THE UNIVERSITY *of* EDINBURGH

The structural role of linker histone H5 tails in the formation of the 30nm fiber of chromatin

Jennifer A. Lynch.

Thesis presented for the degree of Master of Philosophy

University of Edinburgh, 2007

Table of Contents

Declaration

I declare that this thesis was composed by myself, and that the work contained herein is my own, except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification, except as specified.

Jennifer A. Lynch

Abstract

Eukaryotic DNA associates with histones to form nucleosomes, the fundamental component of chromatin. There are two functionally distinct classes of histones, core histones and linker histones. The core histones form an octamer around which DNA winds and a molecule of linker histone associates with this complex completing the nucleosome. The binding of linker histones to chromatin appears to be directed by the interaction of two sites in the folded globular domain with a site or sites close to the dyad axis of the core particle at the point where the DNA enters and exits after making two turns around the core histone octamer. This interaction must determine the location and orientation of the 'tails' of the linker histones allowing them to neutralize the linker DNA and permit the nucleosomal chain to adopt higher-order folding. However, there is little evidence of how the tails are organized or how they affect this folding process. It is toward this problem that this project is directed.

The tails of the linker histones are rich in basic amino acids with the majority of the lysines and arginines found in the C-terminal and N-terminal tails. The tails also remain unstructured in contrast to the globular domain, which is folded even in low ionic strength buffers. Taking advantage of the structured and unstructured nature of the linker histone the proteolytic enzyme trypsin will be used as a probe to assess the structure of the tails. The aim is to mildly cleave the molecule and identify the sites of the cutting under a variety of relevant conditions using mass spectrometry to measure the mass of the digest products.

The organization of the tails was probed under the following conditions; by first digesting the H5 alone, then digesting the H5 in the presence of DNA and finally by digesting the H5 in the presence of chromatin, previously stripped of linker histones, in varying ionic strength buffers. The digest products were run on SDS/PAGE gel and by direct mass spectrometric analysis.

The aim of investigating the linker histone tails is to build upon the body of evidence regarding the higher-order organization of chromatin.

Aims of Thesis

The aims of the thesis were to use mass spectrometry to identify the trypsin sensitive sites of the linker histone H5 tails as they impact the formation of the 30nm fiber of chromatin.

Dedication:

If you want to improve, be content to be thought foolish and stupid with regard to external things. Don't wish to be thought to know anything; and even if you appear to be somebody important to others, distrust yourself. For, it is difficult to both keep your faculty of choice in a state conformable to nature, and at the same time acquire external things. But while you are careful about the one, you must of necessity neglect the other.

For my Mother:

Natures's first green is gold,
Her hardest hue to hold
Her early leaf's a flower,
But only so an hour.
Then leaf subsides to leaf,
So Eden sank to grief.
So dawn goes down to day,
Nothing gold can stay.

Robert Frost

For my Father:

Jenny kissed me when we met
Jumping from the chair she sat in
Time you thief who love to get
Sweets into your list
Put that in.
Say I'm weary say I'm sad
Say that health and
wealth have missed me
Say I'm growing old but add
Jenny kissed me.

Leigh Hunt

I would like to thank Dr. Jim Allan and Professor John Monaghan for their support and assistance throughout this process.

I would also like to thank all my friends and family for their encouragement and support.

List of Figures:

Chapter 4: Results

List of Tables:

Chapter 4: Results

List of Equations:

Chapter 1: Introduction

Equation 1.1 $fv=qE$

Equation 1.2 $\mu= v/E=q/f=Ze/f$

Chapter 2: Mass Spectrometry

Equation 2.1 $$E = \frac{\Delta V}{s_a}$$

Equation 2.2 $$KE = q\Delta V = qEs_a = \tfrac{1}{2} mu_D^2$$

$$u_D = \sqrt{\frac{2qEs_a}{m}}$$

Equation 2.3 $$F = Eq$$

Equation 2.4 $$F = ma \quad \therefore \quad a = \frac{Eq}{m}$$

Equation 2.5 $$a = \frac{\mathrm{d}u}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{Eq}{m}\right)$$

$$\therefore \quad u = \int\left(\frac{Eq}{m}\right)\mathrm{d}t$$

$$\therefore \quad u = u_o + \left(\frac{Eq}{m}\right)t$$

Equation 2.6 $$t_a = \frac{u_D - u_o}{E}\left(\frac{m}{q}\right)$$

Equation 2.7 $$t_D = \frac{D}{u_D}$$

$$t_D = \left(\frac{m}{2qEs_a}\right)^{1/2} D \quad = \quad \left(\frac{m}{2q\Delta V}\right)^{1/2} D$$

Equation 2.8
$$TOF = t_o + t_a + t_D + t_d$$
$$= t_o + \left[\frac{u_D - u_o}{E}\left(\frac{m}{q}\right)\right] + \left[\left(\frac{m}{2q\Delta V}\right)^{1/2} D\right] + t_d$$

Equation 2.9
$$V_+ = U + V\cos(\omega t)$$

Equation 2.10
$$V_- = -U + V\cos(\omega t + \pi)$$
$$\therefore \quad V_- = -\left[U + V\cos(\omega t)\right]$$

Equation 2.11
$$R = \frac{m_1}{\Delta m}$$

Equation 2.12
$$R(ppm) = 10^6 \frac{\Delta m}{m_1}$$

Equation 2.13
$$m = \mathrm{A}t^2 \quad \therefore \quad \frac{\mathrm{d}m}{\mathrm{d}t} = 2\,\mathrm{A}\,t$$

Equation 2.14
$$\Rightarrow \quad \frac{\mathrm{d}m}{m} = \frac{2\,\mathrm{d}t}{t} \quad \therefore \quad \frac{m}{\Delta m} = \frac{t}{2\Delta t}$$

1. Introduction

DNA is a dynamic structure packaged into chromatin in eukaryotic cell nuclei. This occurs as a result of electrostatic interaction of DNA and nuclear proteins. The packaging of the DNA by proteins occurs at several different levels, each level conferring successive degrees of condensation. Understanding of the condensation of DNA came about first through work with nucleases in which DNA was found to fragment into discrete sizes. Electron microscopic studies, Figure 1.2 (Bednar et al. 1998), further showed the nature of the protein/DNA complex with pictures of the 'beads on a string' model. Chemical cross-linking experiments determined that the proteins and DNA making up the nucleosome exist in the nucleus in a precise stoichiometry (Wolffe, 1998). The histones, a family of small, highly conserved proteins, are the nuclear proteins essential in the establishment and modulation of chromatin structure. Discovered over a century ago, it is only recently that high-resolution structures of the histones and the core particle have become available allowing for better understanding of the role they play in the higher-order structure of chromatin.

1.1 ***Heirarchical orders of chromatin structure/Higher-Order Structure***

The Nucleosome Core Particle

The nucleosome is the fundamental unit of chromatin, comprised of two molecules of each of the four core histones (H3, H4, H2A and H2B), and ~146 base pair of DNA.

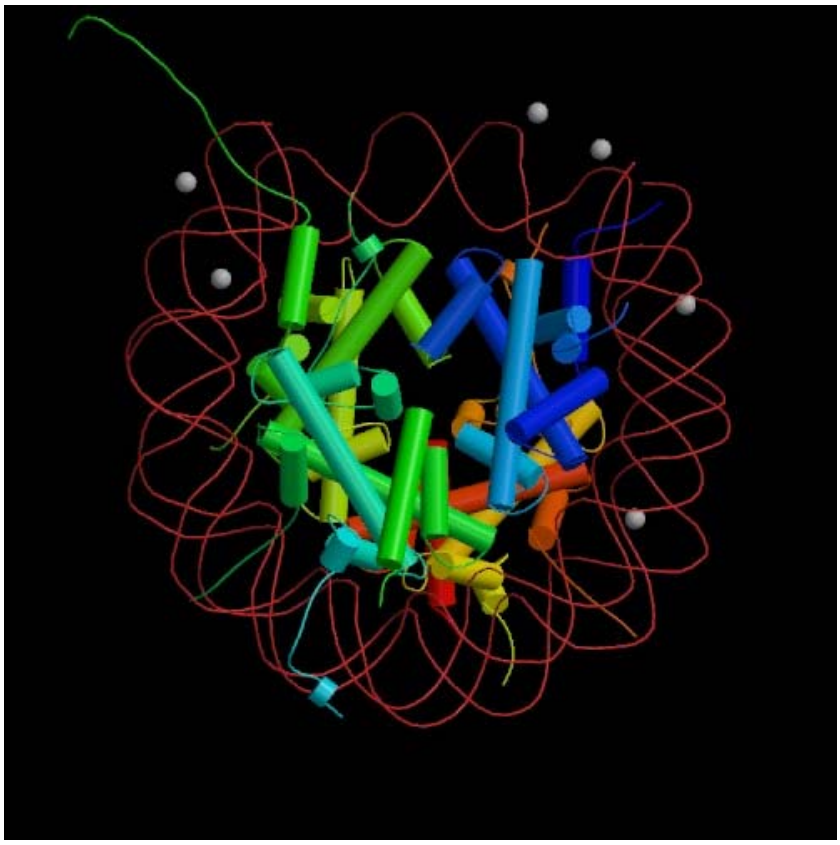

Figure 1.1 Taken from (Luger et al. 1997)Crystal structure of the nucleosome core particle at 2.8Å resolution. Yellow-H2A, Red-H2B, Blue-H3, Green-H4 silver spheres-$Mn2^{+}$ and the thin red ribbon-DNA

The H4/H3 dimer and the H2A/H2B dimer: The core histones associate with one another in a distinct relationship; H2A and H2B form a dimer in a so called handshake motif. The histone-fold portion of each protein is related to its partner in the dimer by an approximate twofold axis of symmetry. The dimer is stabilized largely through hydrophobic interactions that span the entire length of the histone-fold regions of each monomer with the long central helices associated in an antiparallel manner. In the dimer, the N- and C-termini of each histone are close

together. The strand in the helix/strand/helix (HSH) motif in the N-terminal half of a monomer pairs with the strand in the HSH motif of the C-terminal half of its partner to form a short two-stranded β-sheet or β-bridge. A similar sheet is found at the other end in the symmetric position. The solvent-exposed surfaces of these sheets consist of conserved basic residues that are likely to be involved in binding DNA.

The tetramer: The tetramer, two of the H4/H3 dimers, is important because it is involved in the initial step in nucleosome assembly and in determining nucleosome positioning. The two halves of the tetramer, which consist of the heterodimers of H4/H3, are related by a crystallographic twofold axis. The shape of the tetramer is relatively flat, resembling a twisted open horseshoe.

The octamer: An H2A/H2B dimer associates on each side of the $(H3/H4)_2$ tetramer to form the histone octamer, both in vivo as part of the nucleosome and in-vitro at high ionic strength. This association occurs primarily via an interaction between the C-terminal halves of the H2B and H4 that mimics the interaction between two H3 molecules in the tetramer (4 helix bundle). The binding of the H2A/H2B dimers to opposite sides of the tetramer results in a tripartite structure for the octamer. Depending on the point of view, the octamer looks like a disk or can resemble the wedge shape previously seen in electron micrographs. Electron microscopy had also detected the presence of a left-handed superhelical ramp on the surface of the octamer, which presumably was the path of interaction with DNA. The high-resolution structure of the octamer confirms the presence of this ramp but also suggests the detailed nature of the interactions with DNA (Luger et al. 1997).

The core particle: The nucleosome core particle consists of about 146bp of DNA wrapped in 1-3/4 turns around the histone octamer (see figure 1.1). The particle was originally produced as the limit digestion of nucleosomes by micrococcal nuclease. The crystal structure of the core particle has been resolved to 2.8Å. At this resolution it has now become possible to visualize the major and minor grooves of the DNA. The histone-fold domains account directly for the organization of 121bp of DNA. Each histone-fold pair, whether half of the H3/H4 tetramer or a H2A/H2B dimer, is associated with 27-28bp of DNA, leaving 4bp linkers between these units. Binding is primarily to the DNA phosphodiester backbones as they face the protein over 2.5 turns of the double helix. Each helical turn of a chain commits a segment of only two

consecutive phosphate groups to hydrogen-bonding interactions with the histone-fold pairs (minor groove of DNA, some contacts with the bases).

The Chromatosome

The nucleosome core particle along with 168 bp of DNA and the linker histone make up the chromatosome, with the linker histone or just the globular domain, binding at the dyad axis of the DNA. Nucleosomes are connected together by lengths of DNA referred to as linker DNA (van Holde 1989). Electron microscopic studies (see figure 1.2) of chromatin (Finch and Klug 1976) and (Thoma, Koller and Klug 1979) in buffers of varying salt concentrations showed that in low concentrations of salt, the 'beads on a string' structure was measured to be 10nm in diameter; this is considered the first level of chromatin condensation.

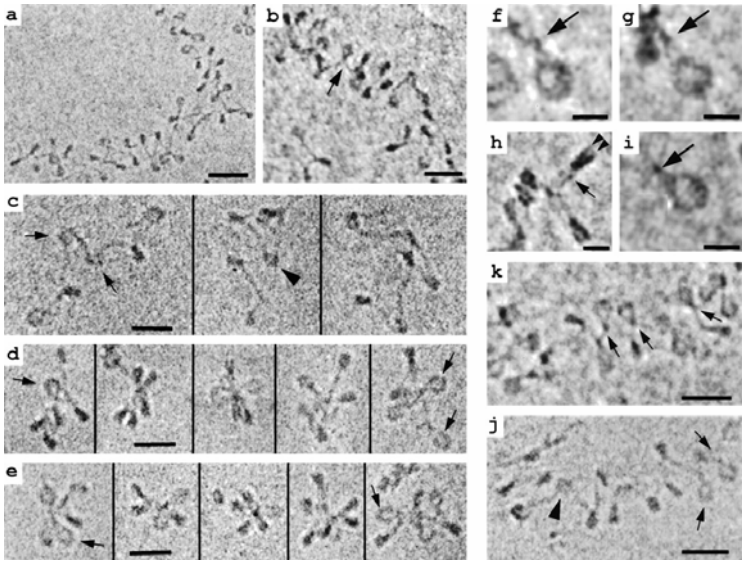

Figure 1.2 Taken from (Bednar et al. 1998) electron micrograph showing the 'beads on a string' of chromatin. (*a*,*b*, and *k*) Soluble chromatin from chicken erythrocyte nuclei vitrified in ~5mM M^+ and imaged unfixed and unstained in the frozen hydrated state. Nucleosomes and linker DNA are seen in many different orientations in these projections of the 3D structure. Arrows in *b* and *k* denote nucleosomes with the linker histone-dependent "stem" conformation. (*c*, *d*, and f-*i)* EC-M of unstained, unfixed chromatin reconstitutued onto tandem DNA sequences containing the nucleosome-positioning sequence of 5S rDNA and vitrified in 10mM M^+. All samples except *c* contain linker histone H5. *En face* views of nucleosome (*f, g*, and *i*) show the linker DNA entering and exiting the nucleosome tangentially, then "intersecting" and remaining apposed for 3-5 nm before diverging (arrows). Edge-on views (*h*) show the two gyres of DNA (arrowheads) and the apposition of linker DNA (arrow_. (*c*) Reconstituted hexanulceosomes without histone H5. In most nucleosomes, the linker DNA segments diverge after leaving the histone core (arrows), but in some cases they appear to "cross over" (arrowhead). (*d*) Reconstituted hexanucleosomes as in *c*, but with added linker histone H5. The chromatin particles are more compact and adopt a star-like conformation in which "stem" structures are common (arrows). (*e*) EC-M images of small oligonucleosomes released from chicken erythrocyte nuclei after micrococcal nuclease digestion. A zigzag, star like conformation is seen, and, like the

reconstituted hexanucleosomes in *d*, linker DNA segments show the "stem" architecture (arrows) (*j*) Chromatin released from COS-7 cells and vitrified in 20mM M^+ again shows a 3D zigzag conformation. While most nucleosomes show the stem motif (arrows), a few (arrowhead) have a linker DNA conformation typical of H1-free chromatin. [Bars = 10 nm (*f-i*); 30 nm for others.]

