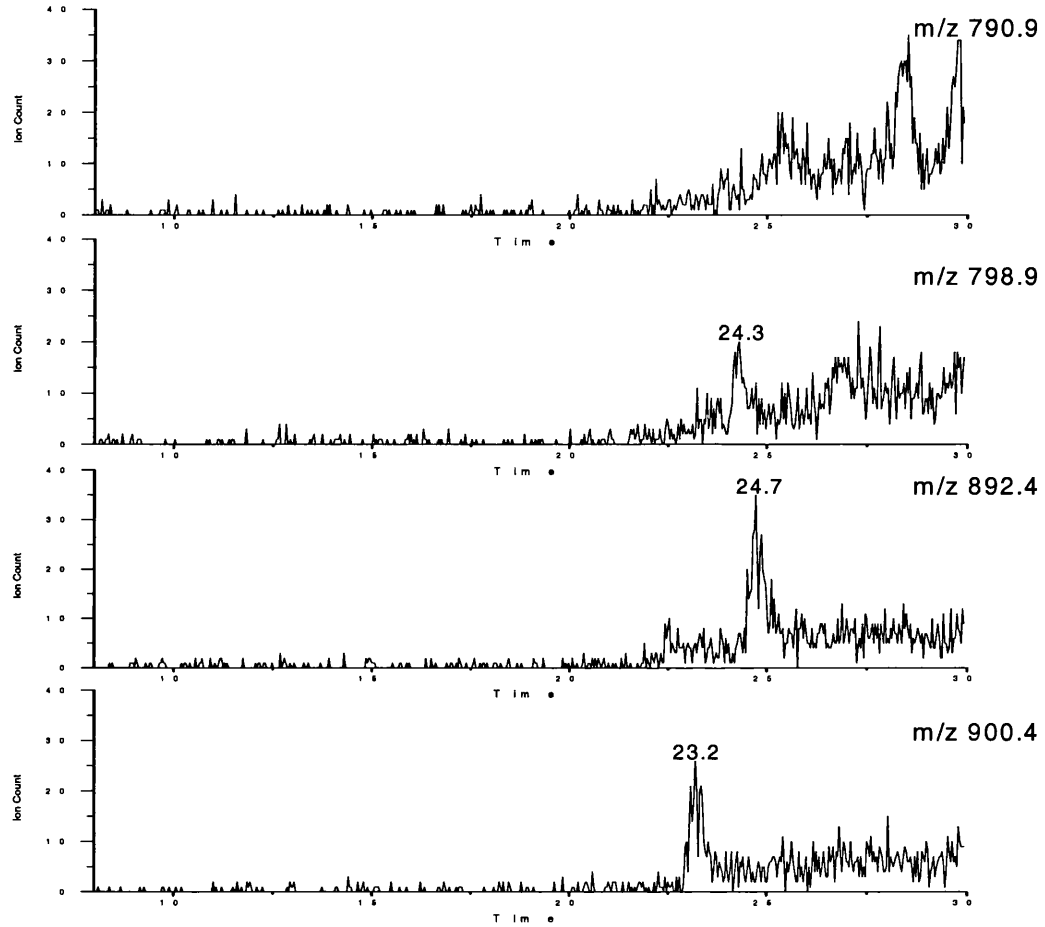


Figure 5.41: Extracted ion chromatograms of doubly-charged versions of the peptide TSSVPTTVGGHMMYP from an LC-MS run of a pro-C digest of HPLC fraction nine from a chymotryptic digest of SRF.

A peak is observed after 24.7 minutes in the chromatogram for the GlcNAc-modified singly methionine-oxidised peptide (m/z 892.4), and a peak is also observed after 23.2 minutes in the chromatogram of the GlcNAcylated doubly methionine-oxidised peptide (m/z 563.3). However, there is no evidence for non-glycosylated versions of this peptide at either oxidation state. The peak after 24.3 minutes in the chromatogram of m/z 798.9 relates to a singly charged peak at m/z 799.39.

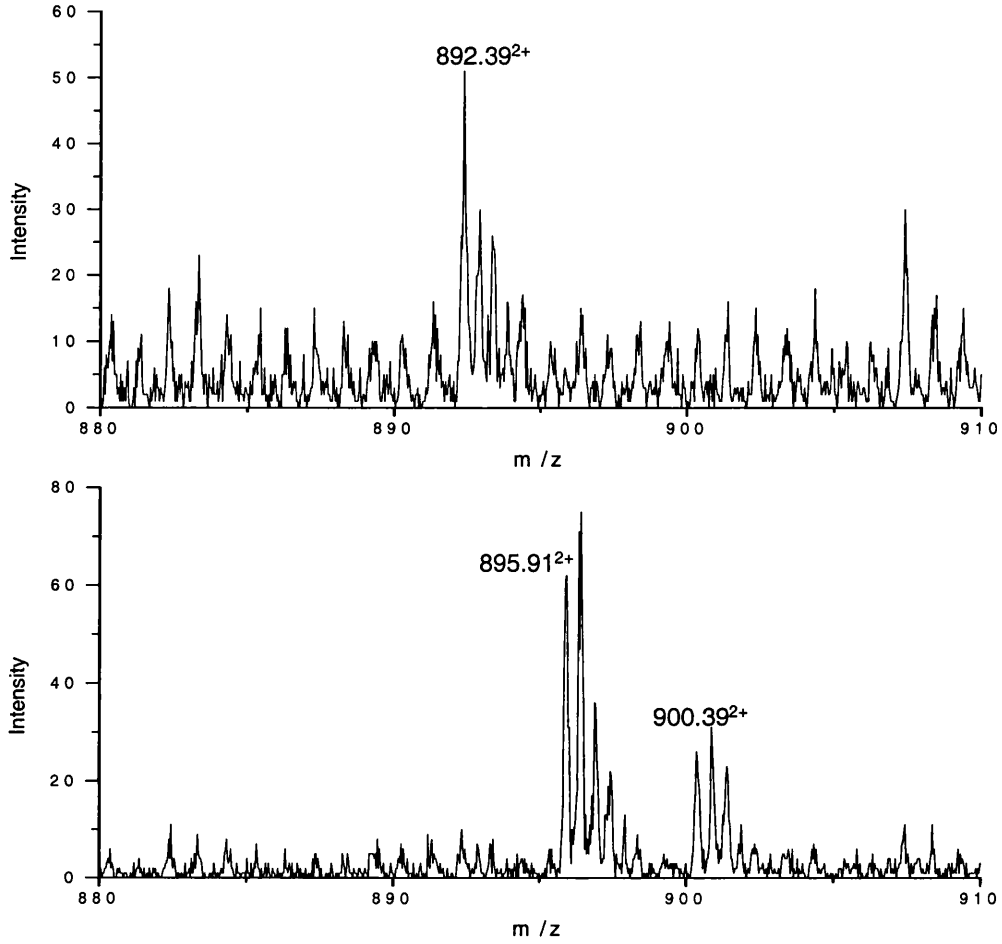


Figure 5.42: ESI-MS spectra from an LC-MS analysis of a pro-C digest of HPLC fraction nine from a chymotryptic digest of SRF.

(A) The peak observed at $[M + 2H]^{2+}$ m/z 892.39 is the predicted mass of the singly methionine-oxidised GlcNAc-modified peptide TSSVPTTVGGHMMYP, and the peak at $[M + 2H]^{2+}$ m/z 900.39 (B) is the doubly methionine-oxidised version of the same peptide.

Together, these results suggest the pro-C enzyme normally cleaves this peptide after proline 400 to produce the peptide residues 401 - 410. However, the presence of the GlcNAc moiety prevents cleavage at this site, leading to the peptide spanning residues 396 - 410 being produced instead. Thus, the GlcNAc residue causes steric hindrance, preventing the enzyme from attacking this site. This strongly suggests the site of

modification is the immediately neighbouring threonine 401. Indeed, Reason *et al.* proposed that a GlcNAc on a neighbouring residue would prevent enzymatic cleavage[92].

This previously unknown site of modification is the first modified threonine identified in SRF; all other known sites are on serine residues.

5.5. Discussion

Between the different digests, 100% of the previously reported sequence of the protein was observed, as well as the N-terminal myc tag (Figure 5.4). This allowed comprehensive analysis of the protein for post-translational modifications. Several known sites of modification were confirmed, new sites were determined, and evidence of further modifications was produced. Most of the modifications characterised in this work were present at reasonably high stoichiometry: for all sites identified greater than 10% of the protein appeared to be modified by comparison of the ratio of intensities of unmodified and modified peaks. This stoichiometry of modification is higher than observed for sites in many proteins. Some of the published modification sites have not been detected in this work. This is likely to be due to the screening process used here not being able to detect low stoichiometry modifications. In order to find these, selective enrichment of modified peptides could be employed to simplify the sample [169, 171].

Of the published phosphorylation sites serine 83 (serine 95 in the myc-tagged sequence) was not observed [188]. This region of the protein was not observed in the tryptic digest. In the trypsin/chymotrypsin combined digest it was observed as a non-phosphorylated peptide in the MALDI mass spectrum at m/z 3552.67, spanning residues 67 – 103 (Figure 5.24). A peak 80 Da larger in mass at m/z 3632.55 was also observed. However, this mass also matches a predicted tryptic peptide spanning residues 108 – 146. There was no data to differentiate between these two peptides, as this peak was not observed in the LC-ESI-MS analysis of the sample. A peptide corresponding to residues 85 – 113 was observed in the chymotryptic digest, but no peak for a phosphorylated peptide was detected, and this region was not identified in the LC-ESI-CID-MS-MS analysis of the sample.

A phosphorylated peptide spanning residues 101 – 135 was analysed by CID-MS-MS (Figure 5.9) and the phosphorylation site was shown to be either serine 101 or serine 103, with serine 103 being more likely, which is a well characterised site of modification [185-187].

The region containing the other published phosphorylation site at serine 435 [189] was only observed in the chymotryptic digest, and there was no evidence for a phosphorylated peptide of this region in this particular protein preparation.

A previously uncharacterised site of phosphorylation was identified in the tryptic digest of SRF at serine 224 in the published sequence (Figure 5.7). This is immediately after a region determined to be important for DNA binding [194]. Thus, phosphorylation at this site may regulate the protein's ability to bind to the SRE.

In the tryptic digest, the peptide spanning residues 261 – 324 was observed in a singly-modified state in the MALDI-MS spectrum (Figure 5.1), but a doubly-GlcNAc-modified state was also seen by ESI-MS (Figure 5.21). This region contains four published GlcNAc modification sites on serines 285, 307, 309 and 316 [92]. It is not surprising that an extra modification was detected using ESI-MS, since it is well known as a softer method of ionisation compared with UV MALDI (see also Section 3.1). Peptides spanning residues 303 – 323 and 303 – 324 were observed in chymotryptic and tryptic/chymotryptic combined digests respectively. However, they were observed with only one GlcNAc modification. Hence, one of the GlcNAc modification sites in peptide 261 – 324 is outside this region. Thus, a GlcNAc modification site was detected in the region 261 – 303, which is probably the published residue serine 285.

A GlcNAc modification site in the peptide 303 – 323 was identified as serine 313. This is a previously undetected site of modification, and is the first novel site of GlcNAc modification in any protein to be identified by mass spectrometry. This site was not observed in the work of Reason *et al.* [92]. One of the glycopeptides they focused on spanned residues 313 – 324, reporting the GlcNAc modification on serine 316. The probable reason they did not observe serine 313 GlcNAc-modified is inherent in the approach they employed to find the GlcNAc-modified peptides. They used a sequential digestion approach by cyanogen bromide followed by trypsin, and then a proline-specific endopeptidase (pro-C) to produce their peptides. This peptide mixture was then

separated by HPLC, and fractions were collected. Fractions were screened for pairs of peaks that differed by 203 Da in mass, which is characteristic of an unmodified and GlcNAc-modified peptide. However, if residue 313 were GlcNAc-modified, this would probably prevent pro-C from cleaving after proline 312 due to steric effects from the proximal sugar residue (further evidence for this effect is in Section 5.4.3). Hence, this site will be cleaved only when serine 313 is unmodified, but if there is a bulky sugar moiety attached it will be inaccessible, so a glycosylated peptide of residues 306 - 324 would be the product of this combined enzymatic digestion. Hence, there would not be a pair of peptides differing by 203 Da in mass, and the modified peptide would not be detected by their screening process.

It is interesting that this site appeared to be heavily modified, but there was no evidence for any modification at serine 316, as there was no sign of a doubly GlcNAc-modified version of this peptide in the extracted ion chromatogram from this digest (data not shown). This suggests that the stoichiometry of modification at serine 313 is considerably higher than at serine 316.

None of the previously identified GlcNAc modification sites in the peptide spanning residues 303-324 were detected, even though these experiments use exactly the same protein preparation of SRF[164]. This shows different selectivity for the two approaches to identifying GlcNAc modification sites. By using radioactivity to find GlcNAc-modified peptides, sites modified at lower stoichiometry could be detected and identified. The pure mass spectrometry based approach used in this chapter, although extremely sensitive, struggles to find sites that are modified at low stoichiometry.

Further sites of GlcNAc modification in SRF were detected during this work. Data from a tryptic digest detected a previously unknown site of GlcNAc modification between residues 395 and 406. This site was narrowed to either threonine 401 or threonine 402 from a CID fragmentation spectrum of a peptide spanning residues 396 - 417 produced by a chymotryptic digest (Figure 5.37), and a pro-C sub-digest of this peptide provided strong evidence that threonine 401 was the modified residue. It may be possible to confirm this site using this same sub-digestion approach, but starting with more sample, such that there would be enough modified peptide after the pro-C digest to get CID-MS-MS data on the newly formed glycopeptide. An alternative approach would

be to sub-digest the purified glycopeptide with a non-specific enzyme such as pepsin, and see if a glycosylated peptide formed by a cleavage between the two threonine residues is observed. However, this approach may not be successful, as the GlcNAc will probably prevent cleavage between the residues in the same way it prevented cleavage by pro-C. Also, a lot of starting material would be required so there is a detectable amount of GlcNAc-modified peptide formed by cleavage between the two threonines, as cleavages at all other sites would also be formed.

A peptide spanning residues 249 to 270 detected in a combined tryptic/chymotryptic digest was observed in a GlcNAc-modified state. The CID spectrum of this peptide (Figure 5.29) contained no glycosylated fragments, so provided no further information on the site. However, an intense peak at m/z 2699.21 in the MALDI-MS spectrum of tryptic SRF (Figure 5.2), and the corresponding peptide in the LC-MS run at $[M + 3H]^{3+}$ m/z 900.40 (corresponding to residues 236 – 260) were not observed GlcNAc-modified, suggesting the site of GlcNAc modification is in the region spanning residues 261 – 270. Interestingly, there are no serine residues in this sequence. Thus, a threonine appears to be the modified residue. This region of the protein proved difficult to observe. Indeed, the peptide spanning residues 249 – 270 was the only time it was assigned in any of the digestions of SRF, and this peptide was a product of a non-specific cleavage.

In this study none of the sites of O-GlcNAc modification and sites of phosphorylation occurred on the same residue. Hence, there was no direct evidence that the two modifications are interacting with each other. Indeed, the GlcNAc modification sites appear to be concentrated in a region of the protein for which there is no evidence of phosphorylation.

Sites of phosphorylation and O-GlcNAcylation were found within the same sample. This suggests that the two modifications are not mutually exclusive at the protein level. Hence, the two modifications probably serve different roles in controlling the function and activity of the protein.

6. PROTEOMICS TO FIND O-GLCNAC MODIFIED PROTEINS

6.1. Introduction

The standard method used today for detecting GlcNAc modification of proteins is essentially the same as was originally employed to identify the modification in 1984[67]. Terminal GlcNAc moieties are *in vitro* labelled by [^{3}H]galactose using a galactosyltransferase[67, 156, 162]. To find GlcNAc-modified proteins from within a mixture, proteins are separated by 1D or 2D PAGE. Modified proteins are detected through the radioactive label by exposing the gel to film. However, this is not a particularly sensitive method of detection as tritium has very low activity. To partly compensate for this, gels are generally soaked in fluorescent scintillants such as EN^{3}HANCE that convert radioactivity into fluorescence and magnify the signal[162]. Despite this, the sensitivity of detection is two hundred fold less than phosphorylation detection using ^{32}P[59].

An alternative method for identifying GlcNAc-modified proteins is through the use of lectins. Wheat germ agglutinin (WGA) binds to terminal GlcNAc residues and also, with lower affinity, sialic acid residues. Its affinity of interaction with conventional branched sugar chains found on extracellular proteins is high. However, its affinity for single GlcNAc moiety is dramatically lower[195]. Nevertheless, if extracellular proteins can be removed from the sample, then the only sugar residues the WGA should bind to will be nuclear and cytoplasmically localised O-GlcNAc-modified proteins. Thus, GlcNAc-modified proteins can be detected by blotting proteins onto membrane, then probing with WGA conjugated to horseradish peroxidase (WGA-HRP) followed by enhanced chemiluminescence (ECL)[88, 156].

WGA has also been used for affinity chromatography to isolate GlcNAc-modified proteins from mixtures[168]. It has been mainly used for purifying over-expressed proteins that are highly modified. Its ability to isolate proteins with only one or two modification sites is not well established, but it isolates a large number of proteins from a whole cell lysate[155].

A different method of affinity purification of GlcNAc-modified proteins is to label the GlcNAc with a galactose residue, then use a different lectin, RCA I, which is specific for terminal galactose residues. However, like WGA, although the affinity of binding is high to multiple residues, the affinity of binding to a singly modified protein or peptide is much lower. Thus, only retardation of singly-modified peptides on the affinity column rather than retention is achieved[169, 171].

6.2. *Nuclear Isolation*

It was decided to isolate nuclei from cells in order to remove other forms of glycosylation, thus facilitating the search for O-GlcNAc-modified proteins. The cell line chosen for this study was the epidermal growth factor receptor (EGFR) over-expressing head and neck tumour carcinoma cell line HN5[166]. A number of nuclear isolation protocols were attempted to assess which protocol was the most effective.

A problem with nuclear isolation is that different protocols work better than others, depending on the cells the nuclei are to be purified from.

All protocols follow a general pattern of firstly swelling the cells in a hypotonic solution. The plasma membrane is then sheared, generally by homogenisation and cytoplasmic and plasma membrane proteins are removed by centrifugation. Some protocols then carry out further steps to wash the nuclei and to remove attached ER, golgi and other cellular debris.

The first few protocols attempted simply used hyptonic solutions to swell the nuclei, then homogenisation to shear the cell membranes. However, the HN5 cells proved too robust for these protocols, as the plasma membranes did not shear, even after repeated passes through a dounce homogeniser. It was thus concluded that a detergent would be necessary to dissolve the plasma membrane and facilitate the lysing of cells. Triton X-100 is a detergent that efficiently lyses the plasma membrane, but leaves the inner nuclear membrane intact[154, 196]. Thus, protocol II for preparation of crude nuclei from Gerner *et al.* was selected, as this is specifically designed for isolating nuclei from tissue culture cells[167]. This protocol efficiently lysed the plasma membranes whilst appearing to maintain the integrity of the nuclei (Figure 6.1B and C). This was confirmed by trypan blue staining being excluded from the nuclei (not shown). Many

nuclei were clumped together and cell debris was visible attached around the nuclear membrane, the majority of which was probably ER, as this is contiguous with the nuclear membrane. Thus, further washing steps were attempted to try to remove some of this contamination.

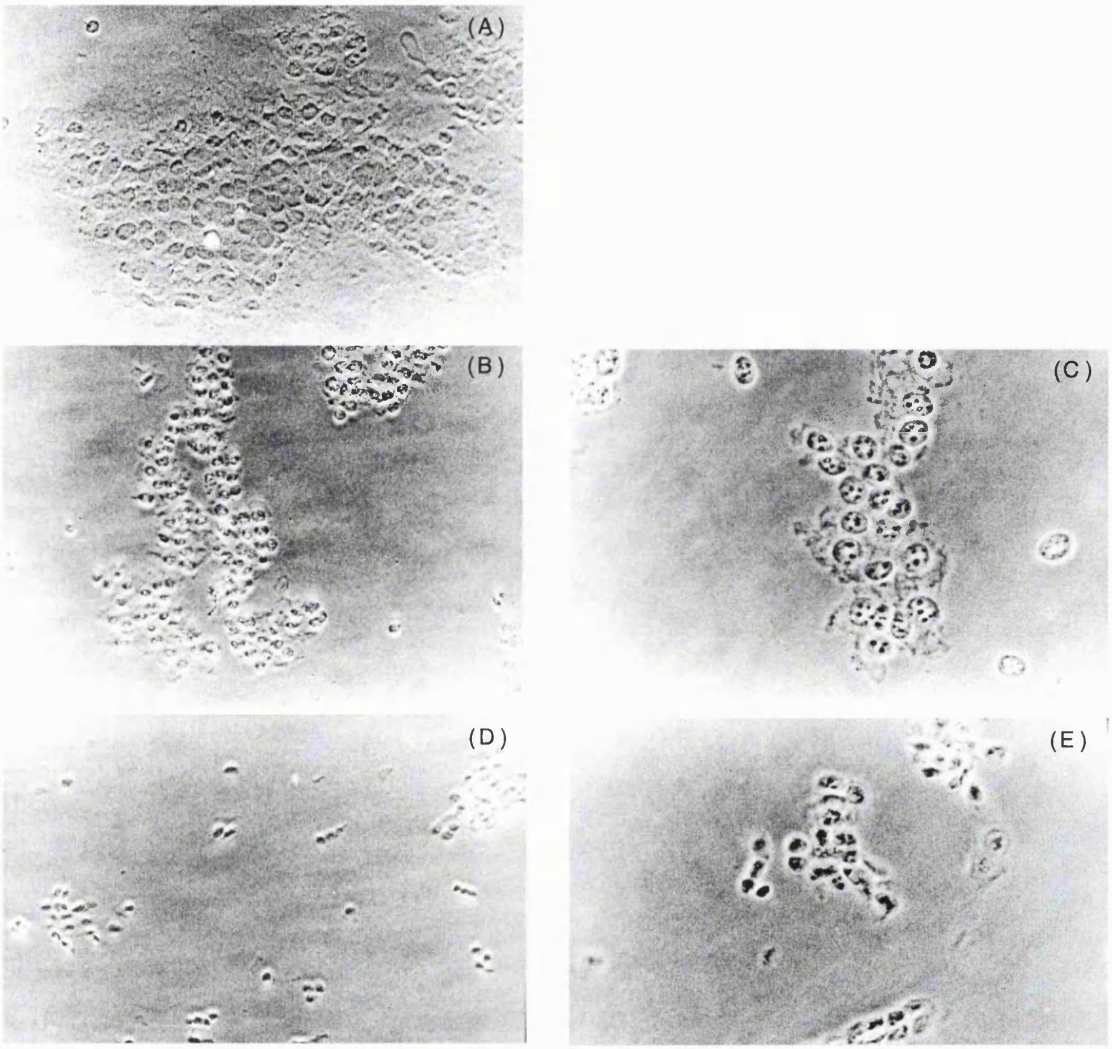


Figure 6.1: Light microscopy photographs of HN5 cells and nuclei at different stages of nuclear isolation.

(A) HN5 cells suspended in a hypotonic solution. (B and C) Crude nuclear fraction after triton X-100 treatment and homogenisation. (D and E) Clean nuclear fraction after tween-40 and sodium deoxycholate treatment in the presence of VRC and further homogenisation. Magnification x200 for (A), (B) and (D), x400 for (C) and (E).

Two different protocols for nuclear purification were compared. The first was based on a protocol used in the laboratory of Dr Ira Mellman (Ludwig Institute for Cancer Research, Yale University School of Medicine, CT), which involves repeated resuspension of the nuclear pellet in a hypotonic solution, homogenisation, then re-pelleting the nuclei. This was compared with the next step of the nuclear purification protocol II from Gerner *et al.*[167]. This protocol involves the addition of vanadyl ribonucleoside complexes (VRCs) to the nuclear washing buffer. VRC is an RNase inhibitor, and also helps to preserve the integrity of the nuclear matrix. This may be due to the nuclear matrix requiring intact RNA for structural integrity[197]. More aggressive washing can thus be used without apparent damage to the nuclei. Hence, this protocol adds the detergents Tween-40 and sodium deoxycholate to the nuclear wash buffer without apparent adverse affects, then further homogenises the nuclei.

In order to compare the two protocols a crude nuclear fractionation of HN5 cells was carried out[167]. The crude nuclear pellet was split in two, and one half was submitted to either the Gerner or Mellman protocol.

At the start there was a substantial white nuclear pellet. Following steps of nuclear purification using the Mellman protocol the pellet decreased dramatically in size. Upon examination of nuclei under the light microscope there appeared to be less contaminating protein surrounding the nuclear membrane, but there were many lysed nuclei, showing there was significant nuclear protein loss during the process.

In contrast the nuclear pellet following purification with the Gerner protocol did not decrease significantly in size. Following washing and homogenisation, nuclei were less clumped (Figure 6.1D) and the surrounding extranuclear debris was reduced (Figure 6.1E). Despite this, under the light microscope the majority of nuclei appeared to be intact.

These two cleaned nuclear fractions were loaded onto a 1D gel and proteins were separated. Proteins were transferred to membrane and probed against E-cadherin (a plasma membrane protein) and lamin B (a nuclear protein). Figure 6.2 shows the result of this blot. The level of E-cadherin decreased slightly after crude nuclear fractionation and was further reduced after nuclear purification to a similar level by both nuclear isolation protocols. The blot for the nuclear protein lamin B showed the level of this

protein decreased upon nuclear isolation, but was equivalent in the Gerner-purified nuclear fraction. However, there was no band visible in the lane corresponding to the Mellman protocol. This indicates that the majority of this protein was lost during the washing steps. This result shows the Gerner protocol to be the more effective preparation protocol for nuclei from HN5 cells.

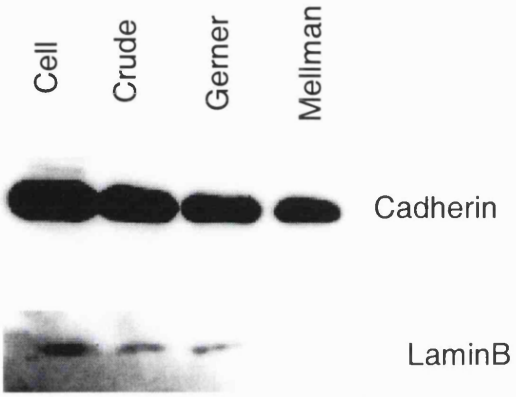


Figure 6.2: Comparative enrichment of nuclear proteins after different nuclear isolation protocols.

Fractions after each step of isolation were resolved by 1D SDS-PAGE, transferred to PVDF and probed for E-cadherin, a plasma membrane protein, and lamin-B, a nuclear protein. Blots were developed using ECL reagents.

The cleanliness of the purified nuclei was assessed in further western blot experiments. In each case the purified nuclei were compared with the crude nuclear preparation and a whole cell lysate. Figure 6.3 shows blots against E-cadherin, a plasma membrane protein; calnexin, a protein of the ER; MEK-2, a cytosolic protein and lamin B from the nucleus. As observed in the comparison of nuclear isolation protocols, the level of E-cadherin decreased upon nuclear isolation, but still gave a strong signal. The signal from calnexin decreased dramatically to almost nothing upon crude nuclear isolation, and the MEK-2 signal disappeared completely after the crude nuclear isolation step. Lamin B signal decreased upon nuclear isolation, but did not further decrease upon washing. Hence, the protocol enriched nuclei, but there was still contamination in the fraction. The

results also suggested the crude nuclear fraction was not significantly more contaminated than the purified fraction.

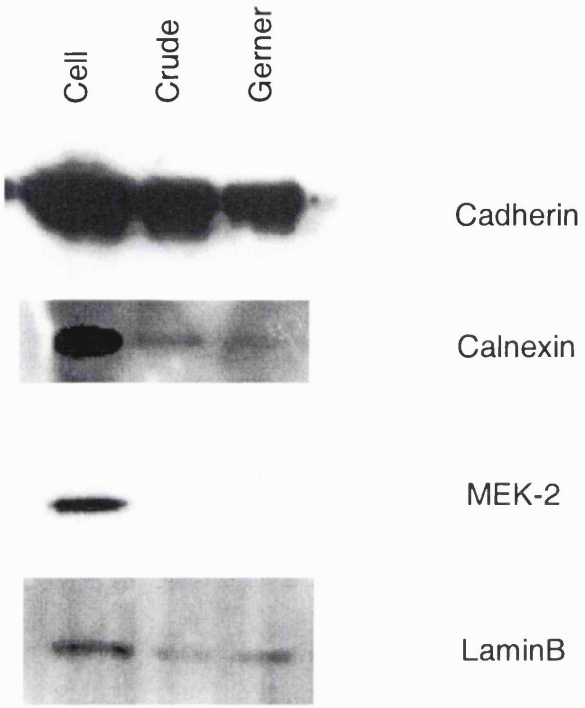


Figure 6.3: Comparative enrichment of nuclear proteins through crude and clean nuclear fractionation steps.

Fractions were separated by 1D SDS-PAGE and then transferred to PVDF. The membrane was then probed against E-caderin (plasma membrane), calnexin (ER), MEK-2 (cytosolic) and lamin B (nuclear). Blots were developed using ECL reagents.

The nuclear isolation protocol was also assessed by 2D polyacrylamide gel electrophoresis. Figure 6.4 shows 2D images of proteins from a whole cell lysate and nuclear fraction lysate of HN5 cells. The nuclear fraction clearly comprised a sub-section of the proteins in the whole cell lysate; many spots present in the whole cell lysate were not in the nuclear fraction, whilst several proteins were more intense in the nuclear fraction. This confirms that the process was selectively enriching certain proteins. Five random spots were excised from the nuclear gel and identified by in-gel digestion followed by peptide mass fingerprinting (Table 6.1). Of these proteins, laminin is an

extracellular glycoprotein, prohibitin is cytoplasmically localised, TAR and hnRNP A1 are nuclear proteins, and eIF-6 is nuclear and cytoplasmically localised[198]. This result suggested the nuclear fraction was partly contaminated by non-nuclear proteins. By comparing the gel images, the laminin series of spots were present at similar intensities in both images. The hnRNP A1 group of spots were marginally more intense in the nuclear fraction, whilst the prohibitin spot was more intense in the whole cell lysate. In contrast, the eIF-6 peak was not visible in the whole cell image and the TAR protein was significantly more intense in the nuclear fraction. Thus, the identified extracellular protein was not enriched in the nuclear fraction, and the intensity of the cytosolic protein has decreased. The protocol has enriched the three proteins expected in a nuclear fraction.