The globular domain of the linker histone is believed to interact with chromatin at the dyad axis; a point where the DNA enters and exits the core particle (Allan et al. 1980). There is now evidence to show the linker histone may be located asymmetrically with respect to the dyad axis (Zhou et al. 1998)

Linker histones

The precise stoichiometry of H1 within the chromatin fiber is uncertain, but in higher eukaryotes, there is, on average, nearly one H1 molecule for each core particle. Most of our knowledge about the role of H1 in chromatin structure is based on *in-vitro* experiments. These studies indicate that two principal functions of linker histones are to organize and stabilize the DNA as it enters and exits the core particle and to facilitate the folding of nucleosome arrays into more compact structures.

The linker histones H1 and H5 (a variant of H1 found in mature nucleated erythrocytes of birds, fish and frogs) are an important component in the formation of the higher-order structure and are involved in the repression of transcription. Digestion of chromatin with micrococcal nuclease generates an intermediate structure between the nucleosome (~200bp DNA) and the core particle (~146bp DNA). This chromatosome is comprised of ~168bp DNA, an octamer of the core histones and a molecule of linker histone. The linker histone protects an extra ~20bp of DNA. The exact location of the linker histone within the chromatosome structure is uncertain although the binding of the linker histones to chromatin appears to be directed by the interaction of the folded globular domain of the linker histone. Linker histones also bind preferentially to 4-way structures, the dyad axis, the entry/exit site of the DNA from the core histones. The linker histone is not essential for folding of the chromatin fiber but it is necessary for the correct folding. Increased salt concentration in the absence of linker histone has been shown to have an effect on the compaction of chromatin. Under these conditions the chromatin lacks the same structure as observed when in the presence of linker histones. (Wolfe 1998).

The globular domain

Of the 189 amino acids in the H5 sequence, 78 of them comprise the globular domain. This globular domain lies between a relatively short highly basic N-terminal tail and a long highly basic C-terminal tail of the protein. (GH1, globular domain of histone H1; GH5, globular domain of histone H5) The structure of the globular domain, referred to as a winged-helix, is comprised of three α-helices attached to a three-stranded β-sheet, see figure 1.3

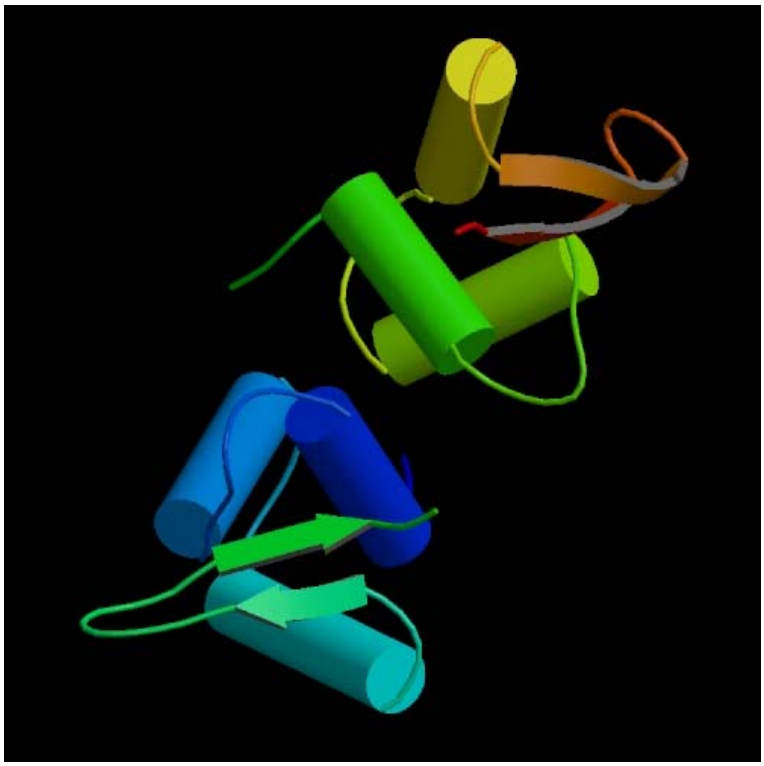

Figure1.3 Crystal structure of a dimer of the globular domain of linker histone. Cylinders-α-helices, arrows-β-sheet. (image taken from Ramakrishnan et al. 1993).

The globular domain interacts with the core particle at either the dyad axis of the DNA entry/exit site or asymmetrically to this site (figure 1.4). In either case it protects ~20 base pair of DNA from micrococcal nuclease digestion. It is not clear if this is symmetric or asymmetric positioning (Thomas, 1999). The binding of the globular domain of the linker histone to the nucleosome has been shown to have two DNA binding sites (Goytisolo et al. 1996). It is these two sites which are involved in the formation of the nucleosome (figure 1.5).

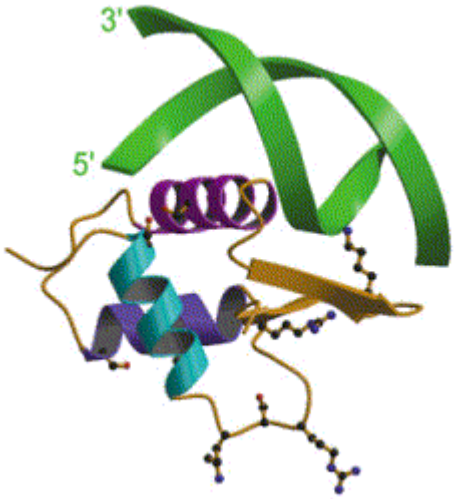

Figure 1.4 Interaction of GH5 DNA-binding site 1 with DNA, modeled by homology to the co-crystal structure of the related transcription factor HNF3. Figure from (Widom 1998) the DNA is shown as the green ribbon with the 3' and 5' labeled, the α-helices of the GH5 in purple, blue and aqua and the β-sheets in gold with the lysine sidechains indicated by the ball and stick models.

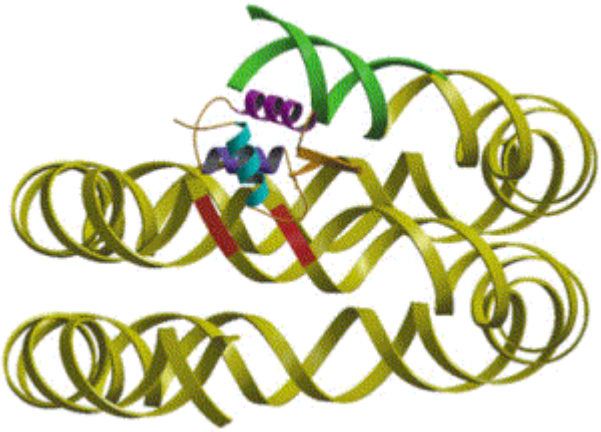

Figure 1.5 Location of GH5 on a model of the nucleosome core-particle DNA superhelix, figure from (Widom 1998) the binding site of the DNA shown in green, the remaining DNA shown in gold and the GH5 α-helices shown in purple, blue and aqua.

With 32 of the 44 lysines and 13 of the 22 arginines located in the C-terminal tail of H5, the role of partially neutralizing the DNA for further condensation in the higher-order structuring of chromatin would appear to lie here. Further information is important for a full understanding of the balance between gene activation and repression at two levels. The first is the relationship between H1-binding and transcription factor binding to nucleosomes and the second is the possible gene-specific effects of H1, and the formation of H1-dependent higher-order structures, the detailed nature of which is still not well defined. GH1 has a lower affinity for

nucleosome binding than GH5. This is due to the sequence at the loop between helix II and helix III. In GH5 it is a histidine and in GH1 contains acidic residues (Ali et al. 2004).

Tails of linker histones

The tails of the linker histones comprise over half of the molecule. The C-terminal tail is required for condensation of nucleosomes into a 30nm fiber (Allan et al. 1986), (Hendzel et al. 2004), (Th'ng et al. 2005). The C-terminal tails may assume a segmented α-helix conformation, (Clark et al. 1988), (Roque et al. 2005) that could tract the grooves on DNA (Bharath et al. 2002), (Roque 2004). The α-helical segments would be broken by the prolines that occur throughout the sequence of the tails. Serines and threonines, which are the sites of phosphorylation, precede these prolines. This sequence relates to the SPKK motif (in this case S/TPXX, where XX is R or K) a sequence shown to be for DNA binding. But an NMR study, (Suzuki et al. 1993), provided only weak evidence for this motif resulting in a β-turn or α-turn, not as part of a turn but as an extended structure where basic residues make contact with phosphate backbone of DNA.

The structure of the 30nm fiber

An accurate model of the 30nm fiber does not exist at present. There are currently three models (see figure 1.6) for the structure of this 30nm fiber, the solenoid model (Finch et al. 1976), the zig-zag model (Woodcock et al. 1993) and the cross-linker model (Staynov 1983).

- In the solenoid model (figure 1.6a) the chain of nucleosomes forms a helical structure with the axis of the core particles being roughly perpendicular to the solenoid axis. The DNA entry-exit site faces into the center of the solenoid axis requiring the linker DNA to bend.
- In the zig-zag model (figure1.6b and figure 1.7) the linker DNA is straight, connecting nucleosomes located on opposite sides of the fiber forming a zig-zag pattern. With the zig-zag model the linker histones may not be as internalized.
- In the cross linker model (figure1.6c) the linker histones would be located within the central core of the solenoid.

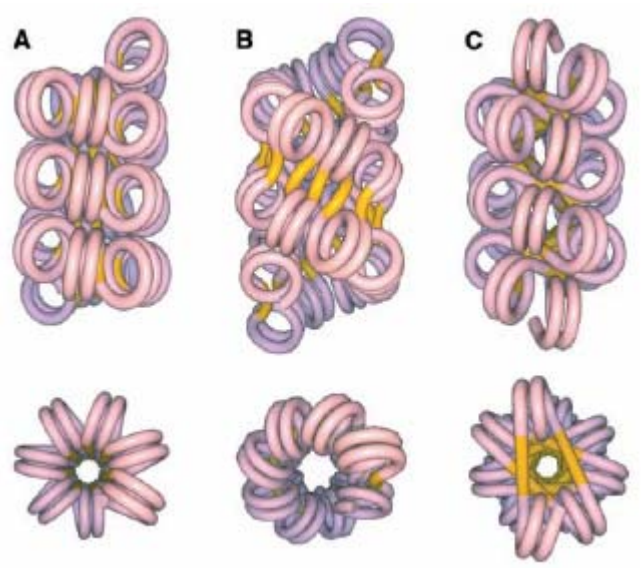

Figure 1.6 Models for the DNA path in the chromatin fiber. Higher order structure models: (A) one-start solenoidal. (B) two-start supercoiled. (C) two-start twisted. Upper views have the fiber axis running vertically; lower views are down the fiber axis. DNA associated with the nucleosome core is red/blue, and linker DNA running between cores is yellow. These models are idealized with nucleosome cores in each start contacting each other. The open three-dimensional zigzag seen in conditions not fully compacting may be a precursor. Taken from reference (Dorigo et al. 2004).

Insight into the construction of the 30nm fiber is a necessary step to understanding the regulation of DNA. In order to understand the 30nm fiber it is necessary to learn the structure and function of its components.

The location of the globular domain must direct the location and orientation of the tails permitting them to neutralize the linker DNA and allowing the nucleosomal chain to adopt higher-order folding.

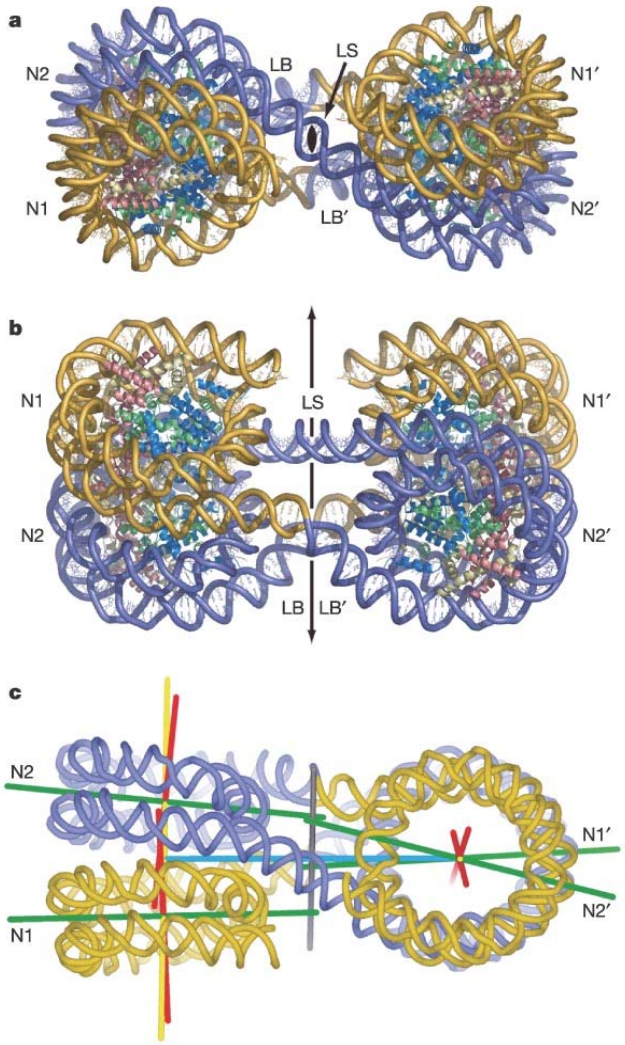

Figure 1.7 taken from (Schalch et al. 2005). X-ray crystal structure at 9Å of the tetranucleosome demonstrating the two-start (zig-zag) model as in figure 1.6a above.

1.2 Linker Histone Mobility

The location of the linker histone is still under investigation but has been shown to be mobile within the nucleus of a living cell. The nucleosome will remain intact, but will still have to be mobile along the DNA; the linker histone in this instance may act as a lock mobilizing an additional 20bp of DNA. This 'lock' will need to be released to allow the core histones movement along the DNA allowing access to specific genes. The linker histones must be mobile along the chromatin fiber. (Lever et al. 2000), (Misteli et al., 2000), (Brown et al. 2006),

As shown in figure 1.8a, using green fluorescent protein (GFP) tagging of H1.1 (a variant of H1) followed by photobleaching monitoring of the system shows the

recovery of fluorescence in the cell nucleus as shown in figure 1.8a where the H1.1 has been tagged and not in figure 1.8b where the H2B has been tagged.

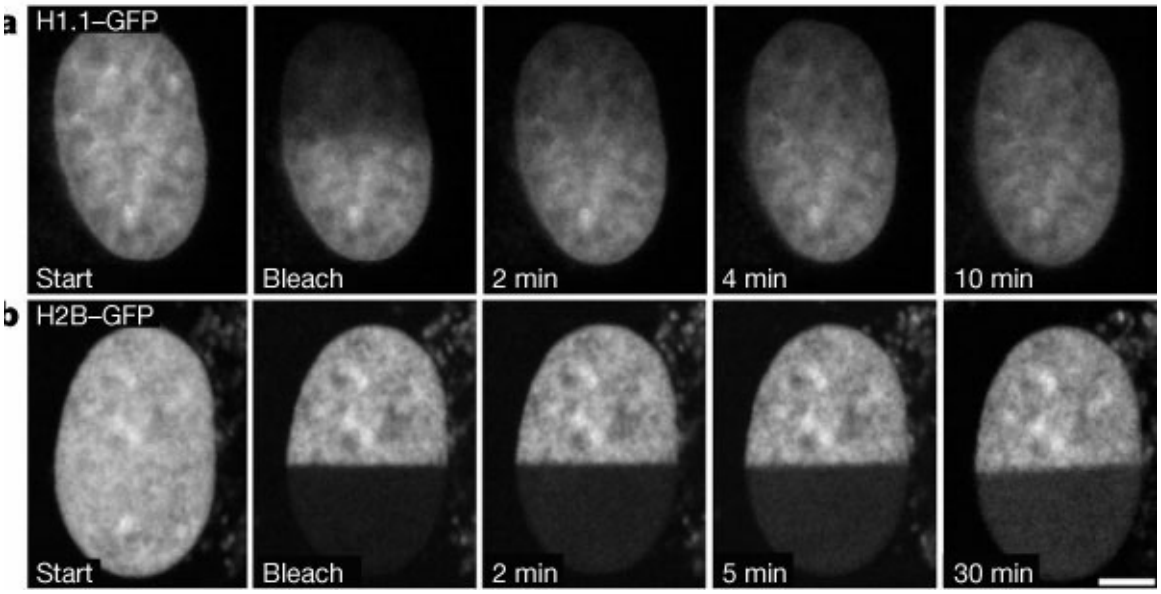

Figure 1.8 (Lever et al. 2000) used green fluorescent protein (GFP) tagging of H1.1 followed by photobleaching, and monitored the recovery of fluorescence within the nucleus. The top series of pictures are the GFP tagged H1.1, note that after 10 minutes the fluorescence again appears throughout the nucleus, as at the start. In fact, after even just 2 minutes the nucleus appears to have fluorescence throughout, although not as equally dispersed as at 10 minutes or at the start. In the bottom series of photographs, the H2B has been tagged with GFP and after 30 minutes the fluorescence has failed to return to the entire nucleus indicating a lack of movement on the part of the H2B. The distribution of the fluorescence remains in the portion of the nucleus where photobleaching did not occur.

This experiment showed the movement of the linker histone through the chromatin fiber. After laser ablation, the GFP loses fluorescence and as a result the only way the fluorescence can return is through the movement (diffusion) of the H1 through the fiber (or in the surrounding medium). The H2B results are a control for an "immobile" component. The lack of movement from H2B in the fiber is demonstrated by the lack of fluorescence recovery in the ablated region of the nucleus even after a 30-minute period. The structure of the nucleosome (see figure 1.1) inhibits any mobility of the H2B.

1.3 Histone modifications

The nucleosome, and therefore the histones, have a second role beyond compaction of DNA but obviously related to it, regulation of gene transcription. Transcription

requires access to the DNA and this means the chromatin must at least partially unfold in order for regulatory proteins to access the DNA (Zlantanova et al. 2000). As this partial unfolding of chromatin occurs there are modifications added to some of the amino acids of histones, particularly those in the 'tails' but it is not clear if the modifications dictate the unfolding or are a result of the unfolding. All modifications can affect one another, and many are positively or negatively correlated with each other. This modification variation is referred to as the 'histone code' (Strahl and Allis 2000). This means of storing and relaying information involves a variety of post-translational modifications of the core and linker histone tails. Lysine, arginine, histidine, serine and threonine are the primary amino acids involved in this process. Each of these amino acids has a possible modification, some more than one. The following list of modifications covers most of these.

- Acetylation: occurs mainly on the tails of the core histones at ε-amino groups in the N-terminal tail domain and a minor one involving the blocking of the α-amino group at the N-terminal residue. While any lysine could be involved in acetylation, only specific residues are involved. The core histone tails are involved in chromatin organization, as shown by Allan et al., who showed that the tails are necessary for fiber stabilization. Acetylation could therefore have the effect of opening the chromatin to associated proteins for transcription. As a result, this loosening of the chromatin fiber by acetylation is believed to be a key factor in gene control.
- Methylation: Lysine residues located in the N-terminal tails of the histones are methylated on the sidechains and are involved the recruitment of chromatin associated proteins.
- Ubiquitination: involved in the mitotic and meiotic processes of the cell, occurs along the C-terminal tail.
- Phosphorylation: Occurs on both core and linker histones in either of two distinct ways; on serine and threonine residues, or on lysine, histidine and possibly arginine residues.

  The main type of phosphorylation occurs on serine and is cell cycle dependent on N-terminal tail domain of core histones with some occurrence in the C-terminal region.

(Rea et al. 2000) have shown a possible regulatory switch between Lys 9 methylation and Ser 10 phosphorylation. In this combination modification example, the Ser 10 inhibits the methylation of the Lys 9; this appears to affect the H3 NH_2-terminus modification states.

Lysine methylation has been of recent focus due to the correlation with transcriptional activity. The methylation of the lysine differs from acetylation in several ways, it does not remove the charge of the lysine and the lysine may have one, two or three methyl groups. Methyltransferase enzymes that differ in the level of methylation they can achieve have been identified. The exact nature of the mono, di or trimethylated lysine is as yet undetermined, but it adds a further layer of complexity to the histone code. (Shi and Whetstine 2007)

On the core histones these include lysine residues that may be acetylated, methylated or coupled to ubiquitin; arginines residues that may be methylated; and serine residues that may be phosphorylated (Felsenfeld et al 2003); on the linker histones these include serine residues that may be phosphorylated (Dou et al 2000), (Dou et al. 2002) and amino terminal ends which are cross-linked by Poly (ADP-ribosylation) (Wong et al 1984).

1.4 Enzyme Trypsin

The catalytic mechanism of proteases is the hydrolysis of the peptide bond, between the amino acids of a protein, forming an ester and an amide (Hedstrom. 2002). There are five types of proteases whose names relate to the chemical nature of the proteolytic site: serine, cysteine, aspartic, metallo types and threonine peptidases (Rawlings and Barrett 1994) and (Rawlings et al 2006). For the purposes of this work serine proteases were used. The serine proteases are referred to as the S1 family: examples of these proteases are trypsin, elastase, chymotrypsin, and subtilisin, a bacterial protease, to name just a few of the more than 20 members. Like all serine proteases, there is a highly conserved amino acid sequence around the active site comprised of serine and histidine residues. The catalytic activity of the serine proteases is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which itself is hydrogen bonded to a serine, the hydroxyl group of this serine is essential for the catalysis. In a nucleophilic attack, the hydroxyl group of the serine attacks the carbonyl group of the protein. A fourth

feature of the active site differs from one serine protease to another, which allows for the specific activity of the protease. This is a "pocket", always located close to the active site serine. In trypsin, the pocket is deep and narrow, with a negatively charged carboxylate at its bottom that interacts with a lysine or arginine (Hedstrom 2002). Trypsin is still the most commonly used enzyme for mass spectrometric analysis because it is specific, readily available, usually gives suitable sized peptides for analysis (15-20 amino acids) and is now available in forms resistant to autolysis. Because of the specific cleavage, lysine or arginine is found at the C-terminus of the peptide, unless the peptide is the C-terminal peptide of the protein. The nature of this cleavage has made trypsin the enzyme of choice for mass spectrometry as each peptide generated from enzymatic digestion will result in a charged residue at the C-terminal end of the protein. This will give rise to a relatively large number of small peptides, which should enable greater coverage of the protein for identification purposes.

1.5 Peptide Identification

Protein identification from 1D and 2D gel analysis has become standard practice in proteomics. In these instances, the purpose of the analysis is identification of the protein using, in some cases, as few as one peptide for identification. The data collected from MS/MS analysis in these experiments are submitted to a search engine, *e.g.* Sequest (Eng et al 1994) and Mascot, (Perkins et al 1999) and by comparing the experimental to theoretical data the peptide is sequenced and related back to a protein for identification. This approach is very effective as it does not require complete protein extraction from the gel in order to identify the components. With the work presented here, it was necessary to extract as complete a representation of the protein in order to successfully identify the lysines and arginines in the peptides from the digestion which were involved in the creation of the peptide and subsequently observed by SDS/PAGE gel analysis.

The objective of the one hit kinetics was to identify the digestion products of the H5 C-terminal tail, which is believed to be the main contributor from the H5 to higher-order chromatin formation, (Allan et al. 1986). Varying salt concentrations; 5mM, 20mM and 80mM, were used to explore the effect of salt concentration on the

cleavage sites of the H5 alone in solution, in the presence of histone stripped DNA, where there should not be higher-order structure and with chromatin which had been stripped of linker histones and reconstituted with H5 to allow for the higher-order structure to form. Varying salt concentration in the H5/chromatin model would show the effects of the buffer in the formation of higher-order chromatin structures.

The digestion of H5 alone and the H5/DNA digestions are controls. Linker histones alone will not form the 30nm fiber with DNA so the sites of the C-terminal tails of the H5 which may be available for cleavage will most likely mimic those sites with H5 digested alone. It is the H5/chromatin combination which will give the information which most likely represents what occurs in the cell.

Work from (Butler 1980) and (Thomas 1980) show the effect of ionic strength of the buffer impacts the higher order structure formation as observed in the sedimentation coefficient of chromatin. The chromatin maintains the 30nm structure at 80mM where as at 20mM and 5mM the chromatin higher-order structure has altered and the fiber has a more open conformation.

1.6 SDS/PAGE Electrophoresis

The migration of proteins in gel electrophoresis works by applying an electric field to a solution and the solute molecules with a positive charge migrate towards the cathode and solute molecules with a negative charge migrate toward the anode. This migration is the electrophoresis. The velocity of the solutes is dependent on two factors: the driving motion is the force qE exerted by the electric field on the molecule, where q is the molecule's charge (in coulombs) and E is the electric field strength (in volts per meter); resisting the motion is the frictional force fv exerted on the molecule by the solvent medium, where v is the velocity of the molecule and f is the frictional coefficient, which depends on the size and shape of the molecule. Larger and less compact molecules encounter more frictional resistance than smaller more compact molecules and thus have larger frictional coefficients.

Upon application of the electric field to the solvent, the solute molecule quickly accelerates to a velocity where these forces balance and then moves steadily at this rate. This steady velocity is determined by the balance of forces:

Equation 1.1 $$fv=qE$$

This equation can be rewritten as $v/E=q/f$, we can now express the rate of motion per unit of field strength, v/E. This ratio is the electrophoretic mobility (μ) of the molecule:

Equation 1.2 $\mu = v/E = q/f = Ze/f$

The charge of the molecule is now expressed as the product of the unit of the electron (or proton) charge (e) times the number of unit charges, Z (a positive or negative integer), where f is the frictional coefficient and depends on the size and shape of the molecule.

When electrophoresis is carried out in a gel, the mobility is lower than would be expected from equation 1 because the gel exhibits a molecular sieving effect. This can be seen by graphing mobility as a function of the concentration of the gel. A graph of log μ versus percent gel is usually linear; this is called a Ferguson plot, the limiting mobility approached as percent gel approaches zero is called the free-mobility. The steepness of the Ferguson plot depends on the size and shape of the macroion, for it reflects the difficulty the macroion experiences in passing through the molecular mesh of the gel. As a result of these factors, different kinds of molecules can exhibit widely different behaviours in gel electrophoresis. However, certain simple cases are of great importance. Polyelectrolytes like DNA or a polylysine molecule have one unit charge on each residue, so each molecule has a charge (Ze) proportional to its length. But the frictional coefficient (f) also increases with molecular length, so to a first approximation, a macroion whose charge is proportional to its length has a free mobility almost independent of its size. In a mixture of such molecules, the molecular sieving effect determines the relative mobilities at any given gel concentration and the sieving effect is proportional to molecular length or molecular weight (*Biochemistry* 2000).

The gel is an acrylamide polymer of bis-acrylamide and acrylamide at a ratio of 1:29; the percentage of the acrylamide determines the sieve size of the gel. With an increase of acrylamide there is a decrease in sieve size. The decrease in acrylamide

increases the sieve size allowing the protein to move further through the gel with less hindrance.

The protein is dissolved in β-mercaptoethanol, which reduces disulfide bonds, and sodium dodecyl sulfate (SDS) a negatively charged molecule that offsets the positive charges of the protein so that the charge is proportional to the length of the protein. In the presence of SDS under these conditions, the protein loses primary, secondary and tertiary structure and the protein chain is surrounded by SDS molecules which are negative and offset the positive charges of the protein.

The interaction of the histones and DNA must determine the location and orientation of the 'tails' of the linker histones allowing them to neutralize the linker DNA and permit the nucleosomal chain to adopt higher-order folding. However, there is little evidence of how the tails are organized or how they affect this folding process. Using mass spectrometry for the identification of the trypsin digest products of histone H5 from in-solution and in-gel digestions it is the goal of this work to clarify the available information about the interface of the linker histone tail and the linker DNA.

Introduction References:

Ali, T, Thomas, JO, Journal of Molecular Biology, **337**, (5) 1123-1135 (2004) *Distinct properties of the two putative 'globular domains' of the yeast linker histone, Hho1p*

Allan, J., Hartman, P.G., Crane-Robinson, C., Aviles, F.X., Nature, **288,** (5792) 675-679 (1980) *The structure of histone H1 and its location in chromatin*

Allan, J, Mitchell, T, Harbone, N, Bohm, L, Crane-Robinson, C, Journal of Molecular Biology, **187**, (4) 591-601 (1986) *Roles of H1 domains in determining higher order chromatin structure and H1 location*

Bharath, MMS, Ramesh, S, Chandra, NR, Rao, MRS, Biochemistry, **41**, 7617-7627 (2002) *Idnetificationo f a 34 amino acid stretch within the C-terminus of histone H1 as the DNA-condensing domain by site-directed mutagenesis*

Bednar, J., Horowitz, RA., Grigoryev, SA., Carruthers, LM., Hansen, JC., Koster, AJ., Woodcock, CL. Proceeds of the National Academy of Sciences, **95,** (24) 14173-14178 (1998) *Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin*

Brown, DT, Izard, T, Misteli, T, Nature Structural and Molecular Biology, **13** (3) 250-255 (2006) *Mapping the interaction surface of linker histone H1(0) with the nucleosome of native chromatin in vivo*

Butler, P.J.G., Thomas, J.O., Journal of Molecular Biology, **140,** 505-529 (1980) *Changes in chromatin folding in solution*

Clark, DJ, Hill, CS, Martin, SR, Thomas, JO, EMBO Journal, **7**, (1) 69-75 (1988) *Alpha-helix in the carboxy-terminal domains of histones H1 and H5*

Dorigo, B. Schalch, T, Kulangara, A, Duda, S, Schroeder, RR, Richmond, TJ, Science, **306**, 1571-1573 (2004) *Nucleosome arrays reveal the two-start organization of the chromatin fiber*

Dou, Y, Gorovsky, MA, Molecular Cell, **6**, 225-231 (2000) *Phosphorylation of linker histone H1 regulates gene expression in vivo by creating a charge patch*

Dou, Y, Gorovsky, MA, Proceeds of the National Academy of Sciences USA, **99**, (9) 6142-6146 (2002) *Regulation of transcription by H1 phosphorylation in Tetrahymena is position independent and requires clustered sites*

Duggan, MM, Thomas, JO, Journal of Molecular Biology, **304**, 21-33 (2000) *Two DNA-binding sites on the globular domain of Histone H5 are required for binding to both bulk and 5 S reconstituted nucleosomes*

Eng, J.K., McCormack, A.L., Yates, J.R., Journal of the Society for Mass Spectrometry **5** (11) 976-989 (1994) *An approach to correlate tandem mass-spectral data of peptides with amino-acid sequences in a protein database*

Felsenfeld, G, Groudine, M, Nature, **421**, 448-453 (2003) *Controlling the double helix*

Finch, J.T. and Klug, A. Proceeds of the National Academy of Sciences USA, **73,** 1897-1901. (1976) *Solenoidal model for superstructure in chromatin*

Goytisolo, FA, Gerchman, S, Yu, X, Rees, C, Graziano, V, Ramakrishnan V, Thomas, JO, EMBO Journal **15**, 3421–3429 (1996) *Identification of two DNA-binding sites on the globular domain of histone H5*

Hendzel, MJ, Lever, MA, Crawford, E, Th'ng, JPH, Journal of Biological Chemistry, **279**, (19) 20028-20034 (2004) *The C-terminal domain is the primary determinant of histone H1 binding to chromatin in vivo*

Hedstrom, L. Chemical Reviews **102** (12) 4501-4524 (2002) *Serine protease mechanism and specificity*

Lever, MA, Th'ng, JPH, Sun XJ, Hendzel MJ, Nature, **408**, (6815) 873-876 (2000) *Rapid exchange of histone H1.1 on chromatin in living human cells*

Lu, X, Hansen, JC, Journal of Biological Chemistry, **279**, (10) 8701-8707 (2004) *Identification of specific functional subdomains within the linker histone H1(0) C-terminal domain*

Luger, K., Mäder, A.W., Richmond, R.K, Sargent, D.F. and Richmond, T.J., Nature, **389**, 251-260 (1997) *Crystal structure of the nucleosome core particle at 2.8 Å resolution*

Mathews, C., van Holde, K.E., Ahern, K., Addison Wesley Longman Inc. *Biochemistry* **3rd Edition** (2000)

Misteli, T, Gunjan, A, Hock, R, Bustin, M, Brown, D.T., Nature, **408**, (6814) 877-881 (2000) *Dynamic binding of histone H1 to chromatin in living cells*