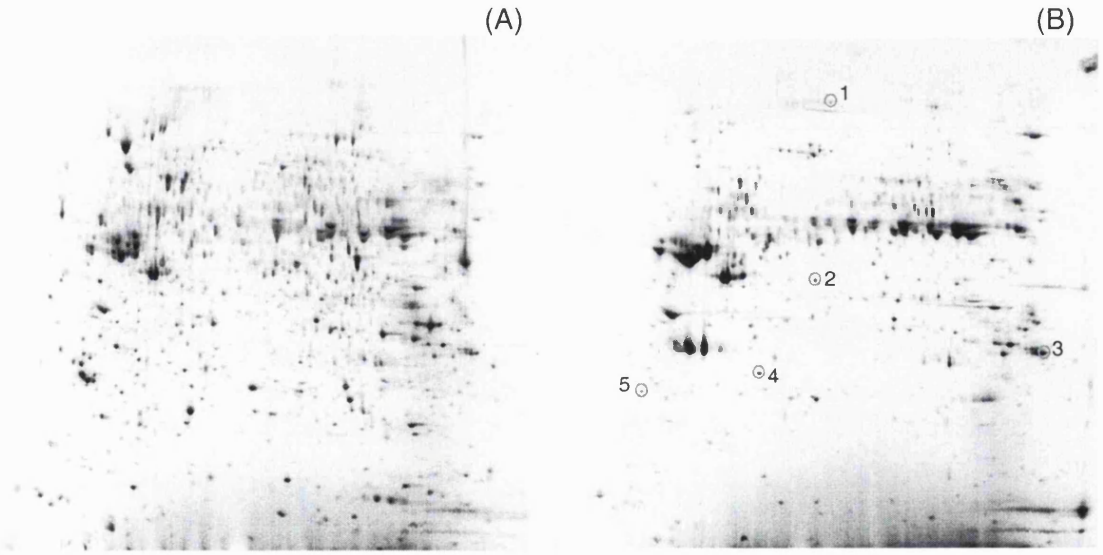


Figure 6.4: Enrichment of proteins following nuclear isolation.

A whole cell lysate (A) and a purified nuclear fraction (B) of HN5 cells were separated by 2D PAGE. Proteins were visualised using Coomassie staining. Five spots were excised from the nuclear fraction gel for subsequent identification by mass spectrometry.

Table 6.1: Protein identifications of random spots excised from a 2D gel of a nuclear fraction (Figure 6.4B).

Spot No	ID	MW / pI	Peak Match	Coverage (%)
1	Laminin (gamma 2 subunit)	130991 / 5.83	19/35	33
2	TAR DNA-binding protein-43	44740 / 5.85	15/49	48
3	TIS / HnRNP 1A	34196 / 9.27	19/83	73
4	Prohibitin	29804 / 5.57	16/53	71
5	eIF-6	26599 / 4.56	10/41	53

6.3. Identifying Radiolabelled Proteins from a Nuclear Fraction

Having isolated nuclei it was hoped to find GlcNAc-modified proteins in this fraction. Initially, a nuclear fraction of HN5 cells was [^{3}H]galactose modified *in vitro* with galactosyltransferase and run on a 1D gel to assess whether proteins were successfully labelled and could be detected. After proteins were separated by 1D PAGE the gel was soaked in EN^{3}HANCE and exposed to film for one month (Figure 6.5). There were a number of bands visible at a range of molecular weights, showing the radiolabelling was effective at labelling certain proteins.

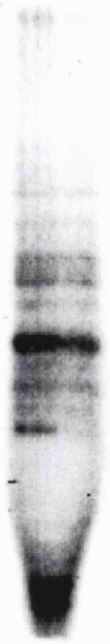


Figure 6.5: Radiolabelling of GlcNAc-modified proteins in a nuclear fraction of HN5 cells.

A nuclear fraction of HN5 cells were [^{3}H]galactose labelled in vitro, *then separated by 1D SDS-PAGE. The gel was soaked in EN^{3}HANCE, then exposed to film.*

Having established the labelling was working, a nuclear fraction was examined by 2D PAGE with the intention of identifying labelled proteins. Thus, a nuclear fraction was split in half, and one half was *in vitro* [^{3}H]galactose labelled. 'Hot' and 'cold' samples were separated by 2D PAGE, and stained using Coomassie (Figure 6.6), showing the two gel images to be very similar. The 'hot' gel was then soaked in EN^{3}HANCE, dried and exposed to film for several weeks.

Upon development a number of very faint signals were observed from labelled proteins, as illustrated in Figure 6.7A. Spots were cut out from the corresponding regions of the duplicate 'cold' gel and analysed by peptide mass fingerprinting. Figure 6.7 indicates the locations of excised spots, some of which do not coincide with locations corresponding to Coomassie stained spots. Spots 2, 4, 6, 8, 10, 11 and 12 were all cut from locations of the 'cold' gel corresponding to radiolabelled areas of the 'hot' gel. Nearby spots were also cut out to see if these matched the same protein. The results from this analysis are given in Table 6.2.

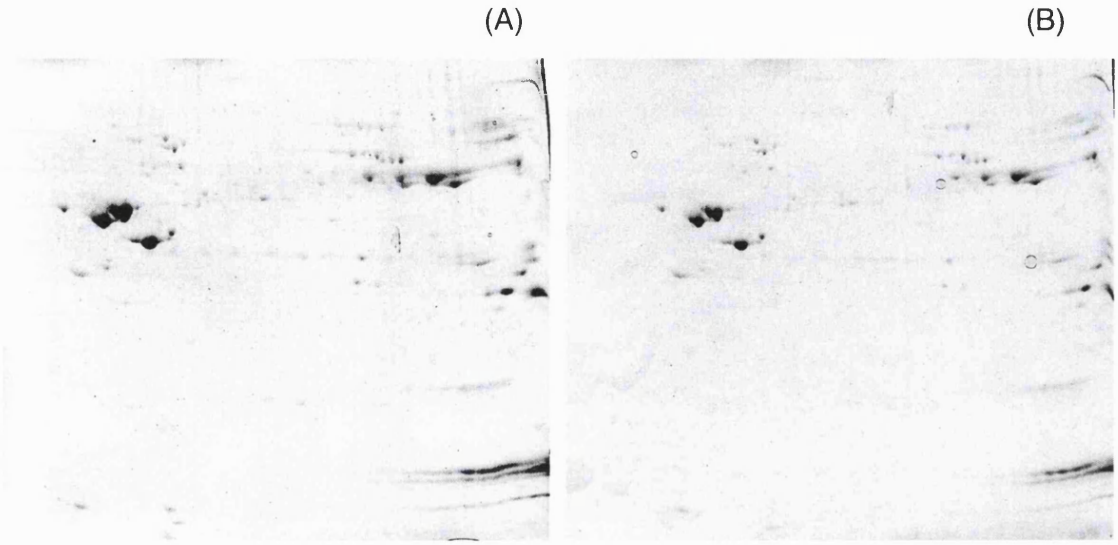


Figure 6.6: 2D-PAGE images of 'hot' and 'cold' nuclear fractions.

(A) Unlabelled and (B) [^{3}H]galactose labelled nuclear fractions were resolved by 2D-PAGE and visualised using Coomassie staining.

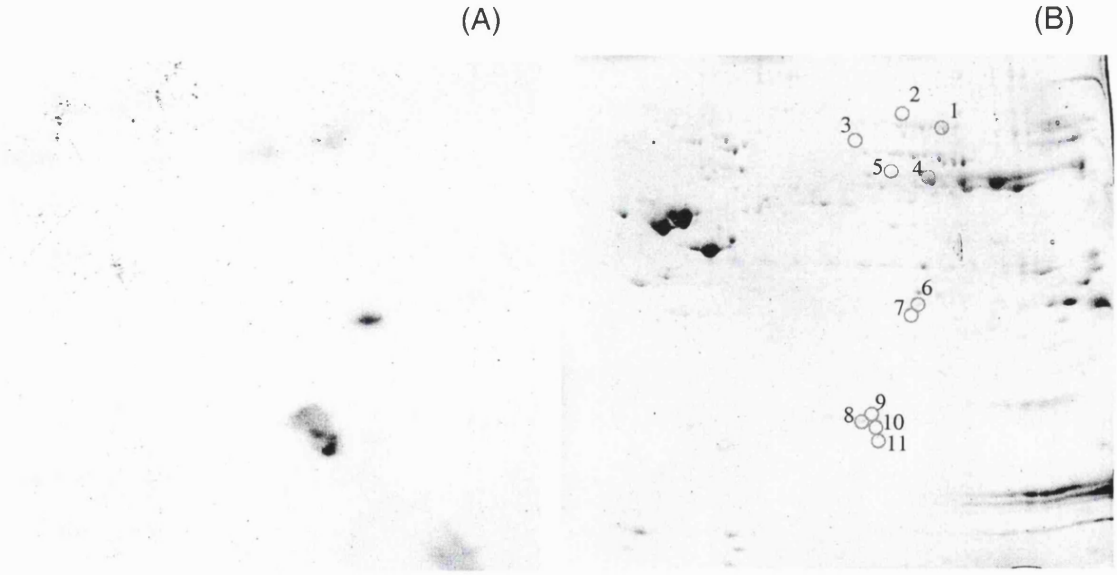


Figure 6.7: Radiolabelled spots on a 2D polyacrylamide gel of 'hot' nuclear fraction and their corresponding positions on the duplicate 'cold' gel.

(A) Exposed film of 2D gel of 'hot' nuclear fraction (Figure 6.6B). (B) Image of 'cold' 2D gel and locations of spots excised for identification by mass spectrometry. Spots 2, 3, 5, 7 and 9-11 relate to radiolabelled spots on (A).

Table 6.2: Protein identifications of spots excised from the duplicate 'cold' gel, relating to radiolabelled proteins in the 'hot' gel.

Spot	ID	MW / pI	Match	Coverage (%)
1	Lamin A	74140 / 6.57	13 / 23	26
2	No match			
3	RED protein + Keratin 6	65517 / 6.14	13/27	28
4	Keratin 5	62462 / 8.14	43 / 52	58
5	Keratin 5	62462 / 8.14	15 / 36	29
6	HnRNP H3a	36927 / 6.37	14 / 22	55
7	HnRNP H3a or b	36927 / 6.37 or 35239 / 6.36	14 / 22	50 or 48
8	Pre-mRNA splicing factor SRP20	18072 / 11.54	6 / 18	29
9	Pre-mRNA splicing factor SRP20	18072 / 11.54	5 /21	29
10	Pre-mRNA splicing factor SRP20	18072 / 11.54	3 / 13	23
11	Keratin contaminants			

Of the 12 gel pieces analysed, positive identifications were determined for ten of these spots: a confident match for spot 2 could not be obtained, and spot 11 matched keratin 1 and other keratins, signifying contaminating protein.

Keratins are cytoplasmically localized and all the other proteins matched are nuclear proteins, suggesting none of the radiolabelled proteins were extracellular glycoproteins, the major potential source of contaminating signal.

None of the proteins matched are previously known O-GlcNAc-modified proteins. However, although keratins 5 and 6 are not known to be GlcNAc-modified, other keratins are known to bear the modification[78, 199]. The mass fingerprint from spot 3 identified two co-eluting proteins, Keratin 6 and RED. RED (or RER protein) is part of a family of nuclear proteins characterised by multiple repeats of the acidic basic residue combination R followed by E or D. It is believed to bind to chromatin and to be involved in transcription[200].

Expression of pre-mRNA splicing factor SRP20 (also known as splicing factor arginine/serine rich 3 (SFR3)) is induced by serum[201], and is thought to be highly phosphorylated on serines in the arginine/serine rich domain.

HnRNP H3 binds pre-mRNA and is involved in RNA splicing. It participates in early heat shock induced splicing arrest[202]. Another member of this family of proteins, hnRNP G, is a previously identified O-GlcNAc-modified protein[105].

All these proteins could potentially be GlcNAc-modified. They are nuclear- or cytoplasmically-localised. The keratins and SRP20 are phosphoproteins. Nothing is known of the post-translational state of RED or hnRNP H3. It is already known that O-GlcNAcylation plays a significant role in the regulation of transcription through modification of the largest subunit of RNA polymerase II and a number of its transcription factors[88, 89]. Hence, regulation of mRNA splicing factors such as SRP20 and hnRNP H3 by O-GlcNAc modification is consistent with GlcNAc playing a global role in the regulation of transcription.

It is unfortunate that the levels of protein were not sufficient to detect the GlcNAc-modified peptides in these protein digests. Thus, one cannot be certain that the identified proteins were the radiolabelled proteins. In each case, the GlcNAc-modified protein migrated to a different point on the gel to the majority of this protein. Consequently, a stained spot relating to the GlcNAc-modified protein was usually not visible. This suggests only a small sub-section of each of these proteins is O-GlcNAc-modified. It may be necessary to enrich the levels of O-GlcNAc modification using a GlcNAcase inhibitor such as PUGNAc[121] or streptozotocin[203] in order to be able to locate the modified peptides and determine their sites of modification.

6.4. Lectin Probing

Although [^{3}H]galactose labelling is specific, there are problems with it. The major drawback is that the signal is very weak and thus long exposures to film are required to detect the modification. It is also difficult to monitor the efficiency of the labelling due to its difficulty of detection. An alternative method for detecting O-GlcNAc-modified proteins is to use the lectin WGA conjugated to horseradish peroxidase (WGA-HRP) to probe proteins transferred onto membrane[88].

A comparison was made of the patterns observed by probing a whole cell lysate and a nuclear fraction of HN5 cells using WGA-HRP (Figure 6.8). The pattern of bands in the two lanes was similar, but there were some differences between the two lanes.

Figure 6.8: Probing cell and nuclear fractions using WGA-HRP.
Whole cell lysate and nuclear lysates were separated by 1D SDS-PAGE, blotted onto PVDF and probed using WGA-HRP. Images were developed using ECL reagents.

In order to assess whether the bands observed are specific and relate to GlcNAc-modified proteins, cell and nuclear fractions were on-blot digested using PNGaseF and sialidase, before WGA-HRP treatment, which would remove all N-linked sugars and sialic acid moieties, for which WGA also has an affinity. As can be seen in Figure 6.9 there were still many bands observed after glycosidase treatment. To confirm the WGA-HRP is binding to GlcNAc residues the membrane was probed using WGA-HRP in the presence of free GlcNAc. This completely destroyed any signal in both the cell and nuclear fraction.

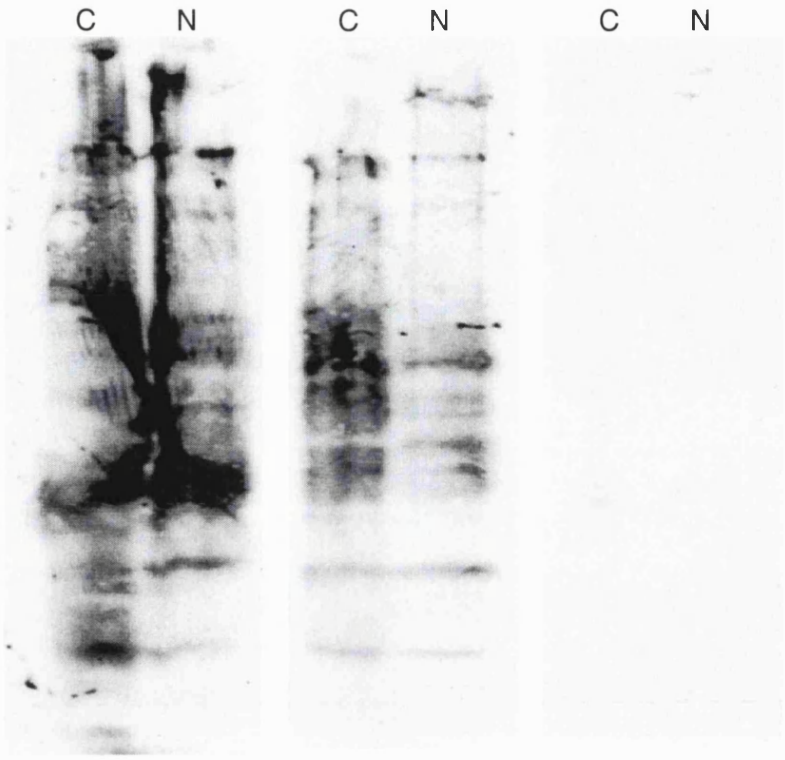


Figure 6.9: Specificity of WGA-HRP binding to GlcNAc-modified proteins.

Whole cell and nuclear fractions were separated by 1D SDS-PAGE, transferred onto PVDF. (A) Membrane was probed using WGA-HRP. (B) Membrane was incubated with PNGaseF and sialidase to remove N-linked carbohydrates and teminal sialic acid residues, and then probed using WGA-HRP. (C) Membrane was probed using WGA-HRP in the presence of 0.5M free GlcNAc. Blots were developed using ECL reagents.

6.5. WGA-Sepharose Affinity Chromatography of Proteins

The ability to affinity purify GlcNAc-modified proteins from a complex mixture such as a nuclear fraction would greatly simplify the sample to analyse. WGA-sepharose affinity chromatography was successfully used to purify the nuclear pore complex from a nuclear envelope extract[156, 168]. This protocol was applied to the HN5 nuclear extract to ascertain whether it could be used to purify GlcNAc-modified nuclear proteins.

A nuclear fraction was loaded onto a WGA-sepharose affinity column. The column was washed twice to remove non-specifically bound proteins, before two elutions of peptides employing a GlcNAc-containing buffer. The eluents after the loading, two wash steps and two elution steps were collected and each were [^{3}H]galactose labelled *in vitro*. These fractions were then separated on a 1D polyacrylamide gel, and proteins visualised by Coomassie staining (Figure 6.10A).

Nearly all the protein bound to the column, and the non-specifically bound protein did not begin to elute until the second wash step. The first elution lane contains the majority of the protein, and there are also several proteins in the second elution lane. Hence, the first elution fraction clearly contains many non-specifically bound proteins, suggesting more washing of the column was required before elution. However, the staining pattern of the elution lanes is not the same as the second wash lane, showing the chromatography has selectively enriched certain proteins.

To ascertain whether these enriched bands correspond to GlcNAc-modified proteins, the gel was soaked in fluorescent scintillant and exposed to film (Figure 6.10B). The major bands enriched in the second elution lane are radiolabelled and hence glycosylated proteins.

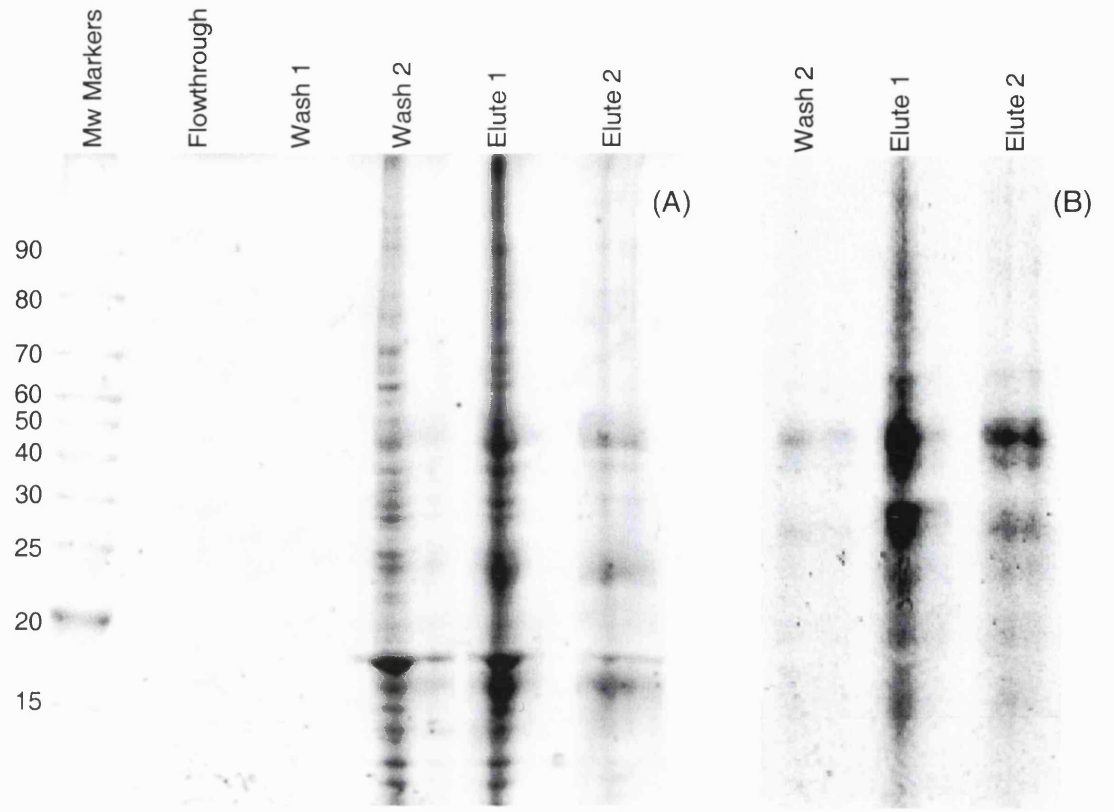


Figure 6.10: WGA-sepharose affinity purification of GlcNAc-modified proteins.
A nuclear fraction of HN5 cells was loaded onto a WGA-sepharose column and the eluent collected (flowthrough). The column was washed twice to remove non-specifically bound proteins (wash 1 and wash 2), then GlcNAc-modified proteins were specifically eluted utilising a GlcNAc containing buffer (elute 1 and elute 2). Fractions were [^{3}H]galactose labelled, then separated by 1D SDS-PAGE and visualised using Coomassie staining(A). The 1D gel was soaked in EN^{3}HANCE and exposed to film (B).

This experiment was repeated employing extra wash steps to remove more of the non-specifically bound protein prior to the elution. However, more washing removed specifically bound GlcNAc-modified proteins as well.

6.6. WGA-Sepharose Affinity Chromatography of Peptides

It was investigated whether WGA-sepharose could be used to enrich GlcNAc-modified peptides from a mixture of unmodified peptides prior to mass spectrometry. The peptide mixture used contained des-arg bradykinin, angiotensin I, neurotensin, G-CTD and G-myc. Peptides were incubated with WGA-sepharose beads in an eppendorf tube, and the supernatant was removed. Beads were then washed using various salt buffers, some containing urea or Triton X-100, before a specific elution step employing a GlcNAc-containing buffer.

A major difficulty of interfacing this affinity chromatography to mass spectrometry is that the salt buffers used for the chromatography are incompatible with subsequent mass spectrometry. Thus, relatively low salt concentration buffers were used, and the samples were subsequently analysed by LC-MS, where there was the opportunity to wash away the majority of the salts, allowing mass spectrometry data to be obtained.

A number of experiments were carried out, but no consistent results could be obtained. It was clear that the retention of GlcNAc-modified peptides on the beads was poor, as the majority of both GlcNAc-modified peptides were lost in every case. Sometimes there was a relative enrichment of one of the GlcNAc-modified peptides, but generally not both G-CTD and G-myc peptides in the same experiment. Samples were still very salty, even after washing on a C18 guard column prior to separation by HPLC, and the major component of each sample turned out to be the glycoprotein wheat germ agglutinin itself at m/z 17089 (Figure 6.11).

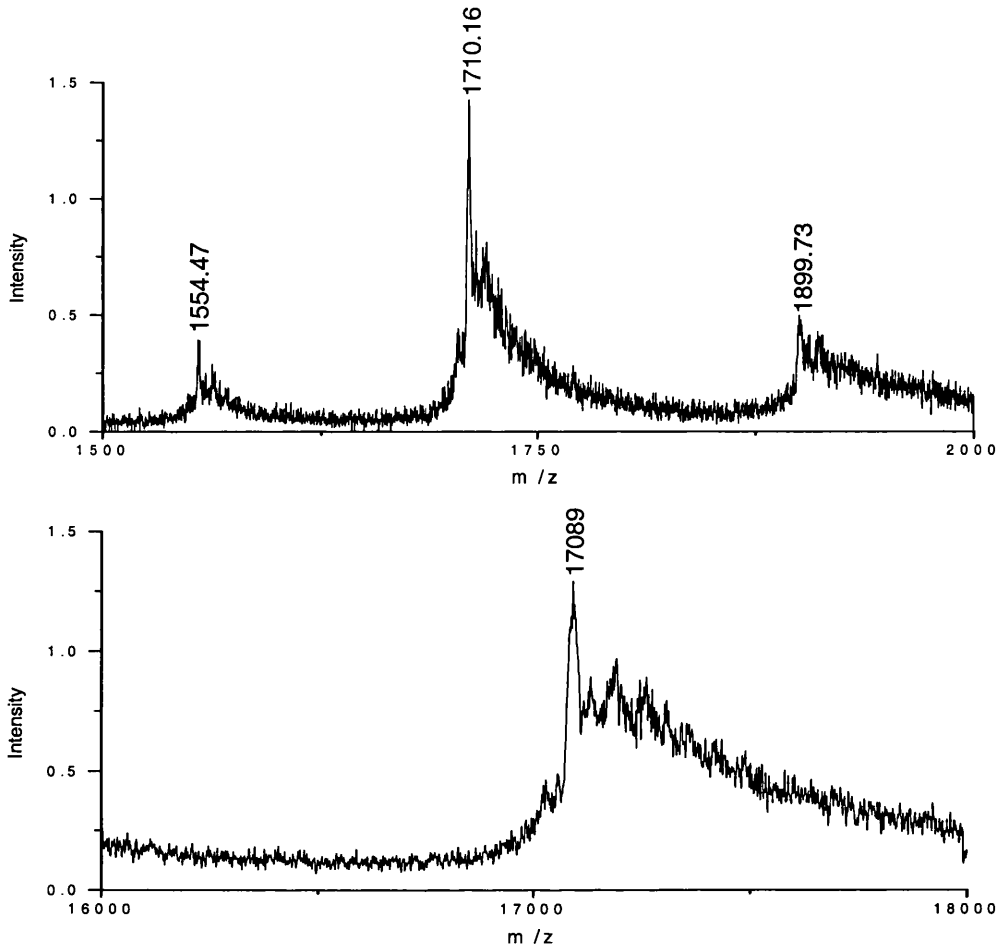


Figure 6.11: Mass spectrum demonstrating WGA is eluting off WGA-sepharose beads.

(A) ESI mass spectrum shows an envelope of multiply charged peaks. Deconvolution of this spectrum shows the peaks relate to a peak of mass 17089 (B). The tailing towards higher mass is due to the different glycoforms of the WGA protein.

6.7. Discussion

Nuclear isolation protocols were assessed for their ability to isolate a pure nuclear fraction from HN5 cells. Fractions were analysed by western blotting and mass spectrometric identification of 2D-PAGE-separated proteins in the nuclear fraction. The chosen protocol produced a fraction containing some contamination with proteins from

other parts of the cell, but all enriched proteins identified by mass spectrometry were nuclear and there appeared to be less loss of nuclear proteins using this method than the alternative protocol it was compared with.

Having selected the most effective protocol, GlcNAc-modified proteins in this nuclear fraction were [^{3}H]galactose-labelled and separated by 2D-PAGE. Proteins from the locations of radiolabelled spots were excised on a duplicate cold gel and identified by peptide mass fingerprinting. None of the identified proteins were previously known to be GlcNAc-modified, but all proteins were nuclear or cytoplasmic, and some belonged to families of proteins of which other member/s are known to be modified. The radiolabelled spots of GlcNAc-modified proteins generally did not co-elute with the majority of the protein, suggesting only a small proportion of the identified protein was GlcNAc-modified. This meant GlcNAc-modified protein levels were low, and no GlcNAc-modified peptides were identified in the digests of these spots.

WGA-HRP was employed to probe membrane blots of nuclear fractions. The WGA does appear to be binding to GlcNAc residues. This method appears to be a valid alternative to the radiolabelling protocol. The advantage of this visualisation method is that it is faster; the images presented here were acquired after only 30 seconds exposure to film. Conversely, the image can suffer from high backgrounds, and does not always produce clear tight bands. It is probably less specific than [^{3}H]galactose labelling. It is also not particularly amenable to subsequent mass spectrometric detection, as the membrane has to be blocked with excess protein before probed using WGA-HRP to prevent non-specific binding to the membrane. Thus, it would be impossible to identify proteins from this membrane blot. Hence, a duplicate version of the sample would have to be run and used for identifying proteins that bind the WGA *i.e.* a similar approach as was used for identifying labelled proteins from the radiolabelling experiment (Section 6.3).

Attempts were made to affinity purify proteins and peptides using WGA affinity chromatography. It was shown that the affinity of interaction between the lectin column and GlcNAc-modified proteins was not high. The affinity column retarded the elution of proteins rather than retaining them. Hence, increasing the number of washing steps caused loss of GlcNAc-modified proteins. This is not a surprise as the affinity of

interaction between WGA and a single GlcNAc moiety is known to be low. The protocol used here was designed for purifying a highly modified protein that would have a higher affinity for the column than most modified proteins[168]. For singly modified proteins milder washing conditions may be required in order to permit retention of these proteins, and other laboratories have used WGA affinity chromatography to isolate GlcNAc-modified proteins[127, 155].

Experiments to affinity enrich GlcNAc-modified peptides using WGA-sepharose beads were inconclusive. The results suggested that the affinity of interaction between the lectin and the singly-modified peptides was too weak to obtain retention of GlcNAc-modified peptides. However, the results do not discount the possibility of obtaining retardation of elution of GlcNAc-modified peptides if they were run through a WGA column in a similar way as observed using RCA I columns for GlcNAc-Gal modified peptides[169, 171]. The use of a column was not investigated in this work, as it would increase sample volumes dramatically, and would make the method less sensitive. It was also disturbing that a significant amount of WGA was being lost off the beads. If the WGA is washing off the beads, then even if there is binding between WGA and GlcNAc-modified peptides, this will not necessarily mean retention of peptides. In conclusion, in these experiments WGA-sepharose chromatography was not a sensitive or reliable affinity purification protocol for enriching GlcNAc-modified peptides. The only demonstrated protocol for affinity enrichment of GlcNAc-modified peptides is to galactose label then purify using RCA I [169, 171].