Perkins, D.N., Pappin, D. J., Creasy, D.M., Cottrell, J. S. Electrophoresis **20 (18)** 3551-3567 (1999) *Probability-based protein identification by searching sequence databases using mass spectrometry data*

Rawlings, N., Morton, F., Barrett, A. Nucleic Acids Res **34** D270-D272 (2006) *MEROPS: the peptidase database http://merops.sanger.ac.uk*

Rawlings, N., Barrett, A. Methods in Enzymology **244**: 19-61 (1994) *Proteolytic enzymes: serine and cysteine peptidases*

Rea, S., Eisenhaber, F., O-Carroll, N., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., Jenuwein, T., Nature **406** (6796) 593-599 (2000) *Regulation of chromatin structure by site-site-specific histone H3 methyltransferases*

Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L. and Sweet, R.M. Nature **362** 219-223 (1993) *Crystal structure of globular domain of histone H5 and its implications for nucleosome binding*

Roque, A., Orrego, M., Ponte, I., Suau, P. Nucleic Acids Research **32** (20) 6111-6119 (2004) *The preferential binding of histone H1 to DNA scaffold-associated regions is determined by its C-terminal domain*

Roque, A, Iloro, I, Ponte, I, Arrondo, JLR, Suau, P, Journal of Biological Chemistry, **280**, (37) 32141-32147 (2005) *DNA-induced secondary structure of the carboxyl-terminal domain of histone H1*

Schalch, T, Duda, S Sargent, DF Richmond TJ Nature, **436** (4047) 138-141 (2005) X-ray structure of a tetranucleosome and its implications for the chromatin fiber

Shi, Y., Whetstine, J.R., Molecular Cell **25** (1) 1-14 (2007) *Dynamic regulation of histone lysine methylation by demethylases*

Staynov, D.V. International Journal of Biological Macromolecules , **5** (1) 3-10 (1983) *Possible nucleosome arrangements in the higher-order structure of chromatin*

Strahl, B.D., Allis, C.D., Nature **403** (6765) 41-45 (2000) *The language of covalent histone modifications*

Suzuki, M, Gerstein, M, Johnson, T, Protein Engineering, **6**, (6) 565-574 (1993) *An NMR study on the DNA-binding SPKK motif and a model for its interaction with DNA*

Thoma, F., Koller, T., Klug, A., Journal of Cell Biology **83** (2) 403-427 (1979) *Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin*

Thomas, J.O., Khabaza, A.J.A., European Journal of Biochemistry **112** (3) 501-511 (1980) *Cross-linking of histone H1 in chromatin*

Thomas, J.O., Wilson, C., EMBO Journal **5** (13) 3531-3537 (1986) *Selective Radiolabelling and identification of a strong nucleosome binding site on the globular domain of histone H5*

Thomas, J.O. Current Opinion in Cell Biology **11** (3) 312-317 (2005) *H1: Location and Role*

Th'ng, JPH, Sung, R, Ye, M, Hendzel, MJ, Journal of Biological Chemistry, **280**, (30) 27809-27814 (2005) The *H1 family histones in the nucleus-Control of binding and localization by the C-terminal domain*

Van Holde, K.E. (1989) Chromatin. Springer-Verlag: New York

Widom, J, Current Biology, **8**, (22) R788-R791 (1998) *Chromatin structure: Linking structure to function with histone H1*

Wolfe, A. (1998) Chromatin Structure and Function third edition. Academic Press London

Wong, M, Allan, J, Smulson, M, The Journal of Biological Chemistry, **259**, (12) (1984) *The mechanism of histone H1 cross-linking by poly(ADP-ribosylation) reconstitutin with peptide domains*

Woodcock, CL, Grigoryev, SA, Horowitz, RA, Whitaker N, Proceedings of the National Academy of Sciences of the United States of America **90** (19): 9021-9025 (1993) *A Chromatin folding model that incorporates linker variability generates fibers resembling the native structures*

Zhou, YB, Gerchman, SE, Ramakrishnan, V, Travers, A, Muyldermans, S, Nature, **365**, (6700) 402-405 (1998) *Position and orientation of the globular domain of linker histone H5 on the nucleosome*

Zlantanova, J, Caiafa, P, van Holde, K, FASEB Journal, **14**, 1697-1704 (2000) *Linker histone binding and displacement: versatile mechanism for transcriptional regulation*

2. Mass Spectrometry

2.1 Mass spectrometry: a brief history

A series of events led to the development of what we now call the mass spectrometer. It was the physicist Eugene Goldstein who, in 1886, first discovered the existence of positively charged particles, generated from gaseous samples as they were exposed to a perforated cathode. Following this in 1898 Wilhelm Wien demonstrated that the rays of positive ions could be deflected in the presence of strong electric and magnetic fields, and be differentiated by their charge and mass, and that different ionic species could be detected simultaneously. In 1905 John Joseph Thomson developed the discoveries of Goldstein and Wien to construct an instrument that could ionise any gas and, using strong electric and magnetic fields, separate the ions by their charge and mass. The ions formed parabolic beams, which were focussed onto photographic plates to record the mass spectrum. The equations of motion that defined the parabolic path of the ions were expressed in terms of the charge-to-mass ratio, *e/m* of the ion (where *e* denotes the charge of an electron and *m* is the mass of the particle).

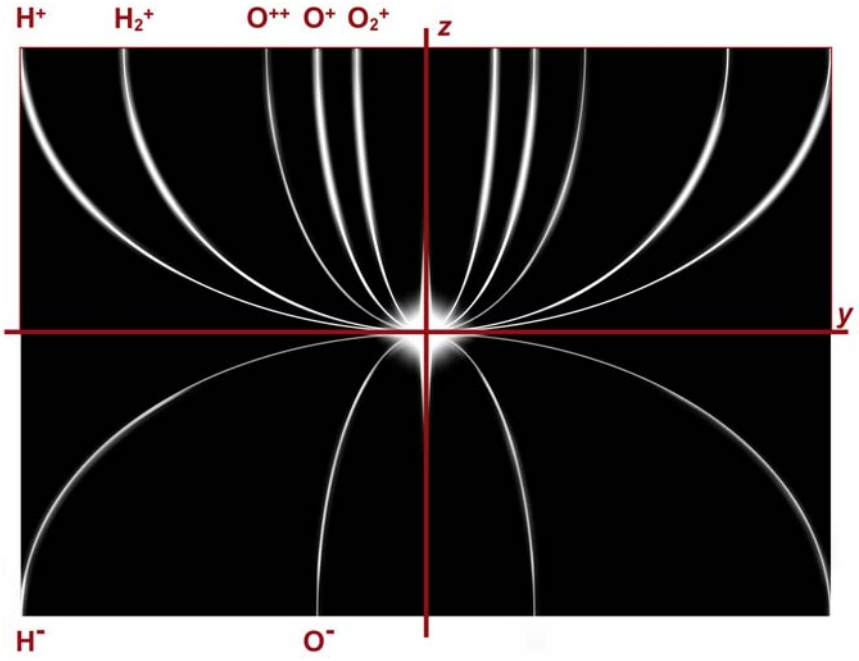

Figure 2.1: A theoretical reproduction of a parabola mass spectrum as recorded for hydrogen and oxygen on a photographic plate. This diagram has been adapted from Thomson (1913).

Within a few years of its creation, the mass spectrometer was instrumental in the discovery and characterization of the stable isotopes of the elements and was soon finding uses for organic identification and structural elucidation in the petrochemical and fine chemicals industries. More recently the mass spectrometer has found a major application as one of the most important analytical tools for the investigation of complex biological systems.

2.2 The mass spectrometer: a general example

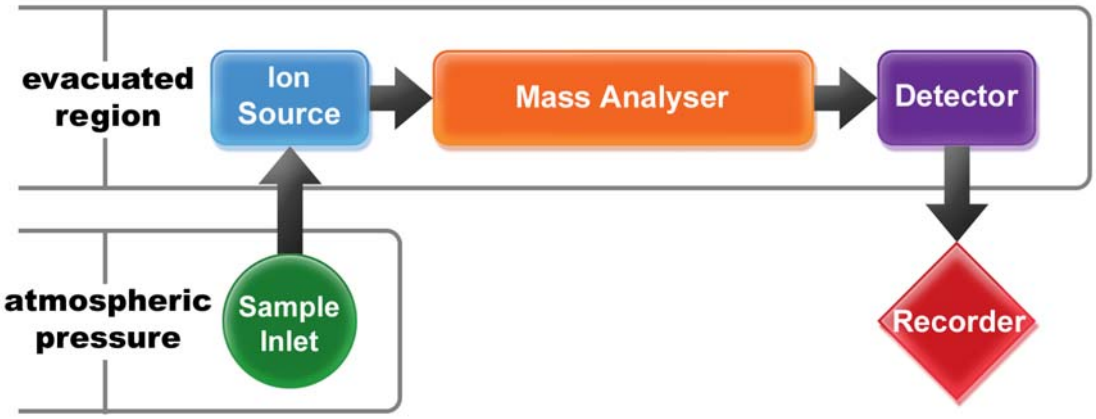

Figure 2.2: A block diagram representing the essential components of every mass spectrometer.

Each mass spectrometer can be separated into five distinct parts:

Sample inlet: Transfers the sample (solid, liquid or gaseous) from the atmospheric pressure of the laboratory to the ion source. The sample inlet may vaporize the sample into the gas phase before entry into the ion source.

Ion source: Converts the sample into gas phase ions, provides the ions with the kinetic energy of either a continuous ion beam or a tight ion packet. Neutral sample molecules are ionized and then accelerated into the mass analyzer.

Mass analyzer: The section separates ions, either in space or in time, according to their mass to charge ration.

Detector: On reaching the detector, in this instance an electron multiplier, the ions strike the surface causing a gradual cascade of electrons to produce a current that can be measured and then amplified. There are some detectors, which can detect and then

amplify the minute current generated from the motion of the ions past a system of electrodes. In all instances, the amplified signal is sent to a data recorder.

Recorder: The output analog signal from the detector is converted to a digital signal. It is here the mass spectrum of a sample is observed as a series of sample ions with different *m/z* values.

Vacuum systems: In order to determine the *m/z* values of the ions generated from sample molecules, it is necessary for the ions to travel through the analyzer without interaction with other molecules or ions; this requires the analyzer to be kept under vacuum. There are often stages of vacuum in the instrument with the lowest pressure being required in the detector region. The vacuum maintained within the instrument can vary with the requirements of each instrument, but is typically in the region of sever millionths to one billionth of atmospheric pressure (10^{-6}-10^{-10} mbar).

The route travelled by the ions, along the length of the mass analyze, is by convention referred to as the *z*-axis. Many mass spectrometers are liear, in that the ion beam (or ion packet) travels along the *z*-axis from the ion source to the detector and all of the above-described components act in parallel to the *z*-axis. There are some instrument that are not linear, and all those components that do not act in the parallel to the *z*-axis are termed to be *off-axis.* Thus the motion of the ions does not absolutely define the *z*-axis of the instrument.

2.3 The units of mass spectrometry

The units used to indicate the mass of th analyte were commonly shown as *m/z*, indicating a mass-to-charge. This results from the manner by which the spectrum is presented, along the *x*-axis of the spectrum the scale is mass-to-charge, denoted as *m/z*. On the *y*-axis of the spectrum units are denoted as relative *% intensity* or *% relative intensity* or *% intensity* of signal generated by the ions.

The abundance of ions detected translates into a spectrum where the *m/z* is plotted against the intensity of the signal generated by the abundance of ions detected.

It is common to find the mass-to-charge ratio taken to be the absolute value of the mass of an analyte species, i.e. equal to *u*, but this neglects the possibility of multiply charged species. An ion in a spectrum at a mass-to-charge value of 44 could be interpreted as being due to singly charge carbon dioxide, CO_2^+ (mass-to-charge = 44/1) or doubly charged pentan-1-ol, $[CH_3CH_2CH_2CH_2CH_2OH]^{2+}$ (mass-to-charge =

88/2 = 44) or indeed singly charged ethylene oxide. Another common unit is the Dalton, *Da*, named after the British chemist John Dalton, and is often used to represent a mass-to-charge unit, and is equal to one gram per mole (also the value of *u*). Once again this is only really equal to mass. To remove any ambiguity it is best to quote the mass-to-charge value, which is also denoted by the Thomson, *Th*, named after the founding figure of mass spectrometry. Thus the doubly charged pentan-1-ol is 44 Th and singly charged benzene is 78 Th.

2.4 Sample introduction and ionisation techniques

Inlet:

In the case of the work performed here, the sample was introduced into the Applied Biosystems MALDI-ToF DE-STR (Matrix-assisted-laser-desorption-ionization, time-of-flight, delayed-extraction) instrument on a gold plated MALDI target plate. For the H5 validation analysis, the analysis was done using a Micromass (now Waters) Q-ToF (hybrid quadrupole – time-of-flight) using electrospray ionisation.

Matrix-Assisted Laser Desorption Ionisation: MALDI

Matrix-assisted laser desorption ionisation (MALDI) was first introduced in 1988 by Hillenkamp and Karas and it has since become a common tool for the analysis of proteins and peptides. In MALDI the analyte is mixed with an excess of matrix, usually an ultraviolet-absorbing weak organic acid and it is this intimate mixture of analyte and matrix that is essential to the MALDI process. The important aspect of the mixture of analyte and matrix is the co-crystallization. The formation of ‘good’ crystals enhances the ionisation process. The matrix acts as an intermediate in energy transfer from the laser beam to the sample. The absorbing matrix can transform the laser energy into excitation energy for the solid sample thereby generating sample ions that move through the time-of-flight analyzer. The proximity of the sample to the matrix could be what is required for the exchange of the proton for ionisation. A matrix is chosen for the ability to absorb the energy from the laser, in this case a nitrogen laser which emits energy at 337nm. The proposed method for ionization in MALDI requires the matrix to absorb the energy of the laser pulse and that energy is transferred to the sample creating the ion.

The MALDI matrix absorbs the laser light energy and it vaporises causing the analyte to vaporise thus entering the gas phase. The matrix can also act as a proton donor and receptor, giving a charge to the analyte.

Electrospray ionization (ESI)

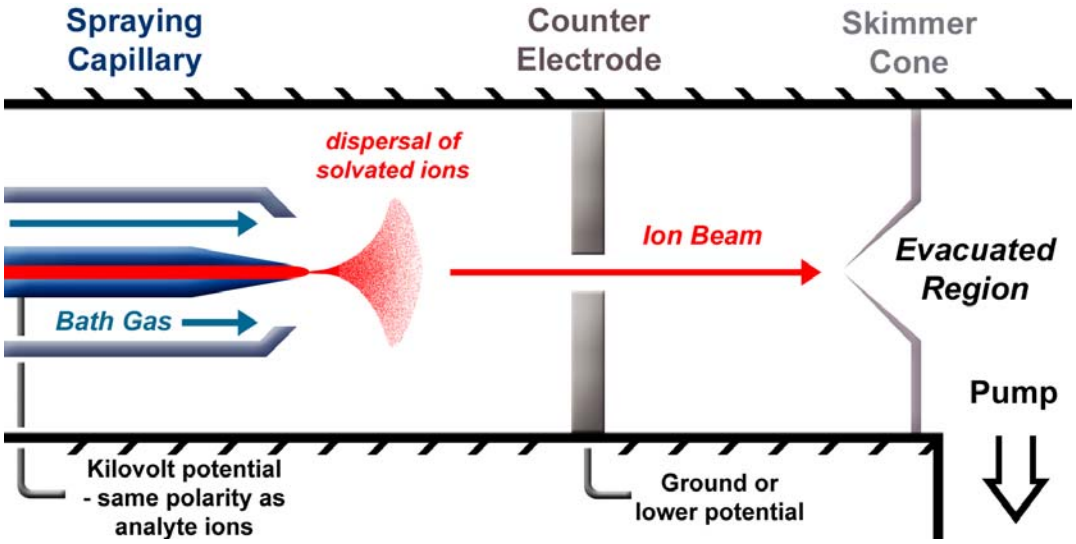

Figure 2.3: A general example of one type of electrospray source where the electric field is generated by the direct application of an electric potential to the spraying capillary. The skimmer cone can also have a potential applied to collimate the ion beam. Diagram adapted from (Chapman 1998).

Electrospray was developed in 1989 (Fenn et al 1989) electrospray and like MALDI, allowed for the analysis of high molecular weight compounds. There are three major steps in the production by electrospray of gas-phase ions from electrolyte ions in solution: (1) production of charged droplets at the electrospray capillary tip; (2) shrinkage of the charged droplets by solvent evaporation and repeated droplet disintegration, leading ultimately to very small highly charged droplets capable of producing gas-phase ions, and (3) the final production of gas-phase ions from the very small and highly charged droplets.

Unlike MALDI, electrospray creates multiply charged analytes, $(M + H)^+$ the singly charged species of MALDI and $(M + nH)^{n+}$ the multiply (n = number of protons added to analyte generating n number of charges).

2.5 Analysers

Analyser:

A time-of-flight analyser was used for this work. With this type of analyzer the ions leave the source region and travel through a field free region (there are no applied magnetic or electric fields in this area) to the detector and the time it takes to travel that known distance is specific to each mass to charge ratio. The ions are separated in time according to their mass to charge ratio.

Time-of-Flight Mass Spectrometer

The time-of-flight mass spectrometer (ToF-MS) is perhaps the most straightforward design of mass spectrometer. Rather than measure the mass to charge of an ion, the ToF-MS measures the momentum of the ion. The principle behind such measurement is that the ions, with a distribution of mass and charge that travel, as a packet, in the same direction with constant kinetic energy will have a corresponding distribution of velocities, where the velocity is inversely proportional to the square root of the mass-to-charge ratio Guilhaus (1995).

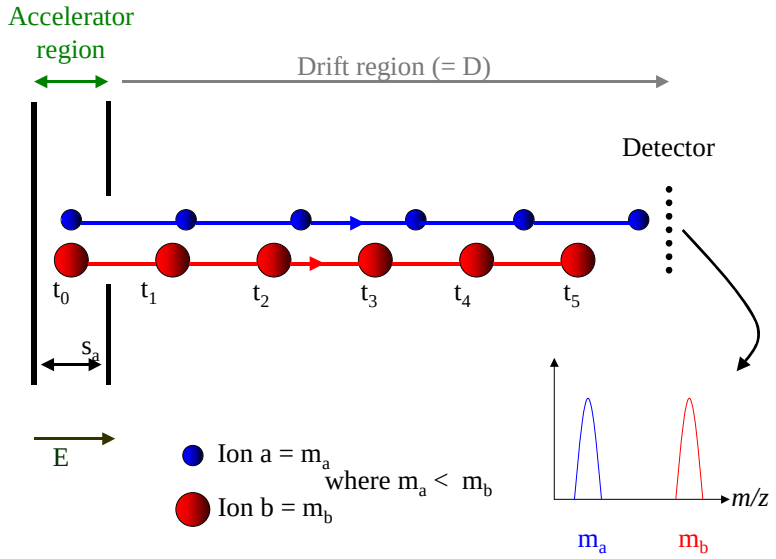

Figure 2.4: A general schematic of a linear time-of-flight mass spectrometer, showing the motion of ions *a* and *b* through the instrument. Note that both ions are singly charged. The term s_a represents the distance travelled within the accelerating region; E the vector of the electric field and D the length of the drift region. The terms $t_{1\text{-}5}$ relates to evenly spaced snapshots in time. Thus the motion of the ions is shown for the progression in time. It can be seen that the lighter ion *a* has a greater velocity and will reach the detector before the heavier ion *b*.

A general schematic of a linear time-of-flight instrument is shown in figure 2.4 analysing two ionic species, *a* and *b*, each carrying unit charge, but having different masses, m_a and m_b respectively. In this scheme, $m_a < m_b$, the mass-to-charge value of the ions can be taken as their mass. Ions of the same mass-to-charge value form discrete ion packets temporally separated from all other ions. The motion of the ions *a* and *b* as they travel through the instrument is shown in figure 2.4 at evenly spaced points in time, denoted from t_1 through to t_5. The mass spectrum is then constructed from the arrival times of the ion packets at the detector. As such the entire procedure of analysis is usually pulsed. The potentials are applied to the accelerating electrodes at time $t = t_o$ for a given length of time, normally of the order of microseconds, before being switched off again. The observed time-of-flight is thus recorded respectively of $t = t_o$. The record of ion abundance against time is then converted into a mass spectrum.

The overall goal is to determine the total time-of-flight of the ion, from the beginning of the pulse at $t = t_o$, to detection.

An ion has charge $q = z.e$, where z is the number of positive or negative charges on the ion, and e is the magnitude of the charge of the electron. The ion is placed between two electrodes at a point denoted as s_o. Potentials applied to the electrodes generate a potential difference, ΔV, which creates an electric field. The electric field strength, E, is equal to the potential difference between the electrodes, divided by the distance travelled by the ion under the influence of the field, denoted by s_a.

Equation 2.1 $$E = \frac{\Delta V}{s_a}$$

The potential difference defines the kinetic energy of the ion, and thus its maximum velocity. Thus the ion is accelerated from an initial starting velocity, u_o, to the maximum allowable velocity before the ion enters the field-free drift region. Therefore the ions travel though the drift region with the maximum possible velocity, which should theoretically remain constant. This is defined as the *drift velocity*, u_D, such that:

Equation 2.2 $$KE = q\Delta V = qEs_a = \tfrac{1}{2} mu_D^2$$

$$u_D = \sqrt{\frac{2qEs_a}{m}}$$

The force, F, experienced by the ion in the presence of an electric field, E, is equal to:

Equation 2.3 $$F = Eq$$

This force accelerates the ion in the direction of the electric field. The acceleration can be determined from Newton's second law of motion:

Equation 2.4 $$F = ma \quad \therefore \quad a = \frac{Eq}{m}$$

The acceleration must be determined so that one can calculate the time spent by the ion accelerating to the drift velocity. The acceleration is defined as the differential of the velocity with respect to time:

Equation 2.5 $$a = \frac{\mathrm{d}u}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{Eq}{m}\right)$$

$$\therefore \quad u = \int\left(\frac{Eq}{m}\right)\mathrm{d}t$$

$$\therefore \quad u = u_o + \left(\frac{Eq}{m}\right)t$$

This equation determines the velocity of the ion at any time, t, under constant acceleration. Since the kinetic energy is fixed by the one electric field the ion can only accelerate to its maximum drift velocity. The time required for acceleration from u_o to u_D can be calculated by:

Equation 2.6 $$t_a = \frac{u_D - u_o}{E}\left(\frac{m}{q}\right)$$

If one can assume that the ions are initially at rest before acceleration, then $u_o = 0$ and the acceleration time is readily determined purely from the electrode parameters and the mass and charge of the ion. For completeness it is often assumed that the ions take a finite time to respond to the action of the force, t_o, from $t = 0$ when the potential is applied to the electrodes. The determination of this value is possible but is often overlooked due to its negligible effect to the total flight time of the ion.

The time taken to traverse the drift region, t_D, is simply the length of the drift region, D, divided by the drift velocity:

Equation 2.7

$$t_D = \frac{D}{u_D}$$

$$t_D = \left(\frac{m}{2qEs_a}\right)^{1/2} D \quad = \quad \left(\frac{m}{2q\Delta V}\right)^{1/2} D$$

Finally, one may wish to also include the response time of the detector, t_d although again this is likely to have a negligible effect upon the overall flight time, and its inclusion is dependent upon the desired accuracy of the research.

Therefore the time-of-flight, TOF, is equal to (Guilhaus1995):

Equation 2.8

$$TOF = t_o + t_a + t_D + t_d$$

$$= t_o + \left[\frac{u_D - u_o}{E}\left(\frac{m}{q}\right)\right] + \left[\left(\frac{m}{2q\Delta V}\right)^{1/2} D\right] + t_d$$

Whereas the ion velocity is inversely proportional to the square root of the mass-to-charge ratio, the dependence of the flight time on the mass-to-charge ratio is more complicated. In most time-of-flight instruments the drift time, t_D, is far greater than the acceleration time, t_a. Indeed the drift time is often far greater than the sum of the acceleration time, the response time, t_o, and the detection time t_d, meaning the relationship for the drift time in equation 2.7 is a good approximation to the flight time of the ions.

Delayed extraction (Vestal et al 1995), samples are ionised, accelerate into the instrument, and pause at a 'grid' allowing for the ion packet to come together and depart at the same time from the source region. The grid, in effect a wire mesh, is located 1-2 mm above the target plate and has a voltage applied to it which allows for the fine tuning of ion acceleration. The grid voltage is set as a % of the voltage applied to the sample plate as the mass of the compound increases the % of the sample plate voltage applied to the grid is decreased.

Quadrupole Mass Spectrometer

A quadrupole mass analyser is an example of a scanning mass analyser. The range of mass-to-charge values analysed is selected before the analysis begins. In 1953 the German physicists Wolfgang Paul and Hans Dehmelt published their research into the use of quadrupolar electrical fields for the separation of ions for mass spectrometry (Paul et al 1953) and (Paul et al 1958) .The magnitude of the band-

pass, Δm/z, can also be controlled, and remains constant throughout the analysis procedure. The band-pass is gradually moved between the upper and lower limits of the experimental window and only those ions that fit within the confines of the band-pass will be detected, while all other ions are lost. A constant flow of ions is necessary for a mass spectrum to be recorded, a fact that reflects the popularity of interfacing electrospray ionisation sources to quadrupole mass analysers. In addition, the quadrupole can only operate with ions injected into the array at relatively low speeds, normally achieved by an accelerating voltage of 10-20 V (Chapman 1998) Again the electrospray ionisation source can fulfil this criterion since the temperature of the bath gas controls the velocity of the ions, and can easily maintain the ions at the necessary low kinetic energies prior to the final evacuated transfer region directly before the analyser. Here another R.F. only multipole, typically a hexapole, will accumulate the ions into a coherent packet for sending into the quadrupole analyser.

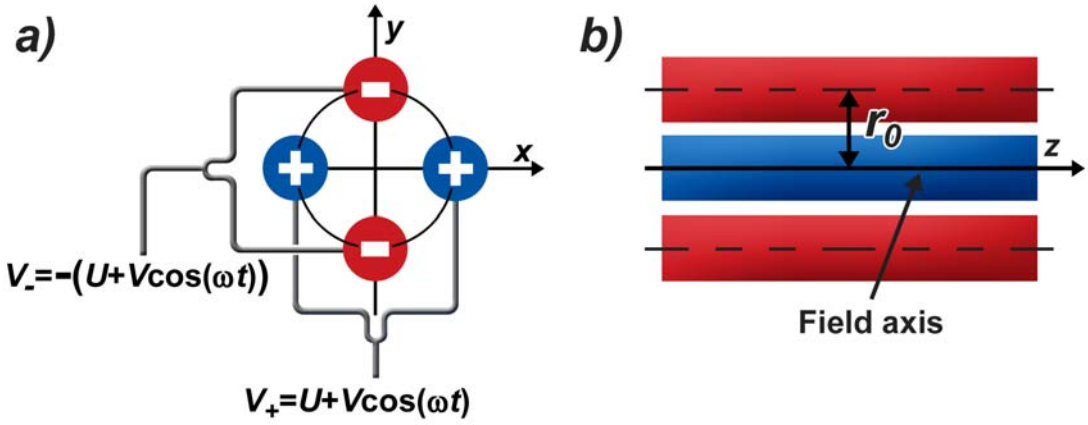

Figure 2.5: A general schematic of a linear quadrupole mass analyser with rods of circular cross-section. a) The arrangement of rods and the respective potentials applied to the vertical and horizontal pair of electrodes are shown. b) The separation between the centre of the rods and the *z*-axis, r_o can also be thought of as the radius of curvature of the quadrupole. Diagram adapted from (Chapman 1998).

In a standard quadrupole the mass spectrum is generated by maintaining a constant r.f. (radio frequency) frequency, ω, and varying both the magnitude of the constant potential, U, and the maximum magnitude of the r.f. potential, V_o simultaneously while keeping the ratio of U/V_o constant (Chapman 1995) the recorded mass-to-charge values are proportional to V_o, such that a linear increase in this parameter provides an easily calibrated linear mass-to-charge scale. It is possible to generate a

mass spectrum by varying the r.f. frequency while keeping both U and V_o constant, but this option is not particularly convenient to perform.

The voltage supplied to the electrode rods has two components:

1. A direct current (d.c.) component, U,
2. A radio-frequency (r.f.) component $Vcos(\omega t)$. This voltage component has a time dependent sinusoidal form, with amplitude that varies between zero and the maximum voltage, denoted by V_o. ω is the angular frequency of this component, such that the frequency is equal to $\omega/2\pi$.

The electrodes can be split into horizontal and vertical pairs. The potential applied to the horizontal pair on the x-axis is (Chapman 1995) and (Hofstadler et al 1996):

Equation 2.9 $$V_+ = U + V\cos(\omega t)$$

Similarly, the potential applied to the vertical pair of electrodes on the y-axis is:

Equation 2.10 $$V_- = -U + V\cos(\omega t + \pi)$$

$$\therefore \quad V_- = -[U + V\cos(\omega t)]$$

As can be seen from these equations, and figure 2.6, opposite sets of electrodes have the same charge, but are biased with d.c. voltages of opposite polarity, and the r.f. components are out of phase by 180^o (Hofstadler et al 1996).

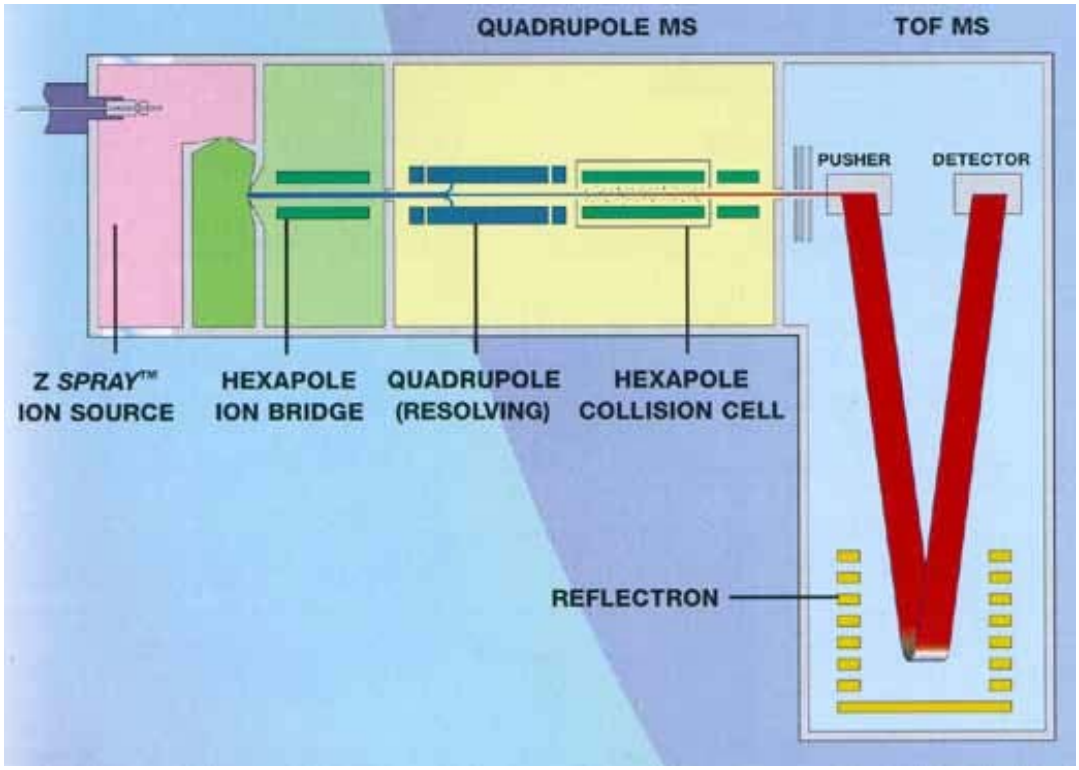

Figure 2.6: A schematic diagram of the hybrid quadrupole (MS1) time-of-flight (MS2) mass spectrometer, Schematic diagram courtesy of Waters Corp.

The Q-ToF (Morris et al 1996) uses first a quadrupole for mass transmission/filter and the time-of-flight for mass analysis (see figure 2.6). The analysis of the H5 as a complete protein was conducted by direct infusion on the Q-ToF, a quadrupole-time-of-flight hybrid mass spectrometer.