7. CONCLUSIONS FROM THIS WORK

There are three stages to the analysis of GlcNAc modification. First modified proteins must be identified and purified. The next step is to locate GlcNAc-modified peptides in the digest of the modified protein. Finally, the site of modification of the glycosylated peptide must be determined.

The use of affinity chromatography to enrich a sample for O-GlcNAc-modified proteins and peptides will be a vital step if sites are modified at low stoichiometry. Affinity chromatography using WGA to isolate GlcNAc-modified proteins has previously been reported [155, 168], and the protocol was repeated in this thesis. However, experiments attempting to enrich for GlcNAc-modified peptides from a protein digest using WGA were unsuccessful in this work. Nevertheless, enrichment of GlcNAc-Gal modified peptides using the lectin RCA I has been demonstrated by others [169, 171]. The affinity of interaction between singly glycosylated peptides and both these lectins is known to be low, as RCA I affinity columns fail to retain singly modified peptides; they merely retard their elution [169], and WGA has been shown to contain four binding sites which act in a cooperative manner [195]. It is likely that WGA may be employed in an affinity column to enrich GlcNAc-modified peptides, but more method development will be required.

Probably the most specific affinity tool for enriching samples for GlcNAc modification is the recently produced monoclonal antibody CTD110.6, which is reported to bind only to GlcNAc-modified serine and threonine residues [155]. If so, this would be more specific than WGA. Thus it could be used for western blotting and for the immunoprecipitation of GlcNAc-modified proteins.

Detection of GlcNAc-modified peptides in an enzymatic digest of a modified protein can be a major challenge. If the site of modification is known, then one can predict the mass of the GlcNAc-modified peptide and search for a specific peak in a MALDI mass spectrum. Instead, LC-MS analysis of the digest could be performed, and then an extracted ion chromatogram of the predicted GlcNAc-modified peptide m/z could be carried out. Alternatively, one could examine the region of the LC-MS run immediately prior to the elution time of the unmodified version of the potentially GlcNAcylated

peptide, since the GlcNAc-modified analogue is more hydrophilic (multiple illustrations of this in Chapter 5).

If the site of modification is not known, the task is much more difficult. When the stoichiometry of modification is relatively high, then an LC-MS analysis using automatic selection of peaks for fragmentation analysis can be performed. All the fragmentation spectra can then be interrogated for the formation of the GlcNAc oxonium ion (Chapter 5). However, this is not a reliable exhaustive technique for finding modified peptides, as if sites are modified at low stoichiometry, the GlcNAc-modified peptides may not be automatically selected for fragmentation analysis. Thus, these GlcNAc-modified peptides would not be located using this screening method. Two of the previously published O-GlcNAc modification sites in SRF [92] were not detected during the work in this thesis, and the probable explanation for this is low stoichiometry of modification at these sites.

A more comprehensive mass spectrometric approach to finding GlcNAc-modified peptides is through the use of precursor ion scanning. GlcNAc-modified peptides produce the GlcNAc oxonium ion at m/z 204 in ESI-CID-MS-MS. Thus, a precursor ion scan for fragments formed at this mass should be able to locate modified peptides, and could detect peptides modified at low stoichiometry. The advantage of precursor ion scanning over examining extracted ion chromatograms of automatically produced MS-MS spectra is this approach will fragment all ions, rather than just the most intense. Therefore, this approach is more likely to detect low stoichiometry modified peptides. The major limitation of this technique is the sensitivity of precursor ion scanning, which is significantly lower than conventional mass spectrometry. There is also the complication that non-glycosylated peptides can also produce a fragment ion of mass 204 Da. Most notably, peptides whose C-terminal residues are GK, a relatively common occurrence in tryptic peptides, will produce a y_2 ion of theoretical mass 204.135, compared with the GlcNAc oxonium ion's theoretical mass of 204.086. A triple quadrupole mass spectrometer is not able to distinguish between these ions in precursor ion scanning experiments. However, precursor ion scanning on a Q-TOF instrument has the resolving power to distinguish between ions that differ by this small amount, and this resolution was demonstrated using the QSTAR in Chapter 4. Precursor ion scanning on a

Q-TOF instrument is the most specific mass spectrometric method for locating GlcNAc-modified peptides in a complex mixture. However, it is less sensitive than other mass spectrometric approaches outlined here, including precursor ion scanning on a triple quadrupole instrument, because of the lower duty cycle. Further investigative work needs to be performed to establish the sensitivity of precursor ion scanning on the QSTAR as a screening technique, but where sample amount is not a major limiting factor, this approach should be the most thorough at detecting all GlcNAc-modified peptides in a digest.

An alternative precursor ion scanning experiment has already been used to locate GlcNAc-modified peptides. GlcNAc residues were enzymatically modified *in vitro* with galactose, then precursor ion scanning was carried out for the GlcNAc-Gal disaccharide oxonium ion at m/z 366 during an LC-MS analysis [169]. This technique was shown to be able to locate a modified peptide in a tryptic digest spiked with a GlcNAc-modified synthetic peptide. However, it was unable to find the modified peptide in αA-crystallin without the use of RCA I affinity chromatography to enrich for GlcNAc-modified peptides.

The classical radioactivity-based method for finding GlcNAc-modified peptides by screening HPLC fractions for [^{3}H]galactose-labelled modified peptides is a sensitive method for finding fractions containing GlcNAc-modified peptides. The use of Edman sequencing to identify the peptide requires the fraction to contain only one peptide, so necessitates further rounds of HPLC purification, making this protocol less sensitive. Using mass spectrometry to identify the labelled peptide in this fraction is a better protocol, and this combined radioactivity and mass spectrometric approach is probably as sensitive and comprehensive as precursor ion scanning.

A significant problem with the study of O-GlcNAc modification of proteins is there is often low stoichiometry of O-GlcNAc modification at any particular site. Hence, even for highly modified proteins, several picomoles of starting material are required. Analysis of GlcNAc-modified proteins from a gel band has been performed in this thesis, and the evidence suggests there is no noticeable loss of the GlcNAc modification through the in-gel digestion protocol (Chapter 4). This is an important result, as gel electrophoresis is a convenient method for purifying GlcNAc-modified proteins.

Work in this thesis has established that the ability to identify sites of O-GlcNAc modification using mass spectrometry is peptide-sequence specific. Due to the high lability of the glycosidic linkage the absence of a glycosylated fragment is not a reliable factor in determining a GlcNAc modification site. In most cases, it is necessary to obtain overlapping N-terminal and C-terminal fragment ion series to identify a site of modification. Hence, enzymes that produce peptides with basic C-termini such as trypsin and Lys-C are often not the best choices of enzyme for determining sites of modification. The occasions when overlapping ion series may not be required are when glycosylated internal fragment ions are produced. GlcNAcylation sites often have a proline residue a few residues N-terminal to the modification, and cleavage N-terminal to proline residues is favoured in CID fragmentation [170]. Thus, if a glycosylated internal ion is formed due to cleavage at this proline, and it contains only one possible modified residue, then this will identify the site of GlcNAc modification (Chapter 4) [107].

The sensitivity of the techniques presented in this thesis is significantly higher than previously published protocols for identifying sites of modification. β-Elimination of the GlcNAc moiety, followed by ESI-CID-MS-MS can identify a site of modification when starting with one pmole of a synthetic peptide [158]. In this thesis, using CID-MS-MS on a Q-TOF instrument, a site of modification of a synthetic GlcNAc-modified peptide was determined by nanospray-MS-MS when only 100 fmoles of sample was loaded, and sample consumption was less than 20 fmoles (Chapter 3).

In the last couple of years the McLafferty group has reported a new ion fragmentation chemistry called electron capture dissociation (ECD) [204, 205], which cleaves peptides to form 'c' and 'z' type ions. Unlike PSD and CID, ECD is a non-ergodic process and fragments the peptide backbone without cleaving the more fragile bonds attaching post-translational modifications such as phosphate groups [206] and O-linked sugars [207]. Hence, ECD spectra of O-glycosylated peptides contain few or no deglycosylated fragments, and sites of modification can be determined with much greater ease. Unfortunately, this fragmentation mechanism has thus far only been employed on FT-ICR instruments, and is not presently commercially available. However, in the future, the use of this fragmentation mechanism could become the method of choice for studying

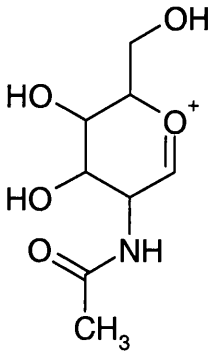
post-translational modifications, especially if it can be applied on less expensive instruments.

The work in this thesis has demonstrated that using quadrupole-oa-TOF-MS instruments it is possible to locate and identify O-GlcNAc modification sites at much higher sensitivity than the classical method combining radioactivity and Edman sequencing [156]. CID-MS-MS spectra acquired using QTOF instruments are also a more sensitive approach than previously demonstrated approaches combining β-elimination with subsequent mass spectrometric fragmentation analysis [157, 158].

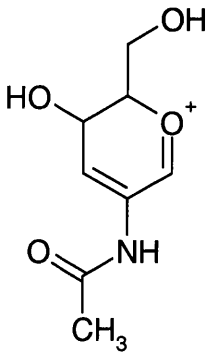
The improved isolation and detection of GlcNAc-modified proteins using the CTD110.6 antibody[155], combined with improved mass spectrometric techniques to detect GlcNAc-modified peptides[169] and determine sites of modification, as demonstrated in this thesis, should facilitate more detailed studies of this modification, and provide information to elucidate the function of this post-translational modification on biological activity.

Appendix 1

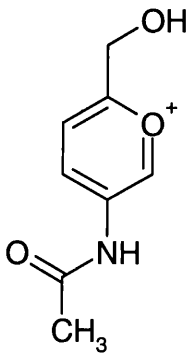
GlcNAc Fragment Ions



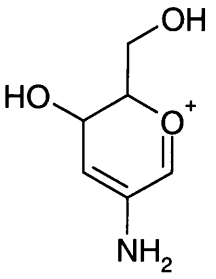
m/z 204



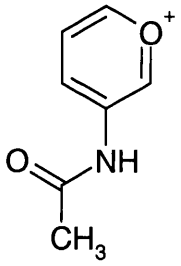
m/z 186



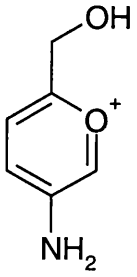
m/z 168



m/z 144



m/z 138



m/z 126

Appendix 2

αA-Crystallin

1

MDIAIQHPWF KRTLGPFYPS RLFDQFFGEG LFEYDLLPFL SSTISPYYRQ

51

SLFRTVLDSG ISEVRSDRDK FVIFLDVKHF SPEDLTVKVQ EDFVEIHGKH

101

NERQDDHGYI SREFHRRYRL PSNVDQSALS CSLSADGMLT FSGPKIPSGV

151

DAGHSERAIP VS[1]REEKPSSA PSS

[1]GlcNAc modification[74, 107]

Appendix 3

Human Serum Response Factor

-12
ME QKLISEEDLN[1]
1
MLPTQAGAAA ALGRGSALGG SLNRTPTGRP GGGGGTRGAN GGRVPGNGAG
51
LGPGRLEREA AAAAATTPAP TAGALYSGSE GDS[2]ESGEEEE LGAERRGLKR
101
SLS[3]EMEIGMV VGGPEASAAA TGGYGPVSGA VSGAKPGKKT RGRVKIKMEF
151
IDNKLRRYTT FSKRKTGIMK KAYELSTLTG TQVLLLVASE TGHVYTFATR
201
KLQPMITSET GKALIQTCLN SPDS[4]PPRSDP TTDQRMSATG FEETDLTYQV
251
SESDSSGETK DTLKPAFTVT NLPGTTSTIQ TAPSTS[5]TTMQ VSSGPSFPIT
301
NYLAPVS[5]AS[5]V SPS[6]AVS[5]SANG TVLKSTGSGP VSSGGLMQLP TSFTLMPGGA
351
VAQQVPVQAI QVHQAPQQAS PSRDSSTDLT QTS[5]SSGTVTL PATIMTSSVP
401
T[7]T[7]VGGHMMYP SPHAVMYAPT SGLGDGSLTV LNAFS[8]QAPSTMQVSHSQVQE
451
PGGVPQVFLT ASSGTVQIPV SAVQLHQMAV IGQQAGSSSN LTELQVVNLD
501
TAHSTKSE

[1]Myc tag

[2]Phosphorylation by CKII[188]

[3]Phosphorylation: by pp90^{rsk}(MAPKAP-K1)[185]

[4]Novel phosphorylation site

[5]GlcNAcylation Sites:[92] One of 307 or 309 is modified.

by CaM Kinases II and IV[187]

in response to stress by MAPKAP-K2:[186]

[6]Novel GlcNAcylation site

[7]Novel GlcNAcylation site: One of 401 or 403 is modified. Evidence favours 401.

[8]Phosphorylation by DNA-PK[189]

8. REFERENCES

1. Knox, B.E. and Vastola, F.J., Laser Focus, 1967. **3**: p. 15.
2. Tanaka, K., Waki, H., Ido, Y., Yoshida, Y., and Yoshda, T., Rapid Comm. in Mass Spec., 1988. **2**: p. 151-153.
3. Karas, M. and Hillenkamp, F., *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons.* Anal Chem, 1988. **60**(20): p. 2299-301.
4. Morris, H.R., *Soft Ionisation Biological Mass Spectrometry.* 1981, London: Hayden and Sons.
5. Talroze, V.L., Jacob, R.J., Burlingame, A.L., and Baldwin, M.A., *Insight into the MALDI Mechanism: Matrix Decomposition and Pneumatic Assistance in Plume Formation*, in *Advances in Mass Spectrometry*, E. Gelpi, Editor. 2001, Wiley & Sons: Chichester. p. 481-482.
6. Beavis, R. and Chait, B., *Matrix-assisted Laser Desorption Mass Spectrometry Using 355nm Radiation.* Rapid Comm. Mass Spec., 1989. **3**(12): p. 436-439.
7. Strupat, K., Karas, M., Hillenkamp, F., Eckerskorn, C., and Lottspeich, F., *Matrix-Assisted Laser Desorption Ionization Mass Spectrometry of Proteins Electroblotted after Polyacrylamide Gel Electrophoresis.* Anal. Chem., 1994. **66**(4): p. 464-470.
8. Schmelzeisen-Redeker, L., Bütfering, L., and Röllgen, F.W., Int. J. Mass Spectrom. Ion Phys., 1989. **50**: p. 139.
9. Iribarne, J.V. and Thomson, B.A., J. Chem. Phys., 1976. **64**: p. 2287.
10. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.
11. Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M., *Femtomole sequencing of proteins from Polyacrylamide gels by nano-electrospray mass spectrometry.* Nature, 1996. **379**: p. 466-469.