2.6 Resolution

Resolution as defined by the International Union of Pure and Applied Chemistry (IUPAC 1997) defines the mass resolving power of a mass spectrometer as: "The ability to distinguish between ions differing in the quotient mass/charge by a small increment."

The resolution of a mass spectrometer can therefore be regarded as the capacity of the system to separate ions of adjacent mass-to-charge ratios, although no standard definition of that separation has been formally adopted by the mass spectrometry community (Chapman 1998). At the most basic level, the resolution, *R*, required to separate two ions of mass-to-charge values m_1 and m_2, where $m_2 = m_1+\Delta m$, as shown in Figure 2.7 is given by:

Equation 2.11
$$R = \frac{m_1}{\Delta m}$$

A large resolution value corresponds to a greater resolving power of the mass spectrometer. The resolution can also be expressed in terms of *parts per million,* leading to the equation

Equation 2.12
$$R(ppm) = 10^6 \frac{\Delta m}{m_1}$$

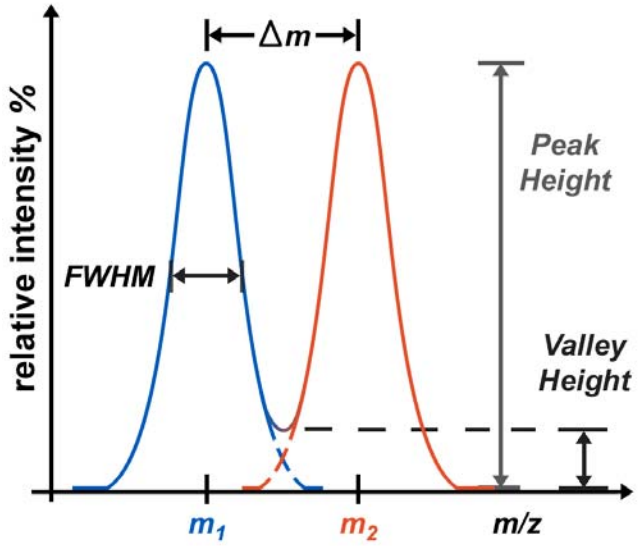

Figure 2.7: A mass spectrum showing two peaks of equal intensity for ions of mass-to-charge values m_1 and m_2. The various parameters for determining the resolution of the spectrum, such as the valley height and the full-width at half-maximum height (FWHM) of m_1 are shown. Diagram adapted from Chapman (1998).

When dealing with parts per million, the lower the value of *R*, the greater the resolving power of the mass spectrometer.

A complication arises in the determining the resolution in that the operating principles of some types of mass analysers force the definition of resolution for that instrument. As an example a quadrupole mass analyser works by maintaining the value of mass separation, Δm, throughout the mass range. As such, if the mass separation is held at m/z 1 Th, then the resolution of a peak at *m/z* 2000 Th will be 2000, and a peak at *m/z* 200 Th will have a resolution of 200. The constant mass separation represents the limiting factor in the resolution of quadrupole analysers. It can be seen that the resolution of a mass spectrometer is dependent upon the mass-to-charge value of the species being measured and the method by which it is being measured. It is valid to say that mass spectrometers should only be classed in terms of their overall resolving abilities, rather than any quantitative value, unless those values are quoted from the analysis of example species.

The method of operation within the quadrupole mass analyser means that resolution can be measured from just one peak. This can be done from the mass spectrum by measuring the width of a peak at half of the height of that peak, known as the Full Width at Half Maximum, of FWHM.

Resolution in time-of-flight mass spectrometry

In time-of-flight mass spectrometry it is just as valid to measure the resolution from the time domain as it is from the mass domain (Guilhaus 1995). Given that the drift time is the dominant factor in the time-of-flight, it can be seen from Equation 2.7 that the mass is directly proportional to the square of the time, thus we it is possible to can define:

$$m = \mathrm{A}t^2 \qquad \therefore \qquad \frac{\mathrm{d}\,m}{\mathrm{d}\,t} = 2\,\mathrm{A}\,t$$

Equation 2.13

where A is a constant of proportionality. It then follows that

$$\Rightarrow \qquad \frac{\mathrm{d}m}{m} = \frac{2\,\mathrm{d}\,t}{t} \qquad \therefore \qquad \frac{m}{\Delta m} = \frac{t}{2\Delta t}$$

Equation 2.14

Both the time interval Δt and the mass interval Δm are measured as the full-width at half maximum of the peak.

The resolution is dependent upon the flight time of the ions, which in itself is dependent upon the geometry of the electrodes, the surrounding volume of the source, and most importantly the kinetic energy of the ions (Chapman 1998) and (Guilhaus 1995). Ideally, each ion of a given type would be accelerated to the exact same kinetic energy and reach the detector at exactly the same time as a very tight ion packet. However, this does not always occur, and the ions are frequently given a range of kinetic energies, producing broad ion packets. The size of the ion packet is important for the resolution of the mass spectrum. A broad ion packet will strike the detector over a greater interval of time than a tight ion packet. The longer this time interval is the broader the corresponding peak appears on the mass spectrum.

Methods References:

Allan, J., Hartman, P.G., Crane-Robinson, C., Aviles, F.X., Nature **288** (5792) 675-679 (1980) *The structure of histone H1 and its location in chromatin*

Gidley, M.J., Sanders, J.K.M., Biochemistry Journal **203** 331-334 (1982) *Reductive methylation of proteins with sodium cyanoborohydride: identification, suppression and possible uses of N-cyanomethyl by products*

Gould, H. Oxford University Press *Chromatin: A Practical Approach* (1998)

Harborne, N., Allan, J., FEBS Letters **155** (1) 88-92 (1983) *Modulation of the relative trypsin sensitivities of the core histone tails*

Maman, J.D., Yager, T.D., Allan, J., Biochemistry **33** (6) 1300-1310 (1994) *Self-association of the globular domain of histone H5*

Jentoft, N., Dearborn, D.G., Journal of Biological Chemistry **254** (11) 4359-4365 (1979) *Labeling of proteins by reductive methylation using sodium cyanoborohydride*

Lambert, S.F., Thomas, J.O., European Journal of Biochemistry **160** 191-201 (1986) *Lysine-containing DNA-binding regions on the surface of the histone ocatmer in the nucleosome core particle*

MacGillivray, A.J., Birnie, GD., Buttersworth *Nuclear Structures: Isolation and Characterization*. (1986)

Taylor, I.A., Webb, M., Humana Press Inc. **2nd edition** Methods in Molecular Biology **148** DNA-Protein interactions: Principles and Protocols *Chemical modification of lysine by reductive methylation*

Thomas, J.O., Methods in Enzymology **170** 369-385 (1989) *Chemical radiolabeling of lysines that interact strongly with DNA in chromatin*

Thomas, J.O. Wilson, C.M., EMBO Journal **5** (13) 3531-3537 (1986) *Selective radiolabelling and identification of a strong nucleosome binding-site on the globular domain of histone H5*

3. Materials and Methods

This chapter includes the general methods and materials used for the preparation and characterization of materials that have been used to generate the work included in this thesis.

3.1 Stock solutions

PMSF (Phenylmethylsulfonyl fluoride) was prepared as 250mM stock in isopropanol and stored at 37ºC. PMSF was added to each buffer just prior to use.

EDTA (ethylenediamine tetraacetic acid) was dissolved in water sodium hydroxide (NaOH) tablets were added to adjust the pH to 8.0, at which point the EDTA is soluble, and adjusted to a final concentration of 0.5M..

Tris (Tris (hydroxymethyl) methylamine) was dissolved in water, the pH adjusted with hydrochloric acid (HCl) to pH7.5 and adjusted to a final volume for a final concentration of 2M.

TE (Tris, EDTA) **Stock** 200mM Tris pH7.5, 2mM EDTA. PMSF was added just prior to use. This is a 10X stock solution

TEP $_{5,\ 20,\ 80}$ 10mM Tris pH7.5, 0.1mM EDTA, 0.2mM PMSF, with 5mM NaCl (TEP_5), 20mM NaCl (TEP_{20}), 80mM NaCl (TEP_{80})

Sinapinic Acid (3,5-Dimethoxy-4-hydroxycinnamic acid) Sinapinic acid was added in excess, ~1g, to a 1:1 mixture of H_2O/Acetonitrile (ACN), 0.1% formic acid.

Phosphate Buffer 35.49g of disodium hydrogenphosphate and 29g of sodium dihydrogen phosphate were combined with 500ml of water. This gave a final concentration of 1M.

10mM PO_4 0.65M NaCl Buffer 10mM PO_4, 0.1mM EDTA, 0.2mM PMSF, 0.65M NaCl.

DNA Gel Loading Buffer 6x loading dye purchased from MBI Fermentas.

Micrococcal nuclease digestion buffer (MN buffer) 5mM Tris pH, 1mM $CaCl_2$, 80mM NaCl, 0.1mM PMSF, 0.2M sucrose. The micrococcal nuclease was available in the laboratory at prescribed concentration of 10,000 units/μL.

Trypsin 1mg/ml of trypsin in TEP was diluted to 100μg/ml in TEP for daily use.

Linker Histone H5 stock 19mg/ml (in water/PMSF) H5 from prepared protein available in Dr. Jim Allan's laboratory.

3.2 Nuclei Preparation

Buffer A: 0.25M Sucrose, 6mM $MgCl_2$, 50mM Tris pH 7.5, 0.5mM EGTA (Ethylene glycol-bis[β-aminoethyl ether]-N,N,N',N'-tetraacetic acid), 0.2mM PMSF

Buffer A + 0.4% NP40 (Nonidet P40) NP40, a detergent, was added w/v to Buffer A. The addition of the NP40 is for cell lysis.

Buffer A + 0.2% NP40 NP40 added to Buffer A w/v.

PBS (Phosphate Buffered Saline) + **Heparin** (used to prevent clotting of the blood) PBS was made by dissolving one PBS tablet in 200ml of water (to obtain: 0.01M phosphate buffer, 0.0027M potassium chloride and 0.137M sodium chloride at pH7.4 at 5000units/ml). 25,000 units of heparin were added to the PBS to prevent coagulation of the blood. (Heparin was purchased from Sigma at 50,000 units, 1ml of water was added to the heparin and 500µl were added to the PBS).

Nuclei Storage Buffer: 25% glycerol, 50mM NaCl, 50mM Tris pH 7.5, 5mM $MgCl_2$, 0.1mM EGTA, 0.1mM PMSF

Approximately 750ml of fresh chicken blood was mixed with an equal volume of PBS + heparin and filtered through 2 layers of PBS + heparin soaked gauze. The chicken blood/PBS solution was then transferred to 400ml plastic flasks and centrifuged at 3,500 rpm for 5 minutes on a centrifuge (Sorval RC-5B). The supernatant and white blood cells, the buffy coat, was removed and the red blood cells were resuspended an additional two times in PBS, containing no heparin, and centrifuged both times as above. The red blood cells form a pellet upon centrifugation, the pellets were resuspended sequentially in 500ml Buffer A and 500ml Buffer A + 0.4% NP40, centrifuged as above, the supernatant removed and the pellets resuspended in 500ml Buffer A + 0.2% NP40. The sample was centrifuged and the pellets were washed twice with Buffer A. The pellets were then resuspended in nuclei storage buffer at 10mg/ml final concentration of DNA and stored in 5ml aliquots at -70°C.

3.3 Determination of nuclei concentration

50µl of nuclei was resuspended in 5M Urea, 2M NaCl to a final volume of 1ml. This solution was sonicated for one minute, centrifuged at 14,000 rpm for one minute and the optical density (OD) read, at a wavelength (λ) of 260 nm,.

The extinction coefficient for a substrate is calculated by reading the optical density at a specified wavelength with the concentration of substrate at 1mg/ml in a 1cm cell., for 1mg/ml of DNA, the OD_{260} = 20 units so 1 OD_{260} unit/ml = 50µg/ml.

3.4 Nuclei Digestion

In order to carry out experiments on the structure of the 30nm fibre, chromatin of approximately 4,000 base pairs (bp) in length was required. Test digestions were conducted to determine the appropriate conditions required to achieve chromatin of the required length; 500µl of 10mg/ml nuclei were initially digested with micrococcal nuclease at 5units/ml (stock 10,000units/µl) for designated time points allowing an appropriate length of time for digestion to be determined.

Once an appropriate digestion time was determined, the nuclei were washed twice in micrococcal nuclease (MN) buffer and resuspended in the same volume of MN buffer as the initial volume, 500µl. The sample was equilibrated to 37ºC in a water bath and digested by the addition of micrococcal nuclease. The digest was stopped by the addition of 200µl of 0.5M EDTA and the sample was placed on ice for 30 minutes. After centrifugation, at 4,000 rpm for 15 minutes, the pellet was resuspended in TEP_{20} and stored at 4ºC overnight to allow lysis to occur. The sample was then centrifuged, conditions as above, and the soluble chromatin in the supernatant was removed. A 1ml aliquot was removed; the DNA was precipitated and run on a 0.8% agarose gel, see figure 3.1, the to determine the time required to generate the required length of chromatin for remaining digestions. An appropriate digestion time was determined and 5ml of nuclei were digested as in above conditions. The supernatant was removed following lysis and run on a sucrose gradient to fractionate the chromatin, (MacGillivray 1986)

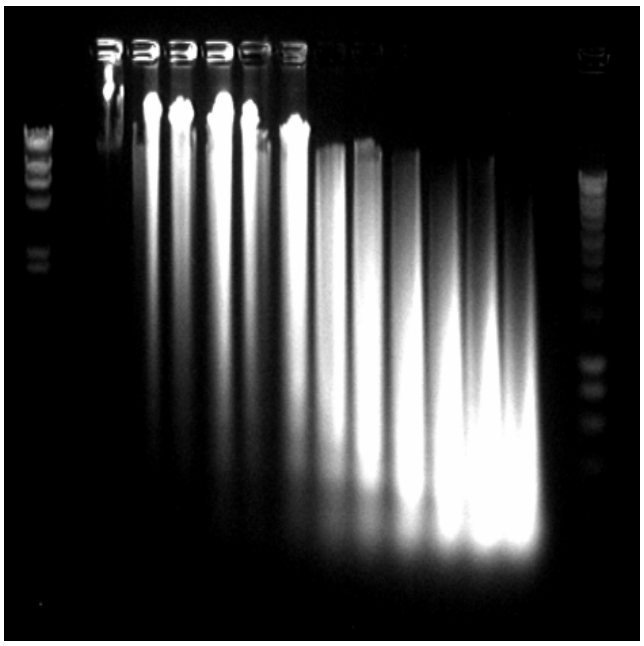

Figure 3.1 0.8% agarose gel of DNA test digestion to determine time for digestion to generate DNA with length ~4,000bp. Reading from left to right, the first lane is a DNA marker, the following 12 lanes are time points 0 to 60 minutes followed again by DNA marker (ladder of DNA from 10,000bp to 250 bp)

3.5 Sucrose Gradient

Chromatin was fractionated on the basis of size by sedimentation through sucrose gradients (Gould 1998). Two concentrations of sucrose (12% and 45%) were added to a mixer forming the gradient. The mixer has two tubes, each with one of the sucrose concentrations and a magnetic stir bar. A valve between the two is open and the tubing runs through a pump which pulls the sucrose through the valve from the two tubes to fill a centrifuge tube to within ~.8cm from the top. The gradient is then allowed to set overnight at 4°C. The following day the chromatin is gently laid onto the top, so that the gradient remains undisturbed. The tubes are balanced by weight and centrifuged at 27,000 rpm for 2.5 hour. Upon completion of centrifugation, the bottom of the centrifuge tube is punctured and 60% sucrose is pumped into the bottom and the gradient in the tube is forced out through a UV detector (see figure 3.2) for monitoring the gradient as it is pumped out of centrifuge tube and the 1.5 ml fractions are collected. Aliquots of the fractions were ethanol precipated and sized on agarose gel (see figure 3.3). Fractions from three separate gradients were pooled and a small aliquot taken to prepare DNA for sizing on agarose gel in order to confirm

the fractions with discrete lengths of approximately 4,000bp (see figure 3.3). The chromatin fractions were stored at –20°C in sucrose.