12. Henion, J.D. and Maylin, G.A., *Drug analysis by direct liquid introduction micro liquid chromatography mass spectrometry.* Biomed Mass Spectrom, 1980. **7**(3): p. 115-21.
13. Thibault, P., Paris, C., and Pleasance, S., *Analysis of peptides and proteins by capillary electrophoresis/mass spectrometry using acidic buffers and coated capillaries.* Rapid Commun Mass Spectrom, 1991. **5**(10): p. 484-90.
14. Cameron, A.E. and Eggers, D.F.J., Rev. Sci. Instr., 1948. **19**: p. 605.
15. Vestal, M.L., Krause, M., Wahrhaftig, A.L., and Johnston, W.H. in *Proceedings of the Eleventh Annual Conference on Mass Spectrometry and Allied Topics.* 1963. San Francisco.
16. Cotter, R., *Time-of-Flight Mass Spectrometry: Instrumentation and Applications in Biological Research.* 1997, Washington, DC: American Chemical Society.
17. Wiley, W.C. and McClaren, I.H., Rev. Sci. Instr., 1955. **26**: p. 1150-1157.
18. Mamyrin, B.A., Karataev, V.I., Shmikk, D.V., and Zagulin, V.A., Sov. Phys. JETP, 1973. **37**: p. 45.
19. Spengler, B., Kirsch, D., Kaufmann, R., and Jaeger, E., *Peptide sequencing by matrix-assisted laser-desorption mass spectrometry.* Rapid Commun Mass Spectrom, 1992. **6**(2): p. 105-8.
20. Verentchikov, A.N., Ens, W., and Standing, K.G., *Reflecting time-of-flight mass spectrometer with an electrospray ion source and orthogonal extraction.* Anal Chem, 1994. **66**(1): p. 126-33.
21. Morris, H.R., Paxton, T., Panico, M., McDowell, R., and Dell, A., *A novel geometry mass spectrometer, the Q-TOF, for low- femtomole/attomole-range biopolymer sequencing.* J Protein Chem, 1997. **16**(5): p. 469-79.
22. Whitehouse, C.M., Gulcicek, E., Andrien, B., Banks, F., and Mancini, R. in *Proc. 46th ASMS Conf. Mass Spectrom. Allied Top.* 1998. Orlando, FL.
23. Thomson, B.A., Lock, C.M., and Chernushevich, I.V. *Enhanced product ion sensitivity in a qqtof for protein identification.* in *Proc. 49th ASMS Conf. Mass Spectrom. Allied Top.* 2001. Chicago.
24. Loboda, A.V., Krutchinsky, A.N., Bromirski, M., Ens, W., and Standing, K.G., *A tandem quadrupole/time-of-flight mass spectrometer with a matrix- assisted laser*

desorption/ionization source: design and performance. Rapid Commun Mass Spectrom, 2000. **14**(12): p. 1047-57.

25. Baldwin, M.A., Medzihradszky, K.F., Lock, C.M., Fisher, B., Settineri, T.A., and Burlingame, A.L., *Matrix-assisted laser desorption/ionization coupled with quadrupole/orthogonal acceleration time-of-flight mass spectrometry for protein discovery, identification, and structural analysis.* Anal Chem, 2001. **73**(8): p. 1707-20.
26. Wasinger, V.C., Cordwell, S.J., Cerpa-Poljak, A., Yan, J.X., Gooley, A.A., Wilkins, M.R., Duncan, M.W., Harris, R., Williams, K.L., and Humphery-Smith, I., *Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium.* Electrophoresis, 1995. **16**(7): p. 1090-4.
27. Godovac-Zimmermann, J. and Brown, L.R., *Perspectives for mass spectrometry and functional proteomics.* Mass Spectrom Rev, 2001. **20**(1): p. 1-57.
28. O'Farrell, P.H., *High resolution two-dimensional electrophoresis of proteins.* J Biol Chem, 1975. **250**(10): p. 4007-21.
29. Klose, J., *Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals.* Humangenetik, 1975. **26**(3): p. 231-43.
30. Gauss, C., Kalkum, M., Lowe, M., Lehrach, H., and Klose, J., *Analysis of the mouse proteome. (I) Brain proteins: separation by two- dimensional electrophoresis and identification by mass spectrometry and genetic variation.* Electrophoresis, 1999. **20**(3): p. 575-600.
31. Unlu, M., Morgan, M.E., and Minden, J.S., *Difference gel electrophoresis: a single gel method for detecting changes in protein extracts.* Electrophoresis, 1997. **18**(11): p. 2071-7.
32. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M., *Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels.* Anal Chem, 1996. **68**(5): p. 850-8.
33. Gharahdaghi, F., Weinberg, C., Meagher, D., Imai, B., and Mische, S., *Mass spectrometric identification of proteins from silver-stained polyacrylamide gel: A*

method for the removal of silver ions to enhance sensitivity. Electrophoresis, 1999. **20**: p. 601-605.

34. Kristensen, D.B., Imamura, K., Miyamoto, Y., and Yoshizato, K., *Mass spectrometric approaches for the characterization of proteins on a hybrid quadrupole time-of-flight (Q-TOF) mass spectrometer.* Electrophoresis, 2000. **21**(2): p. 430-9.
35. Sumner, L., White, S., Wolf-Sumner, B., and Asirvatham, V. *Silver Stain Removal Using H2O2 for Enhanced Peptide Mass Mapping by MALDI-TOF-MS.* in *Proc. 49th ASMS Conf. Mass Spectrom. Allied Top.* 2001. Chicago.
36. Rabilloud, T., Strub, J.-M., Luche, S., van Dorsselaer, A., and Lunardi, J., *A comparison between Sypro Ruby and ruthenium II tris (bathophenanthroline disulfonate) as fluorescent stains for protein detection in gels.* Proteomics, 2001. **1**(5): p. 699-704.
37. O'Farrell, P.Z., Goodman, H.M., and O'Farrell, P.H., *High resolution two-dimensional electrophoresis of basic as well as acidic proteins.* Cell, 1977. **12**(4): p. 1133-41.
38. Santoni, V., Molloy, M., and Rabilloud, T., *Membrane proteins and proteomics: un amour impossible?* Electrophoresis, 2000. **21**(6): p. 1054-70.
39. Nilsson, C.L. and Davidsson, P., *New separation tools for comprehensive studies of protein expression by mass spectrometry.* Mass Spectrom Rev, 2000. **19**(6): p. 390-7.
40. Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., and Yates, J.R., 3rd, *Direct analysis of protein complexes using mass spectrometry.* Nat Biotechnol, 1999. **17**(7): p. 676-82.
41. Washburn, M.P., Wolters, D., and Yates, J.R., 3rd, *Large-scale analysis of the yeast proteome by multidimensional protein identification technology.* Nat Biotechnol, 2001. **19**(3): p. 242-7.
42. Chong, B.E., Hamler, R.L., Lubman, D.M., Ethier, S.P., Rosenspire, A.J., and Miller, F.R., *Differential screening and mass mapping of proteins from premalignant and cancer cell lines using nonporous reversed-phase HPLC coupled with mass spectrometric analysis.* Anal Chem, 2001. **73**(6): p. 1219-27.

43. Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nat Biotechnol, 1999. **17**(10): p. 994-9.

44. http://www.prospector.ucsf.edu.

45. http://www.mann.embl-heidelberg.de/GroupPages/PageLink/peptidesearchpage.html.

46. http://fields.scripps.edu/sequest/.

47. http://www.matrixscience.com.

48. Huang, L., Jacob, R.J., Pegg, S.C., Baldwin, M.A., Wang, C.C., Burlingame, A.L., and Babbitt, P.C., *Functional assignment of the 20 S proteasome from Trypanosoma brucei using mass spectrometry and new bioinformatics approaches.* J Biol Chem, 2001. **276**(30): p. 28327-39.

49. Krishna, R. and Wold, F., *Post-translational modifications of proteins*, in *Methods in Protein Sequence Analysis*, I. Imahori and F. Sakiyama, Editors. 1993, Plenum Press: New York. p. 167-172.

50. Brown, J.M. and Firtel, R.A., *Phosphorelay signalling: new tricks for an ancient pathway.* Curr Biol, 1998. **8**(18): p. R662-5.

51. Sanders, D.A., Gillece-Castro, B.L., Stock, A.M., Burlingame, A.L., and Koshland, D.E., Jr., *Identification of the site of phosphorylation of the chemotaxis response regulator protein, CheY.* J Biol Chem, 1989. **264**(36): p. 21770-8.

52. Carr, S.A., Huddleston, M.J., and Annan, R.S., *Selective detection and sequencing of phosphopeptides at the femtomole level by mass spectrometry.* Anal Biochem, 1996. **239**(2): p. 180-92.

53. Janek, K., Wenschuh, H., Bienert, M., and Krause, E., *Phosphopeptide analysis by positive and negative ion matrix-assisted laser desorption/ionization mass spectrometry.* Rapid Commun Mass Spectrom, 2001. **15**(17): p. 1593-9.

54. Zhang, X., Herring, C., Romano, P., Szczepanowska, J., Brzeska, H., Hinnebusch, A., and Qin, J., *Identification of phosphorylation sites in proteins separated by polyacrylamide gel electrophoresis.* Anal. Chem., 1998. **70**: p. 2050-2059.

55. Xhou, W., Merrick, B.A., Khaledi, M.G., and Tomer, K.B., *Detection and sequencing of phosphopeptides affinity bound to immobilized metal ion beads by*

matrix-assisted laser desorption/ionization mass spectrometry. J Am Soc Mass Spectrom, 2000. **11**(4): p. 273-82.

56. Zhou, H., Watts, J.D., and Aebersold, R., *A systematic approach to the analysis of protein phosphorylation.* Nat Biotechnol, 2001. **19**(4): p. 375-8.

57. Goshe, M.B., Conrads, T.P., Panisko, E.A., Angell, N.H., Veenstra, T.D., and Smith, R.D., *Phosphoprotein isotope-coded affinity tag approach for isolating and quantitating phosphopeptides in proteome-wide analyses.* Anal Chem, 2001. **73**(11): p. 2578-86.

58. Stimson, E., Hope, J., Chong, A., and Burlingame, A.L., *Site-specific characterization of the N-linked glycans of murine prion protein by high-performance liquid chromatography/electrospray mass spectrometry and exoglycosidase digestions.* Biochemistry, 1999. **38**(15): p. 4885-95.

59. Hart, G., Cole, R., Kreppel, L., Arnold, S., Comer, F., Iyer, S., Cheng, X., Carroll, J., and Parker, G., *Glycosylation of proteins – A major challenge in mass spectrometry and proteomics*, in *Mass Spectrometry in Biology and Medicine*, A. Burlingame, S. Carr, and M. Baldwin, Editors. 2000, Humana Press: Totowa, NJ. p. 365-381.

60. Dell, A. and Morris, H.R., *Glycoprotein structure determination by mass spectrometry.* Science, 2001. **291**(5512): p. 2351-6.

61. Medzihradszky, K.F., Gillece-Castro, B.L., Townsend, R.R., Burlingame, A.L., and Hardy, M.R., *Structural Elucidation of O-Linked Glycopeptides by High Energy Collision-Induced Dissociation.* J Am Soc Mass Spectrom, 1996. **7**: p. 319-28.

62. Medzihradszky, K.F., Gillece-Castro, B.L., Settineri, C.A., Townsend, R.R., Masiarz, F.R., and Burlingame, A.L., *Structure determination of O-linked glycopeptides by tandem mass spectrometry.* Biomed Environ Mass Spectrom, 1990. **19**(12): p. 777-81.

63. Alving, K., Körner, R., Paulsen, H., and Peter-Katalinic, J., *Nanospray-ESI low-energy CID and MALDI post-source decay for determination of O-glycosylation sites in MUC4 peptides.* J. Mass Spec., 1998. **33**(11): p. 1124-33.

64. Alving, K., Paulsen, H., and Peter-Katalinic, J., *Characterization of O-Glycosylation sites in MUC2 glycopeptides by nanoelectrospray QTOF mass spectrometry.* J. Mass Spectrom., 1999. **34**: p. 395-407.

65. Hofsteenge, J., Huwiler, K.G., Macek, B., Hess, D., Lawler, J., Mosher, D.F., and Peter-Katalinic, J., *C-mannosylation and O-fucosylation of the thrombospondin type 1 module.* J Biol Chem, 2001. **276**(9): p. 6485-98.

66. Huddleston, M.J., Bean, M.F., and Carr, S.A., *Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests.* Anal Chem, 1993. **65**(7): p. 877-84.

67. Torres, C. and Hart, G., *Topography and polypeptide distribution of terminal N-Acetylglucosamine residues on the surface of intact lymphocytes.* J. Biol. Chem., 1984. **259**(5): p. 3308-3317.

68. Goletz, S., Hanisch, F.G., and Karsten, U., *Novel alphaGalNAc containing glycans on cytokeratins are recognized invitro by galectins with type II carbohydrate recognition domains.* J Cell Sci, 1997. **110**(Pt 14): p. 1585-96.

69. Hart, G.W., *Dynamic O-linked glycosylation of nuclear and cytoskeletal proteins.* Annu Rev Biochem, 1997. **66**: p. 315-35.

70. Kearse, K. and Hart, G., *Lymphocyte activation induces rapid changes in nuclear and cytoplasmic glycoproteins.* Proc. Natl. Acad. Sci. USA, 1991. **88**: p. 1701-1705.

71. http://www.cbs.dtu.dk/services/YinOYang/.

72. Cheng, X., Cole, R.N., Zaia, J., and Hart, G.W., *Alternative O-Glycosylation/O-phosphorylation of the murine estrogen receptor beta.* Biochemistry, 2000. **39**(38): p. 11609-20.

73. Jiang, M. and Hart, G., *A Subpopulation of Estrogen Receptors are Modified by O-linked N-Acetylglucosamine.* J. Biol. Chem., 1997. **272**(4): p. 2421-2428.

74. Roquemore, E.P., Dell, A., Morris, H.R., Panico, M., Reason, A.J., Savoy, L.A., Wistow, G.J., Zigler, J.S., Jr., Earles, B.J., and Hart, G.W., *Vertebrate lens alpha-crystallins are modified by O-linked N- acetylglucosamine.* J Biol Chem, 1992. **267**(1): p. 555-63.

75. Roquemore, E.P., Chevrier, M.R., Cotter, R.J., and Hart, G.W., *Dynamic O-GlcNAcylation of the small heat shock protein alpha B- crystallin.* Biochemistry, 1996. **35**(11): p. 3578-86.

76. Griffith, L.S., Mathes, M., and Schmitz, B., *Beta-amyloid precursor protein is modified with O-linked N- acetylglucosamine.* J Neurosci Res, 1995. **41**(2): p. 270-8.

77. Chou, T., Hart, G., and Dang, C., *c-Myc is Glycosylated at Threonine 58, a known Phosphorylation Site and a Mutational Hot Spot in Lymphomas.* J. Biol. Chem., 1995. **270**(32): p. 18961-18965.

78. King, I.A. and Hounsell, E.F., *Cytokeratin 13 contains O-glycosidically linked N-acetylglucosamine residues.* J Biol Chem, 1989. **264**(24): p. 14022-8.

79. Ku, N.O. and Omary, M.B., *Expression, glycosylation, and phosphorylation of human keratins 8 and 18 in insect cells.* Exp Cell Res, 1994. **211**(1): p. 24-35.

80. Mullis, K.G., Haltiwanger, R.S., Hart, G.W., Marchase, R.B., and Engler, J.A., *Relative accessibility of N-acetylglucosamine in trimers of the adenovirus types 2 and 5 fiber proteins.* J Virol, 1990. **64**(11): p. 5317-23.

81. Holt, G.D., Snow, C.M., Senior, A., Haltiwanger, R.S., Gerace, L., and Hart, G.W., *Nuclear pore complex glycoproteins contain cytoplasmically disposed O-linked N-acetylglucosamine.* J Cell Biol, 1987. **104**(5): p. 1157-64.