3.6 Phenol/Chloroform Extraction, DNA Precipitation

DNA was purified from chromatin fractions by extraction with an equal volume of buffered phenol. After centrifugation the aqueous layer was removed and extracted with an equal volume of 24:1 chloroform: isoamyl alcohol. The aqueous layer was removed and made 0.3 M in sodium acetate before the DNA was precipitated with 2 volumes of 100% ethanol and storage at -70º C for one hour. The DNA was pelleted by centrifugation at 1,400 rpm on a bench top centrifuge, washed with 70% ethanol, and then with ethanol:ether (1:1). The pellet was dried under vacuum and then dissolved in DNA gel loading buffer for analysis.

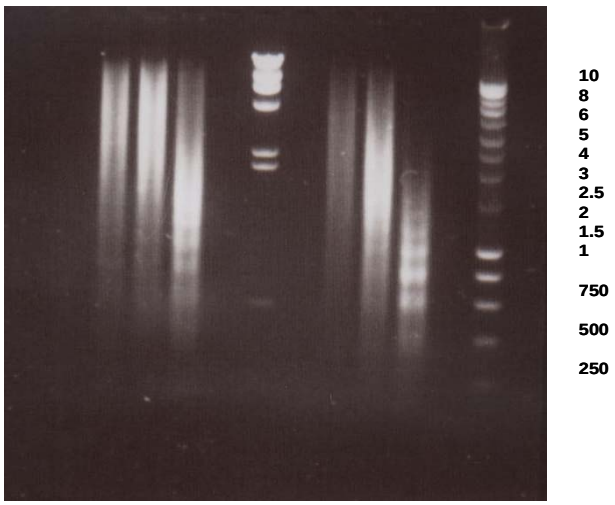

Figure 3.2 DNA recovered from fractions of chromatin 4, 5 and 6 from sucrose gradient, run on 0.8% agarose gel. Center M on gel, M marker lane comprised of DNA bp ranging from 40,000 bp to 5,000 bp. Far right marker lane, DNA from 10,000 bp) to 250 bp.

3.7 DNA Cellulose Chromatography

Chromatin fractions were stripped of linker histones, H1 and H5, by chromatography on DNA cellulose. The chromatin was dialysed overnight in TEP_{80} to remove the sucrose and then applied to a DNA cellulose column equilibrated with TEP_{80} and was overlaid with approximately 1.5ml of TEP_{80}. Care was taken to ensure that the

total volume of sample and overlay was less than the excluded volume of the column. The column was closed and left in the cold room overnight to strip the linker histones from the chromatin. TEP_{80} was added to the column to elute the chromatin and 1ml fractions were collected and monitored at OD_{260} (see figure 3.5). When collection was complete, the column was washed with 2M NaCl to remove the bound linker histones. The efficiency of linker histone depletion from chromatin was monitored by SDS-gel electrophoresis.

3.8 HAP Column/DNA Isolation/Core Histone Isolation

HAP Buffer 10mM PO_4, 0.1mM EDTA, 0.2mM PMSF, 0.65M NaCl

20g of hydroxyapatite (HAP) were suspended in 10mM PO_4 0.65M NaCl buffer and chromatin from the nuclei digestion (see section 3.4) was added to the suspension.

The mixture was poured into a column and allowed to settle and then washed with 10mM PO_4 0.65M NaCl buffer to elute unbound proteins (linker histones). The core histones were then eluted by washing the column with 10mM PO_4 2M NaCl buffer 1ml fractions were collected and monitored at OD_{280} (see figure 3.3) and fractions containing core histones were collected, the concentration of the protein was determined to be 653µg/mland the fractions were stored at -70ºC. The column was then washed with 1M PO_4 buffer to recover the DNA.

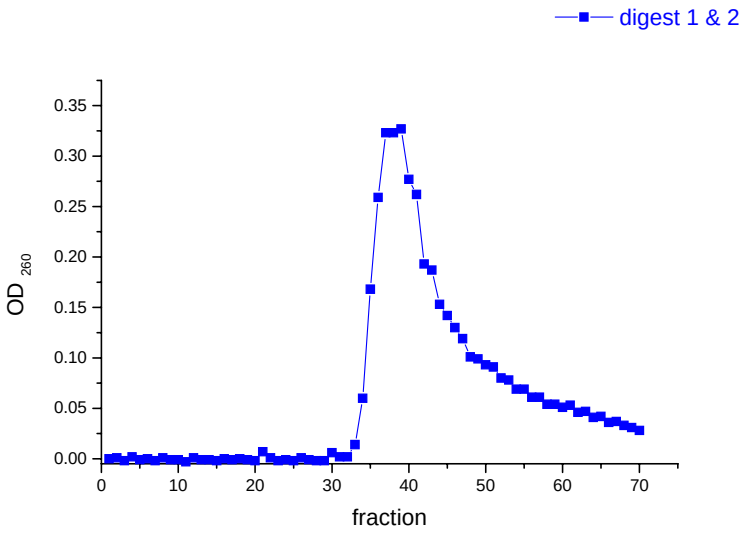

Figure 3.3 Core histone isolation on HAP column. DNA preparation from HAP column. Digest 1 and 2 refers to two separate digestions of chromatin loaded onto the HAP column for core histone isolation, UV monitorying at OC_{260} fractions 35 to 50 collected and stored at -80°C.

3.9 Sepharose Column

The sepharose buffer is TEP 0.65M NaCl. Chromatin was prepared for loading onto the sepharose (Q Sepharose fast flowing, Pharmacia) column with the addition of 5M NaCl to provide a final concentration adjusted to 0.65M NaCl. The chromatin was loaded onto the column, an equivalent volume of buffer was then added and 4ml fractions were monitored by UV at OD_{300}. The samples were stored at -70ºC.

The column was rinsed with TEP 2M NaCl to strip linker histones from column.

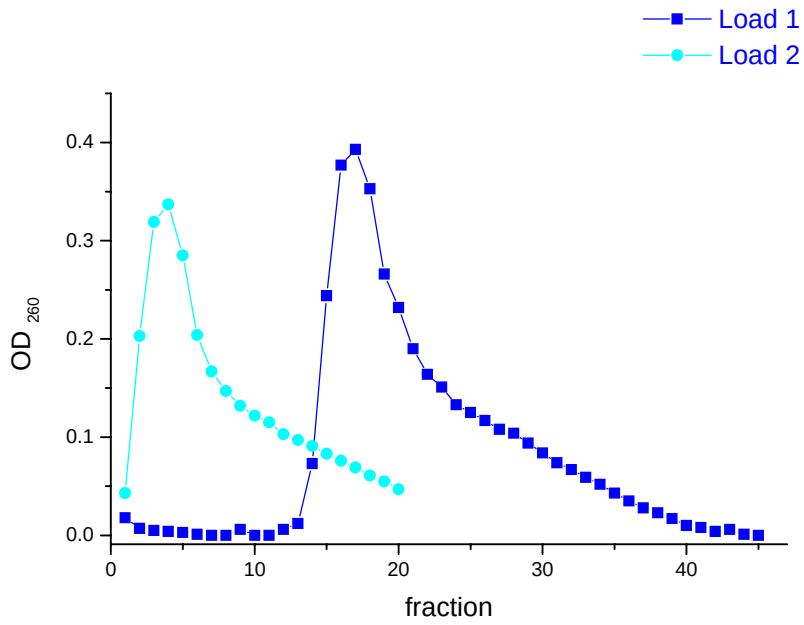

Figure 3.4 Chromatin from nuclei digestion in TEP_{80} stripped of linker histones on sepharose column. Eluted with 0.65M NaCl TEP. Load 1 and 2 refer to separate loadings of chromatin from nuclei digestion onto the sepharose column. Earlier fractions of load 2, which would correspond to fractions 1-13 from load 1, were not collected based on the UV reading from load 1.

3.10 Agarose gels

Agarose gels for DNA analysis: 0.8% agarose Tris-borate ethylenediaminetetracetate TBE (45mM Tris-borate, 1mM EDTA [(TBE) (0.045M Tris-borate, 0.001M EDTA)] + ethidium bromide 0.5ug/ml.

3.11 Gel Electrophoresis of Proteins

Proteins were routinely assayed by sodium dodecyl sulfate/polyacrylamide gel electrophoresis (SDS/PAGE) in a Tris-glycine discontinuous buffer system, according to the method of (Laemmli 1970) and (King and Laemmli 1973).

18% separating gel: polyacrylamide/bisacrylamide (40% 29:1), 1.5M Tris pH8.8, water, 10% SDS, APS (75mg/ml ammonium persulfate), TEMED (N,N,N',N'-Tetramethylenediamine)

3% stacking gel: polyacrylamide/bisacrylamide (40% 29:1), 1M TRIS pH 6.8, water, 10% SDS, APS, TEMED

Protein Gel Loading Buffer purchased from New England Biolabs

Gel running buffer TRIS/Glycine/SDS, 25 mM TRIS pH 7.5, 192 mM Glycine and 0.1% SDS.

A 15.5cm x 11.5cm slab gel of 1mm thickness was routinely used. Glass gel plates were thoroughly cleaned and assembled. An 18% separating gel was routinely used and prepared as above. The separating gel was poured between the glass plates and overlaid with butanol saturated with 1.5M Tris pH 8.8 and allowed to set. The saturated butanol was poured off and the stacking gel was poured on top of the separating gel. The 3% stacking gel was prepared as above. Wells for protein samples were prepared by inserting a comb into the stacking gel. The gel was then placed into a vertical tank and the reservoirs were filled with gel running buffer. Voltage was applied at 100V through the stacking gel and at 200V through the separator until the dye (bromophenol blue) runs off the end of the gel at which point the voltage is applied for an additional hour.

3.12 Protein visualization

Coomassie brilliant blue stain: 0.05% coomassie brilliant blue G250 in 45% methanol, 9% acetic acid.

Destaining solution: 5% methanol, 7.5% acetic acid.

3.13 Trypsin Digestion

Porcine pancreatic trypsin was prepared at 1mg/ml and used as a stock for the digestion. H5 was prepared at 19mg/ml and used as the stock. The H5 was available in gram quantities in the laboratory and it was therefore unnecessary to isolate more. The stocks were both stored at -20ºC. The digestion was carried out on ice, and was stopped with a variety of PMSF concentrations (10mM to 250mM) and precipitated overnight at 4ºC in six times initial aliquot volume with acetone.

H5 was prepared to 50µg/ml in TE buffer with varying concentrations of NaCl, 10µl of 100µg/ml of trypsin and digestion initiated. Time points from 0-60 minutes.

3.14 Protein extraction from gel

With a razor blade, bands of interest were excised from the gel. These slices were then chopped into small pieces and placed in ependorf tubes. 200µl of 1:1 acetonitrile (ACN)/water was added to each ependorf and the mixture was vortexed. The samples were left to stand for 2-4 hours depending on amount of dye. This step is intended to remove the coomassie blue dye from the gel slices. The ACN/Water was pipetted off and 200µl of 1:3:2 formic acid/isopropyl alcohol/water (FIAW) was then added to the gel slices. This mixture was then vortexed, and allowed set aside for at least 2 hours and preferably overnight. The solvent was transferred to new eppendorfs. The sample was then placed under vacuum to remove the solvent. Once the sample had dried to completion, 10µl of MALDI matrix was added and 1µl of sample/matrix mixture was spotted onto MALDI target plate.

3.15 Reductive Methylation

Reductive methylation was first performed on 100µg/ml H5 alone in 5mM HEPES 80mM NaCl buffer. The ratio of formaldehyde to lysine concentration needed to be 12:1; (Thomas and Wilson 1986) this meant that a 20µM concentration of formaldehyde was required for initial testing of the reductive methylation process. H5 has a mass of 20,605 Da and the 44 lysines present on H5 have a mass of 6,424 Da, which means a 1.527mM concentration of lysine (K) present in 100µg/ml of H5. Therefore, in order to have the appropriate concentration of formaldehyde in the reaction, 40µl of 50mM formaldehyde was added to the 10ml reaction volume for a final concentration of 20µM.

The $NaCNBH_4$ concentration was shown to be most effective at 15mM (Gidley et al. 1982). The reaction was performed at 24°C over the course of 2 hours, 500µl aliquots were removed and the reaction was stopped with 1/10 volume of 1M glycine. The protein was precipitated with trichloroacetic acid added to give a 25% solution in each aliquot. After ½ hour on ice the precipitated was pelletted by centrifugation and the supernatant removed by aspiration. Each pellet was then washed with 600µl acetone/0.1M HCl. The sample was again centrifuged and the solvent removed by aspiration. The pellet was then washed with 600µl of acetone.

The sample was treated as above and the remaining acetone evaporated under vacuum. The pellet was re-suspended in 50µl 0.1% ammonium bicarbonate (NH_4HCO_3) solution. A 5µl aliquot was removed and prepared for SDS/PAGE analysis. The remaining sample was digested with trypsin, 5µl of 100µg/ml trypsin added to each aliquot. The digestion was run overnight to allow for complete digestion.

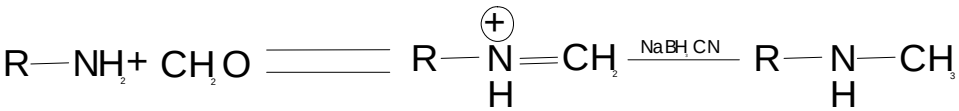

Figure 3.5 Schematic representation of the reductive methylation reaction.

3.16 Mass Spectrometry

The following instrument settings were employed and constitute one run: the laser was set to acquire 500 scans from mass 6000-25,000 at an intensity of 2600, grid voltage 94%, guide wire voltage 0.15%, accelerating voltage 25000V, the extraction delay time was 1500 nsec and the laser repetition rate was 20.0Hz.

Using the time point of 2 minutes, the settings, necessary for the ionization of the peptides from the digest, were determined. This allowed for the same instrument settings to be used for all samples and it was determined at the time that the above settings allowed for maximum ionization of the peptides. The only alteration to the settings was the mass range. The software allows for multiple scans to be combined, this would mean that the total number of scans would be anywhere from 500 to 2,000, where more than 1 analysis comprised of 500 scans is combined to produce a final spectrum, for the data observed in order to improve the signal to noise ratio for a particular sample.

The samples were initially run as the complete digest from the individual time points. These scans were then overlaid and compared. While this was illuminating because the comparison of the scans showed an increase of some peptides over time as other peptides decreased, it was not beneficial in determining where along the protein the sites of greatest sensitivity to trypsin were located.