82. Dong, D.L., Xu, Z.S., Chevrier, M.R., Cotter, R.J., Cleveland, D.W., and Hart, G.W., *Glycosylation of mammalian neurofilaments. Localization of multiple O-linked N-acetylglucosamine moieties on neurofilament polypeptides L and M.* J Biol Chem, 1993. **268**(22): p. 16679-87.

83. Dong, D., Xu, Z., Hart, G., and Cleveland, D., *Cytoplasmic O-GlcNAc Modification of the Head Domain and the KSP Repeat Motif of the Neurofilament Protein Neurofilament-H.* J. Biol. Chem., 1996. **271**(34): p. 20845-20852.

84. Dieckmann-Schuppert, A., Bause, E., and Schwarz, R.T., *Studies on O-glycans of Plasmodium-falciparum-infected human erythrocytes. Evidence for O-GlcNAc and O-GlcNAc-transferase in malaria parasites.* Eur J Biochem, 1993. **216**(3): p. 779-88.

85. Shaw, P., Freeman, J., Bovey, R., and Iggo, R.O., *Regulation of specific DNA binding by p53: evidence for a role for O-glycosylation and charged residues at the carboxy-terminus.* 1996. **12**: p. 921-930.

86. Lubas, W.A. and Hanover, J.A., *Functional expression of O-linked GlcNAc transferase. Domain structure and substrate specificity.* J Biol Chem, 2000. **275**(15): p. 10983-8.

87. Datta, B., Ray, M., Chakrabarti, D., Wylie, D., and Gupta, N., *Glycosylation of Eukaryotic Peptide chain Initiation Factor 2 (eIF-2) – associated 67kDa Polypeptide (p67) and its possible role in the inhibition of eIF-2 Kinase-catalyzed phosphorylation of the eIF-2 (-subunit.* J. Biol. Chem., 1989. **34**: p. 20620-20624.

88. Kelly, W., Dahmus, M., and Hart, G., *RNA Polymerase II is a Glycoprotein.* J. Biol. Chem, 1993. **268**(14): p. 10416-10424.

89. Comer, F.I. and Hart, G.W., *Reciprocity between O-GlcNAc and O-Phosphate on the Carboxyl Terminal Domain of RNA Polymerase II.* Biochemistry, 2001. **40**(26): p. 7845-7852.

90. Luthi, T., Haltiwanger, R.S., Greengard, P., and Bahler, M., *Synapsins contain O-linked N-acetylglucosamine.* J Neurochem, 1991. **56**(5): p. 1493-8.

91. Nyame, K., Cummings, R., and Damian, R., *Schistosoma mansoni Synthesises Glycoproteins Containing Terminal O-linked N-Acetylglucosamine Residues.* J. Biol. Chem., 1987. **262**(17): p. 7990-7995.

92. Reason, A.J., Morris, H.R., Panico, M., Marais, R., Treisman, R.H., Haltiwanger, R.S., Hart, G.W., Kelly, W.G., and Dell, A., *Localization of O-GlcNAc modification on the serum response transcription factor.* J Biol Chem, 1992. **267**(24): p. 16911-21.

93. Ding, M. and Vandre, D.D., *High molecular weight microtubule-associated proteins contain O-linked- N-acetylglucosamine.* J Biol Chem, 1996. **271**(21): p. 12555-61.

94. Handman, E., Barnett, L.D., Osborn, A.H., Goding, J.W., and Murray, P.J., *Identification, characterisation and genomic cloning of a O-linked N-acetylglucosamine-containing cytoplasmic Leishmania glycoprotein.* Mol Biochem Parasitol, 1993. **62**(1): p. 61-72.

95. Medina, L., Grove, K., and Haltiwanger, R., *SV40 large T antigen is modified with O-linked N-Acetylglucosamine but not with other forms of glycosylation.* Glycobiology, 1998. **8**(4): p. 383-391.

96. Arnold, C., Johnson, G., Cole, R., Dong, D.L., M., and Hart, G., *The microtubule-associated protein Tau is extensively modified with O-linked N-Acetylglucosamine.* J. Biol. Chem., 1996. **271**(46): p. 28741-28744.

97. Greis, K.D., Gibson, W., and Hart, G.W., *Site-specific glycosylation of the human cytomegalovirus tegument basic phosphoprotein (UL32) at serine 921 and serine 952.* J Virol, 1994. **68**(12): p. 8339-49.

98. Privalsky, M.L., *A subpopulation of the avian erythroblastosis virus v-erbA protein, a member of the nuclear hormone receptor family, is glycosylated.* J Virol, 1990. **64**(1): p. 463-6.

99. Hagmann, J., Grob, M., and Burger, M.M., *The cytoskeletal protein talin is O-glycosylated.* J Biol Chem, 1992. **267**(20): p. 14424-8.

100. Jackson, S. and Tijan, R.C., *O-Glycosylation of Eukaryotic Transcription Factors: Implications of Mechanisms of Transcriptional Regulation.* 1988. **55**: p. 125-133.

101. Murphy, J.E., Hanover, J.A., Froehlich, M., DuBois, G., and Keen, J.H., *Clathrin assembly protein AP-3 is phosphorylated and glycosylated on the 50-kDa structural domain.* J Biol Chem, 1994. **269**(33): p. 21346-52.

102. Zhang, X. and Bennett, V., *Identification of O-linked N-acetylglucosamine modification of ankyrinG isoforms targeted to nodes of Ranvier.* J Biol Chem, 1996. **271**(49): p. 31391-8.

103. Kelly, W.G. and Hart, G.W., *Glycosylation of chromosomal proteins: localization of O-linked N- acetylglucosamine in Drosophila chromatin.* Cell, 1989. **57**(2): p. 243-51.

104. Miller, M.W., Caracciolo, M.R., Berlin, W.K., and Hanover, J.A., *Phosphorylation and glycosylation of nucleoporins.* Arch Biochem Biophys, 1999. **367**(1): p. 51-60.

105. Soulard, M., Della Valle, V., Siomi, M.C., Pinol-Roma, S., Codogno, P., Bauvy, C., Bellini, M., Lacroix, J.C., Monod, G., Dreyfuss, G., and et al., *hnRNP G:*

sequence and characterization of a glycosylated RNA-binding protein. Nucleic Acids Res, 1993. **21**(18): p. 4210-7.

106. Ku, N.O. and Omary, M.B., *Identification and mutational analysis of the glycosylation sites of human keratin 18.* J Biol Chem, 1995. **270**(20): p. 11820-7.
107. Chalkley, R. and Burlingame, A., *Identification of GlcNAcylation sites of peptides and alpha-crystallin using Q-TOF mass spectrometry.* J. Am. Soc. Mass Spec., 2001. **12**: p. 1106-1113.
108. D'Onofrio, M., Starr, C.M., Park, M.K., Holt, G.D., Haltiwanger, R.S., Hart, G.W., and Hanover, J.A., *Partial cDNA sequence encoding a nuclear pore protein modified by O- linked N-acetylglucosamine.* Proc Natl Acad Sci U S A, 1988. **85**(24): p. 9595-9.
109. Inaba, M. and Maede, Y., *O-N-acetyl-D-glucosamine moiety on discrete peptide of multiple protein 4.1 isoforms regulated by alternative pathways.* J Biol Chem, 1989. **264**(30): p. 18149-55.
110. Cheng, X. and Hart, G.W., *Glycosylation of the murine estrogen receptor-alpha.* J Steroid Biochem Mol Biol, 2000. **75**(2-3): p. 147-58.
111. Cheng, X. and Hart, G.W., *Alternative O-Glycosylation/ O-Phosphorylation of Serine-16 in Murine Estrogen Receptor Beta. Post-translational regulation of turnover and transactivation activity.* J Biol Chem, 2001. **9**: p. 9.
112. Roos, M., Su, K., Baker, J., and Kudlow, J., *O-Glycosylation of an Sp1-Derived Peptide Blocks Known Sp1 Protein Interactions.* Mol. Cell. Biol., 1997. **17**(11): p. 6472-6480.
113. Haltiwanger, R.S., Blomberg, M.A., and Hart, G.W., *Glycosylation of nuclear and cytoplasmic proteins. Purification and characterization of a uridine diphospho-N-acetylglucosamine:polypeptide beta-N-acetylglucosaminyltransferase.* J Biol Chem, 1992. **267**(13): p. 9005-13.
114. Kreppel, L.K., Blomberg, M.A., and Hart, G.W., *Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats.* J Biol Chem, 1997. **272**(14): p. 9308-15.

115. Kreppel, L.K. and Hart, G.W., *Regulation of a cytosolic and nuclear O-GlcNAc transferase. Role of the tetratricopeptide repeats.* J Biol Chem, 1999. **274**(45): p. 32015-22.

116. Goebl, M. and Yanagida, M., *The TPR snap helix: a novel protein repeat motif from mitosis to transcription.* TIBS, 1991. **16**: p. 173-177.

117. Das, A., Cohen, P., and Barford, D., *The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions.* EMBO J, 1998. **17**(5): p. 1192-1199.

118. Shafi, R., Iyer, S.P., Ellies, L.G., O'Donnell, N., Marek, K.W., Chui, D., Hart, G.W., and Marth, J.D., *The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny.* Proc Natl Acad Sci U S A, 2000. **97**(11): p. 5735-9.

119. Dong, D. and Hart, G., *Purification and Characterisation of an O-GlcNAc Selective N-Acetyl-(-D-glucosaminidase from Rat Spleen Cytosol.* J. Biol. Chem., 1994. **30**: p. 19321-19330.

120. Gao, Y., Wells, L., Comer, F.I., Parker, G.J., and Hart, G.W., *Dynamic O-glycosylation of nuclear and cytosolic proteins: cloning and characterization of a neutral, cytosolic beta-N-acetylglucosaminidase from human brain.* J Biol Chem, 2001. **276**(13): p. 9838-45.

121. Haltiwanger, R., Grove, K., and Philipsberg, G., *Modulation of O-linked N-acetylglucosamine levels in nuclear and cytoplasmic proteins in vivo using the peptide O-GlcNAc-(-N-acetylglucosaminidase Inhibitor O-(2-Acetamido-2-deoxy-D-glucopyranosylidene)amino-N-phenylcarbamate.* J. Biol. Chem., 1998. **273**(6): p. 3611-3617.

122. Akimoto, Y., Kreppel, L.K., Hirano, H., and Hart, G.W., *Increased O-GlcNAc transferase in pancreas of rats with streptozotocin- induced diabetes.* Diabetologia, 2000. **43**(10): p. 1239-47.

123. Comer, F.I. and Hart, G.W., *O-Glycosylation of nuclear and cytosolic proteins. Dynamic interplay between O-GlcNAc and O-phosphate.* J Biol Chem, 2000. **275**(38): p. 29179-82.

124. Thornton, T.M., Swain, S.M., and Olszewski, N.E., *Gibberellin signal transduction presents the SPY who O-GlcNAc'd me.* Trends Plant Sci, 1999. **4**(11): p. 424-428.

125. Jacobsen, S.E., Binkowski, K.A., and Olszewski, N.E., *SPINDLY, a tetratricopeptide repeat protein involved in gibberellin signal transduction in Arabidopsis.* Proc Natl Acad Sci U S A, 1996. **93**(17): p. 9292-6.

126. Swain, S.M., Tseng, T., and Olszewski, N.E., *Altered expression of spindly affects gibberellin response and plant development.* Plant Physiol, 2001. **126**(3): p. 1174-85.

127. Lefebvre, T., Alonso, C., Mahboub, S., Dupire, M.J., Zanetta, J.P., Caillet-Boudin, M.L., and Michalski, J.C., *Effect of okadaic acid on O-linked N-acetylglucosamine levels in a neuroblastoma cell line.* Biochim Biophys Acta, 1999. **1472**(1-2): p. 71-81.

128. Griffith, L. and Schmitz, B., *O-linked N-Acetylglucosamine levels in cerebellar neurons respond reciprocally to perturbations of phosphorylation.* Eur. J. Biochem., 1999. **262**: p. 824-831.

129. Ledesma, M.D., Bonay, P., Colaco, C., and Avila, J., *Analysis of microtubule-associated protein tau glycation in paired helical filaments.* J Biol Chem, 1994. **269**(34): p. 21614-9.

130. Lu, H., Zawel, L., Fisher, L., Egly, J.M., and Reinberg, D., *Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II.* Nature, 1992. **358**(6388): p. 641-5.

131. Marshall, N.F., Peng, J., Xie, Z., and Price, D.H., *Control of RNA polymerase II elongation potential by a novel carboxyl- terminal domain kinase.* J Biol Chem, 1996. **271**(43): p. 27176-83.

132. Hirose, Y. and Manley, J.L., *RNA polymerase II and the integration of nuclear events.* Genes Dev, 2000. **14**(12): p. 1415-29.

133. Laybourn, P.J. and Dahmus, M.E., *Phosphorylation of RNA polymerase IIA occurs subsequent to interaction with the promoter and before the initiation of transcription.* J Biol Chem, 1990. **265**(22): p. 13165-73.

134. Freytag, S.O., Dang, C.V., and Lee, W.M., *Definition of the activities and properties of c-myc required to inhibit cell differentiation.* Cell Growth Differ, 1990. **1**(7): p. 339-43.

135. Evan, G.I., Wyllie, A.H., Gilbert, C.S., Littlewood, T.D., Land, H., Brooks, M., Waters, C.M., Penn, L.Z., and Hancock, D.C., *Induction of apoptosis in fibroblasts by c-myc protein.* Cell, 1992. **69**(1): p. 119-28.

136. Lutterbach, B., Hann, S.M., and Biol., C., *Hierarchical Phosphorylation at N-Terminal Transformation-Sensitive Sites in c-Myc Protein Is Regulated by Mitogens and in Mitosis.* 1994. **14**(8): p. 5510-5522.

137. Frykberg, L., Graf, T., and Vennstrom, B., *The transforming activity of the chicken c-myc gene can be potentiated by mutations.* Oncogene, 1987. **1**(4): p. 415-22.

138. Raffeld, M., Yano, T., Hoang, A.T., Lewis, B., Clark, H.M., Otsuki, T., and Dang, C.V., *Clustered mutations in the transcriptional activation domain of Myc in 8q24 translocated lymphomas and their functional consequences.* Curr Top Microbiol Immunol, 1995. **194**: p. 265-72.

139. Hann, S.R., Thompson, C.B., and Eisenman, R.N., *c-myc oncogene protein synthesis is independent of the cell cycle in human and avian cells.* Nature, 1985. **314**(6009): p. 366-9.

140. Armstrong, S.A., Barry, D.A., Leggett, R.W., and Mueller, C.R., *Casein kinase II-mediated phosphorylation of the C terminus of Sp1 decreases its DNA binding activity.* J Biol Chem, 1997. **272**(21): p. 13489-95.

141. Yang, X., Su, K., Roos, M.D., Chang, Q., Paterson, A.J., and Kudlow, J.E., *O-linkage of N-acetylglucosamine to Sp1 activation domain inhibits its transcriptional capability.* Proc Natl Acad Sci U S A, 2001. **98**(12): p. 6611-6.

142. Han, I. and Kudlow, J.E., *Reduced O-glycosylation of Sp1 is associated with increased proteasome susceptibility.* Mol Cell Biol, 1997. **17**(5): p. 2550-8.

143. Su, K., Roos, M.D., Yang, X., Han, I., Paterson, A.J., and Kudlow, J.E., *An N-terminal region of Sp1 targets its proteasome-dependent degradation in vitro.* J Biol Chem, 1999. **274**(21): p. 15194-202.

144. Fanning, E. and Knippers, R., *Structure and function of simian virus 40 large tumor antigen.* Annu Rev Biochem, 1992. **61**: p. 55-85.

145. Chakraborty, A., Saha, D., Bose, A., Chatterjee, M., and Gupta, N.B., *Regulation of eIF-2 alpha-subunit phosphorylation in Reticulocyte Lysate.* 1994. **33**: p. 6700-6706.

146. Chang, Q., Su, K., Baker, J.R., Yang, X., Paterson, A.J., and Kudlow, J.E., *Phosphorylation of Human Glutamine:Fructose-6-phosphate Amidotransferase by cAMP-dependent Protein Kinase at Serine 205 Blocks the Enzyme Activity.* J Biol Chem, 2000. **275**(29): p. 21981-21987.

147. Liu, K., Paterson, A.J., Chin, E., and Kudlow, J.E., *Glucose stimulates protein modification by O-linked GlcNAc in pancreatic beta cells: linkage of O-linked GlcNAc to beta cell death.* Proc Natl Acad Sci U S A, 2000. **97**(6): p. 2820-5.

148. Hebert, L.F., Jr., Daniels, M.C., Zhou, J., Crook, E.D., Turner, R.L., Simmons, S.T., Neidigh, J.L., Zhu, J.S., Baron, A.D., and McClain, D.A., *Overexpression of glutamine:fructose-6-phosphate amidotransferase in transgenic mice leads to insulin resistance.* J Clin Invest, 1996. **98**(4): p. 930-6.

149. Akimoto, Y., Kreppel, L.K., Hirano, H., and Hart, G.W., *Hyperglycemia and the O-GlcNAc transferase in rat aortic smooth muscle cells: elevated expression and altered patterns of O-GlcNAcylation.* Arch Biochem Biophys, 2001. **389**(2): p. 166-75.

150. King, G.L., Kunisaki, M., Nishio, Y., Inoguchi, T., Shiba, T., and Xia, P., *Biochemical and molecular mechanisms in the development of diabetic vascular complications.* Diabetes, 1996. **45 Suppl 3**: p. S105-8.

151. Finlay, D.R., Newmeyer, D.D., Price, T.M., and Forbes, D.J., *Inhibition of in vitro nuclear transport by a lectin that binds to nuclear pores.* J Cell Biol, 1987. **104**(2): p. 189-200.

152. Miller, M.W. and Hanover, J.A., *Functional nuclear pores reconstituted with beta 1-4 galactose-modified O-linked N-acetylglucosamine glycoproteins.* J Biol Chem, 1994. **269**(12): p. 9289-97.

153. Haurum, J.S., Hoier, I.B., Arsequell, G., Neisig, A., Valencia, G., Zeuthen, J., Neefjes, J., and Elliott, T., *Presentation of cytosolic glycosylated peptides by*

human class I major histocompatibility complex molecules in vivo. J Exp Med, 1999. **190**(1): p. 145-50.

154. Snow, C.M., Senior, A., and Gerace, L., *Monoclonal antibodies identify a group of nuclear pore complex glycoproteins.* J Cell Biol, 1987. **104**(5): p. 1143-56.

155. Comer, F.I., Vosseller, K., Wells, L., Accavitti, M.A., and Hart, G.W., *Characterization of a mouse monoclonal antibody specific for o-linked n-acetylglucosamine.* Anal Biochem, 2001. **293**(2): p. 169-77.

156. Roquemore, E.P., Chou, T.Y., and Hart, G.W., *Detection of O-linked N-acetylglucosamine (O-GlcNAc) on cytoplasmic and nuclear proteins.* Methods Enzymol, 1994. **230**: p. 443-60.

157. Greis, K.D., Hayes, B.K., Comer, F.I., Kirk, M., Barnes, S., Lowary, T.L., and Hart, G.W., *Selective detection and site-analysis of O-GlcNAc-modified glycopeptides by beta-elimination and tandem electrospray mass spectrometry.* Anal Biochem, 1996. **234**(1): p. 38-49.

158. Rademaker, G., Pergantis, S., Blok-Tip, L., Langridge, L., Kleen, A., and Thomas-Oates, J., *Mass spectrometric determination of the sites of O-Glycan attachment with low picomolar sensitivity.* Anal. Biochem., 1998. **257**: p. 149-160.

159. Shen, M., Johnson, R., and Wallace, A. *Isotope Coded Reducible Affinity Phosphate Detection in Peptides involving Ser/Thr.* in *Proc. 49th ASMS Conf. Mass Spectrom. Allied Top.* 2001. Chicago.

160. Pulverer, B., Fisher, C., Vousden, K., Littlewood, T., Evan, G., and Woodgett, J.O., *Site-specific modulation of c-Myc co-transformation by residues phosphorylated in vivo.* 1994. **9**: p. 59-70.

161. Bosc, D., Slominski, E., Sichler, C., and Litchfield, W., *Reflectron time-of-flight mass spectrometry and laser excitation for the analysis of neutrals, ionized molecules and secondary fragments.* J. Biol. Chem., 1995. **270**(43): p. 25872-25878.

162. Patterson, S.D. and Garrels, J.I., *Two Dimensional Gel Analysis of Posttranslational Modifications*, in *Cell Biology: A Laboratory Handbook.* 1994, Academic Press. p. 249-257.

163. http://www.protana.com/products/applicationnotes/purification/default.asp.

164. Marais, R.M., Hsuan, J.J., McGuigan, C., Wynne, J., and Treisman, R., *Casein kinase II phosphorylation increases the rate of serum response factor-binding site exchange.* EMBO J, 1992. **11**(1): p. 97-105.

165. http://ncbi.nlm.nih.gov/blast/db/nr.Z.

166. Easty, D.M., Easty, G.C., Carter, R.L., Monaghan, P., Pittam, M.R., and James, T., *Five human tumour cell lines derived from a primary squamous carcinoma of the tongue, two subsequent local recurrences and two nodal metastases.* Br J Cancer, 1981. **44**(3): p. 363-70.

167. Gerner, C., Holzmann, K., Grimm, R., and Sauermann, G., *Similarity between nuclear matrix proteins of various cells revealed by an improved isolation method.* J. Cellular. Biochem., 1998. **71**: p. 363-374.

168. Kita, K., Omata, S., and Horigome, T., *Purification and characterization of a nuclear pore glycoprotein complex containing p62.* J. Biochem., 1993. **113**: p. 377-382.

169. Haynes, P.A. and Aebersold, R., *Simultaneous detection and identification of O-GlcNAc-modified glycoproteins using liquid chromatography-tandem mass spectrometry.* Anal Chem, 2000. **72**(21): p. 5402-10.

170. Oudejans, R.C., Kooiman, F.P., Heerma, W., Versluis, C., Slotboom, A.J., and Beenakkers, M.T., *Isolation and structure elucidation of a novel adipokinetic hormone (Lom-AKH-III) from the glandular lobes of the corpus cardiacum of the migratory locust, Locusta migratoria.* Eur J Biochem, 1991. **195**(2): p. 351-9.

171. Hayes, B.K., Greis, K.D., and Hart, G.W., *Specific isolation of O-linked N-acetylglucosamine glycopeptides from complex mixtures.* Anal Biochem, 1995. **228**(1): p. 115-22.

172. Lagerwerf, F.M., van de Weert, M., Heerma, W., and Haverkamp, J., *Identification of oxidized methionine in peptides.* Rapid Commun Mass Spectrom, 1996. **10**(15): p. 1905-10.

173. Clark, J.I. and Muchowski, P.J., *Small heat-shock proteins and their potential role in human disease.* Curr Opin Struct Biol, 2000. **10**(1): p. 52-9.

174. Derham, B.K. and Harding, J.J., *Alpha-crystallin as a molecular chaperone.* Prog Retin Eye Res, 1999. **18**(4): p. 463-509.

175. Klemenz, R., Frohli, E., Steiger, R.H., Schafer, R., and Aoyama, A., *Alpha B-crystallin is a small heat shock protein.* Proc Natl Acad Sci U S A, 1991. **88**(9): p. 3652-6.

176. Litt, M., Kramer, P., LaMorticella, D.M., Murphey, W., Lovrien, E.W., and Weleber, R.G., *Autosomal dominant congenital cataract associated with a missense mutation in the human alpha crystallin gene CRYAA.* Hum Mol Genet, 1998. **7**(3): p. 471-4.

177. Hunter, C., Fell, L., Pace, N., and Jones, E. *The Enhancement of Sensitivity and Dynamic Range due to Q2 Pulsing on a hybrid Quadrupole-Time of Flight Instrument.* in *Proc. 49th ASMS Conf. Mass Spectrom. Allied Top.* 2001. Chicago.

178. Steen, H., Kuster, B., Fernandez, M., Pandey, A., and Mann, M., *Detection of tyrosine phosphorylated peptides by precursor ion scanning quadrupole TOF mass spectrometry in positive ion mode.* Anal Chem, 2001. **73**(7): p. 1440-8.

179. Treisman, R., *The SRE: a growth factor responsive transcriptional regulator.* Semin Cancer Biol, 1990. **1**(1): p. 47-58.

180. Hill, C.S., Wynne, J., and Treisman, R., *The Rho family GTPases RhoA, Rac1, and CDC42Hs regulate transcriptional activation by SRF.* Cell, 1995. **81**(7): p. 1159-70.

181. Arsenian, S., Weinhold, B., Oelgeschlager, M., Ruther, U., and Nordheim, A., *Serum response factor is essential for mesoderm formation during mouse embryogenesis.* EMBO J, 1998. **17**(21): p. 6289-99.

182. Croissant, J.D., Kim, J.H., Eichele, G., Goering, L., Lough, J., Prywes, R., and Schwartz, R.J., *Avian serum response factor expression restricted primarily to muscle cell lineages is required for alpha-actin gene transcription.* Dev Biol, 1996. **177**(1): p. 250-64.

183. Gauthier-Rouviere, C., Cavadore, J.C., Blanchard, J.M., Lamb, N.J., and Fernandez, A., *p67SRF is a constitutive nuclear protein implicated in the modulation of genes required throughout the G1 period.* Cell Regul, 1991. **2**(7): p. 575-88.

184. Pellegrinin, L., Tan, S., and Richmond, T.J., *Structure of serum response factor core bound to DNA.* Nature, 1995. **376**: p. 490-498.

185. Rivera, V.M., Miranti, C.K., Misra, R.P., Ginty, D.D., Chen, R.H., Blenis, J., and Greenberg, M.E., *A growth factor-induced kinase phosphorylates the serum response factor at a site that regulates its DNA-binding activity.* Mol Cell Biol, 1993. **13**(10): p. 6260-73.

186. Heidenreich, O., Neininger, A., Schratt, G., Zinck, R., Cahill, M.A., Engel, K., Kotlyarov, A., Kraft, R., Kostka, S., Gaestel, M., and Nordheim, A., *MAPKAP kinase 2 phosphorylates serum response factor in vitro and in vivo.* J Biol Chem, 1999. **274**(20): p. 14434-43.

187. Miranti, C.K., Ginty, D.D., Huang, G., Chatila, T., and Greenberg, M.E., *Calcium activates serum response factor-dependent transcription by a Ras- and Elk-1-independent mechanism that involves a Ca2+/calmodulin- dependent kinase.* Mol Cell Biol, 1995. **15**(7): p. 3672-84.

188. Manak, J.R., de Bisschop, N., Kris, R.M., and Prywes, R., *Casein kinase II enhances the DNA binding activity of serum response factor.* Genes Dev, 1990. **4**(6): p. 955-67.

189. Liu, S.H., Ma, J.T., Yueh, A.Y., Lees-Miller, S.P., Anderson, C.W., and Ng, S.Y., *The carboxyl-terminal transactivation domain of human serum response factor contains DNA-activated protein kinase phosphorylation sites.* J Biol Chem, 1993. **268**(28): p. 21147-54.

190. Schroter, H., Mueller, C.G., Meese, K., and Nordheim, A., *Synergism in ternary complex formation between the dimeric glycoprotein p67SRF, polypeptide p62TCF and the c-fos serum response element.* EMBO J, 1990. **9**(4): p. 1123-30.

191. Reason, A.J., Blench, I.P., Haltiwanger, R.S., Hart, G.W., Morris, H.R., Panico, M., and Dell, A., *High-sensitivity FAB-MS strategies for O-GlcNAc characterization.* Glycobiology, 1991. **1**(6): p. 585-94.

192. Geiger, T. and Clarke, S., *Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation.* J Biol Chem, 1987. **262**(2): p. 785-94.

193. Thiede, B., Lamer, S., Mattow, J., Siejak, F., Dimmler, C., Rudel, T., and Jungblut, P.R., *Analysis of missed cleavage sites, tryptophan oxidation and N-terminal pyroglutamylation after in-gel tryptic digestion.* Rapid Commun Mass Spectrom, 2000. **14**(6): p. 496-502.

194. Norman, C., Runswick, M., Pollock, R., and Treisman, R., *Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element.* Cell, 1988. **55**(6): p. 989-1003.

195. Wright, C.S. and Kellogg, G.E., *Differences in hydropathic properties of ligand binding at four independent sites in wheat germ agglutinin-oligosaccharide crystal complexes.* Protein Sci, 1996. **5**(8): p. 1466-76.

196. Gerace, L., Ottaviano, Y., and Kondor-Koch, C., *Identification of a major polypeptide of the nuclear pore complex.* J Cell Biol, 1982. **95**(3): p. 826-37.

197. Fey, E.G., Krochmalnic, G., and Penman, S., *The nonchromatin substructures of the nucleus: the ribonucleoprotein (RNP)-containing and RNP-depleted matrices analyzed by sequential fractionation and resinless section electron microscopy.* J Cell Biol, 1986. **102**(5): p. 1654-65.

198. Biffo, S., Sanvito, F., Costa, S., Preve, L., Pignatelli, R., Spinardi, L., and Marchisio, P.C., *Isolation of a novel beta4 integrin-binding protein (p27(BBP)) highly expressed in epithelial cells.* J Biol Chem, 1997. **272**(48): p. 30314-21.

199. Chou, C.F., Smith, A.J., and Omary, M.B., *Characterization and dynamics of O-linked glycosylation of human cytokeratin 8 and 18.* J Biol Chem, 1992. **267**(6): p. 3901-6.

200. Assier, E., Bouzinba-Segard, H., Stolzenberg, M.C., Stephens, R., Bardos, J., Freemont, P., Charron, D., Trowsdale, J., and Rich, T., *Isolation, sequencing and expression of RED, a novel human gene encoding an acidic-basic dipeptide repeat.* Gene, 1999. **230**(2): p. 145-54.

201. Jumaa, H., Guenet, J.L., and Nielsen, P.J., *Regulated expression and RNA processing of transcripts from the Srp20 splicing factor gene during the cell cycle.* Mol Cell Biol, 1997. **17**(6): p. 3116-24.

202. Mahe, D., Mahl, P., Gattoni, R., Fischer, N., Mattei, M.G., Stevenin, J., and Fuchs, J.P., *Cloning of human 2H9 heterogeneous nuclear ribonucleoproteins.*