Chemical suppliers list:

PMSF: Sigma

Sodium chloride: Sigma

EGTA: Sigma

PBS tablet: Sigma

Polyacrylamide 29:1 National Diagnostics and Sigma

TEMED: Sigma

Ammonium persulphate: Sigma

DNA cellulose: Sigma

HEPES: Sigma

Formaldehyde: Aldrich

Sodium cyanoborohydride: Sigma

Chloroform: Aldrich

Beta mercaptoethanol: Sigma

Porcine pancreatic trypsin: Sigma

Sodium dihydrogen phosphate: Fisher

Potassium chloride: Fisher

EDTA: Fisher

Sucrose: Fisher

Disodium phosphate: Fisher

Tris: Fisher

Ethydium bromide: Fisher

SDS: Fisher

Glacial Acetic acid: Fisher

Methylene chloride: Fisher

Urea: Fisher

Isopropanol: Fisher

Ethanol: BDH

Glycerol: BDH

Magnesium chloride: BDH

NP40: BDH

Isoamyl alcohol: BDH

Ether: BDH

Bromophenol blue: BDH

Acetone: BDH

Butanol: BDH

Methanol: BDH

Glycine: BDH

Agarose: FMC Bioproducts

Acetonitrile: Rathburn

Agarose: FMC Bioproducts

Phenol: Fluka

Sepharose: Pharmacia

Dialysis tubing: Biodesign Inc. of New York

Micrococcal nuclease: Available in lab

Gel Kit supplier/manufacturer: ATTO, BioRad power supply

HAP: Calbio Chem

Supplier addresses:

Sigma-Aldrich Company Ltd.
The Old Brickyard
New Road
Gillingham
Dorset
SP8 4XT

Rathburn Chemicals Ltd
Walkerburn
Scotland
EH43 6AU

FMC BioProducts
Risingevej 1 DK-2665
Vallensbaek Strand
Denmark

Fisher Scientific UK Ltd
Bishop Meadow Road,
Loughborough,
Leicestershire
LE11 5RG

BioDesign Inc. of New York
P.O. Box 1050
Carmel, New York
10512
USA

Methods References:

Allan, J., Hartman, P.G., Crane-Robinson, C., Aviles, F.X., Nature **288** (5792) 675-679 (1980) *The structure of histone H1 and its location in chromatin*

Gidley, M.J., Sanders, J.K.M., Biochemistry Journal **203** 331-334 (1982) *Reductive methylation of proteins with sodium cyanoborohydride: identification, suppression and possible uses of N-cyanomethyl by products*

Gould, H. Oxford University Press *Chromatin: A Practical Approach* (1998)

Harborne, N., Allan, J., FEBS Letters **155** (1) 88-92 (1983) *Modulation of the relative trypsin sensitivities of the core histone tails*

Maman, J.D., Yager, T.D., Allan, J., Biochemistry **33** (6) 1300-1310 (1994) *Self-association of the globular domain of histone H5*

Jentoft, N., Dearborn, D.G., Journal of Biological Chemistry **254** (11) 4359-4365 (1979) *Labeling of proteins by reductive methylation using sodium cyanoborohydride*

Lambert, S.F., Thomas, J.O., European Journal of Biochemistry **160** 191-201 (1986) *Lysine-containing DNA-binding regions on the surface of the histone ocatmer in the nucleosome core particle*

MacGillivray, A.J., Birnie, GD., Buttersworth *Nuclear Structures: Isolation and Characterization*. (1986)

Taylor, I.A., Webb, M., Humana Press Inc. **2nd edition** Methods in Molecular Biology **148** DNA-Protein interactions: Principles and Protocols *Chemical modification of lysine by reductive methylation*

Thomas, J.O., Methods in Enzymology **170** 369-385 (1989) *Chemical radiolabeling of lysines that interact strongly with DNA in chromatin*

Thomas, J.O. Wilson, C.M., EMBO Journal **5** (13) 3531-3537 (1986) *Selective radiolabelling and identification of a strong nucleosome binding-site on the globular domain of histone H5*

4 Results

Our aim was to identify and quantitated the susceptibility of the (trypsin) cleavage sites in linker histones (H5). Our primary source of this information would be the bands observed in SDS acrylamide gels. However, to identify the nature of these bands (which part of H5 they represent) SDS gel electrophoresis alone would not be adequate because (i) products of H5 run anomalously and therefore their sizes can't be assessed reliably and (ii) size alone will not allow us to identify the peptide (sequence). It is for this reason that we wanted to use mass spec approaches.

4.1 Preparation and characterization of materials

Chromatin was prepared from chicken nuclei which was digested with micrococcal nuclease and chromatin with approximate length of 4,000 bp was isolated by centrifugation on in a sucrose gradient, see figure 4.1. The length of the chromatin in the 2.5 ml fractions was determined by analysis on 0.8% agarose gel, see figure 4.2. Fractions containing the desired length of chromatin were selected for the time course digestion work.

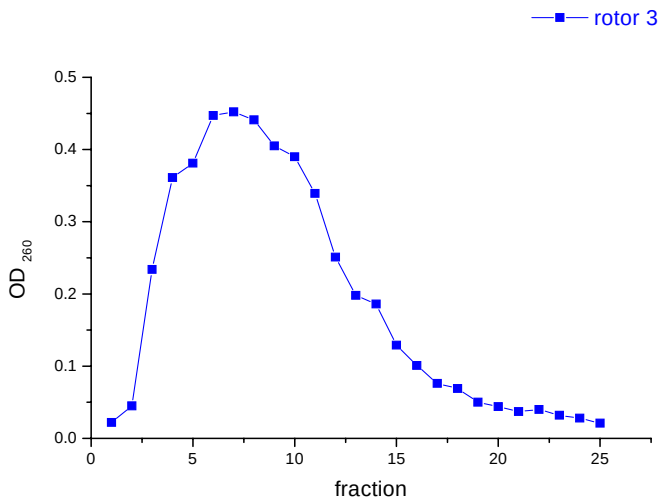

Figure 4.1 UV trace of sucrose gradient 12.5% to 44% of chromatin readings taken at OD_{260} and 2.5 ml fractions were collected from the top of the gradient, fractions 4 through 12 were collected across the curve and the approximate length of chromatin contained was identified by agarose gel electrophoresis. Selected fractions with desired chromatin of desired length, approximately 4,000 bp, were used for further work. Rotor 3 indicates these readings are from the chromatin layered on the sucrose gradient loaded in rotor 3, 4 rotors were loaded in total and balanced for centrifugation by weight.

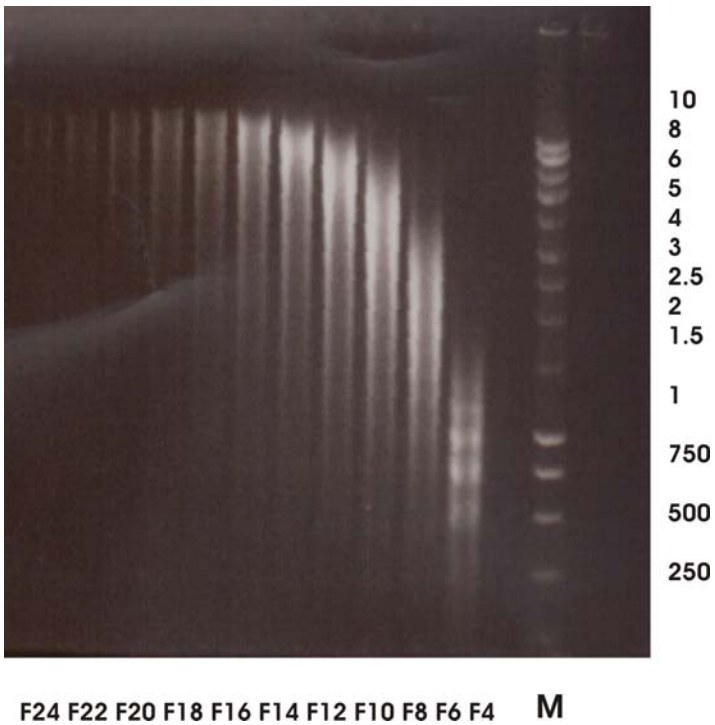

Figure 4.2 DNA recovered from fractions 4 through 24 of chromatin collected from sucrose gradient. The samples were run on 0.8% agarose gel containing ethidium bromide for visualization. M, marker lane, DNA from 10,000 base pairs (bp) to 250 bp. Fractions 12, 11, 10 and 9 contain DNA of appropriate length and these were used for further experiments.

Selected fractions; 4, 5 and 6, were shown to have chromatin with an approximate length of 4,000 bp and were then stripped of linker histones on a DNA cellulose column. 1ml fractions were collected from the DNA cellulose column and when needed the chromatin was mixed with linker histone H5 for the digestion experiments. Figure 4.3 shows the optical density readings of the fractions collected from the DNA cellulose column fraction 3 to fraction 13 were used for the digestion work.

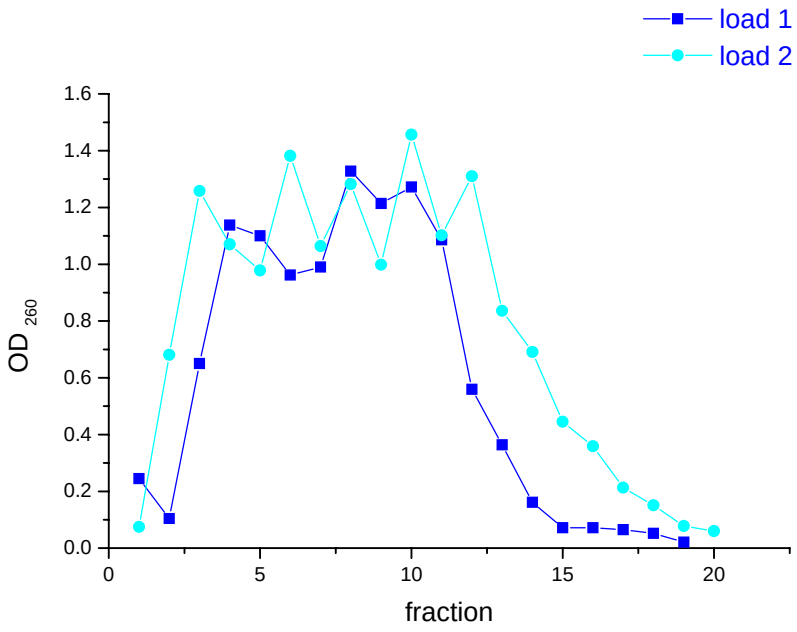

Figure 4.3 Chromatin of required length, approximately 4,000 bp, were stripped of linker histones on DNA cellulose column, the chromatin was split into two fractions for the loading on the column and added in two separate loadings. 1 ml fractions were manually collected from both loadings and the optical density measured at OD_{260}. The fractions containing the stripped chromatin, fractions 5 to 13, were set aside for future work.

The chromatin was stripped of the linker histones on the DNA cellulose column, see figure 4.3. The core histones were isolated from the DNA and analysed by SDS/PAGE analysis. The first lane of the gel in figure 4.4 shows the core histones and no presence of the linker histone H5. The chromatin was reconstituted with a known amount of linker histone H5 for the digestion work.

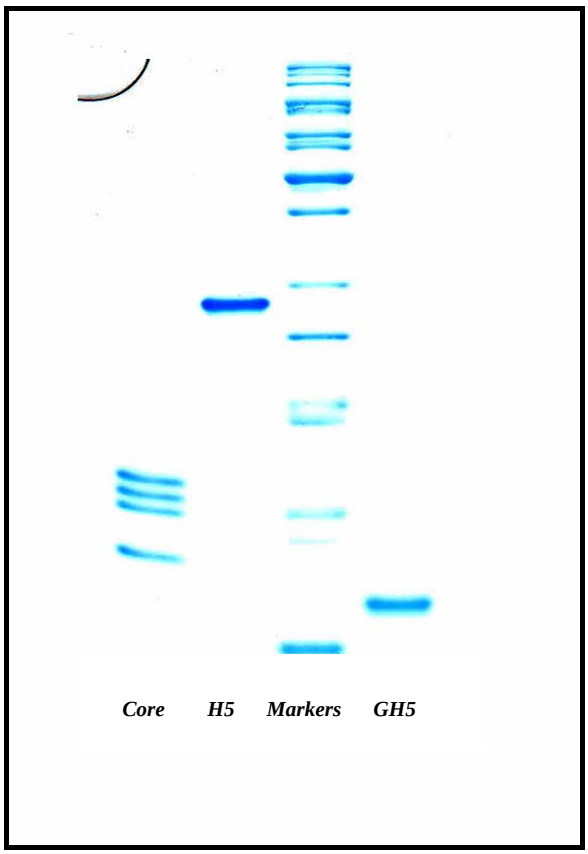

Figure 4.4 SDS/PAGE Protein gel of: in the first lane core histones isolated from chromatin which has been stripped of linker histones on the DNA cellulose column; in the second lane isolated H5; the third lane protein markers; the fourth lane the globular domain of H5. Note the mark in top left corner of gel is the edge of the gel.

4.2 H5 Analysis

Initial H5 identification by electrospray (on a Micromass Q-ToF 1) was done to show the quality the previously purified H5 protein, see figures 4.5, 4.6 and 4.7. The H5 was supplied by Dr. Jim Allan and had been previously purified and as can be seen by SDS/PAGE (see figure 4.4) visualizes as a single band. Each principal group of peaks in the full spectrum comprises of a number of individual peaks, which could be due to modifications (post-translational), sequence variations or the most likely adducts (formate, potassium or sodium).

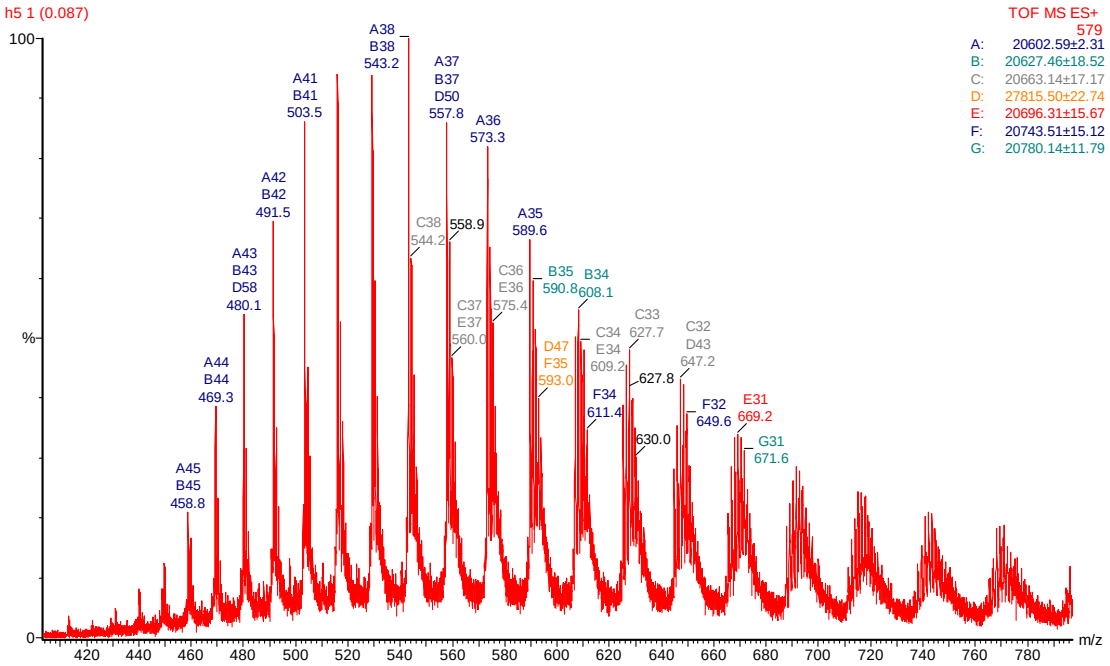

Figure 4.5: Direct infusion ESI-MS Q-Tof spectrum for chicken erythrocyte H5. The spectrum shows the multiply charged peak envelope characteristic of electrospray ionization.

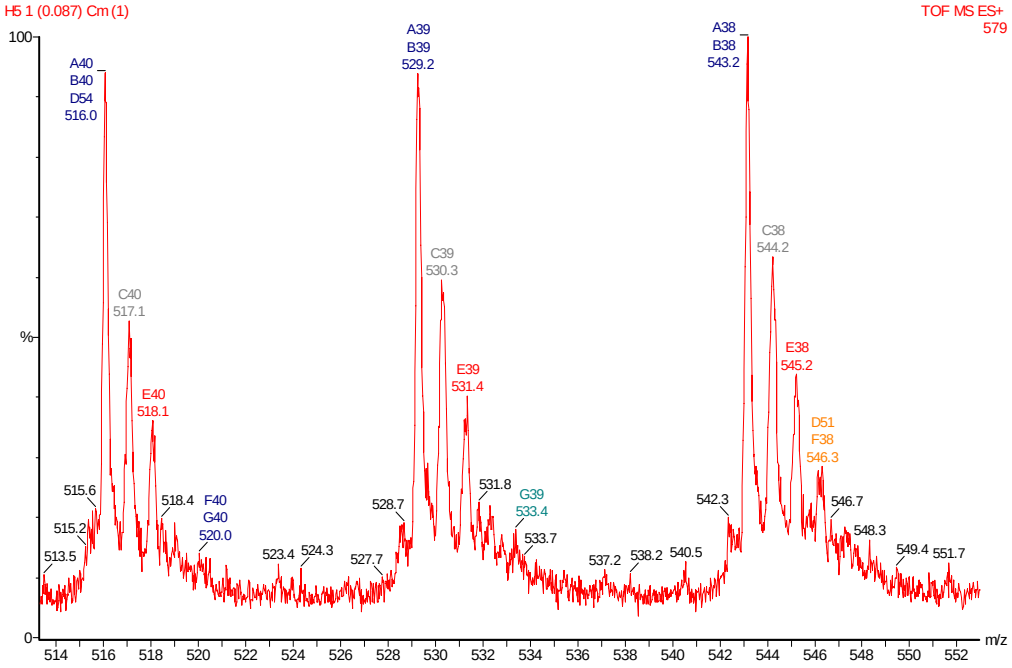

Figure 4.6 Expanded view of the ESI-MS spectrum for chicken erythrocyte H5.

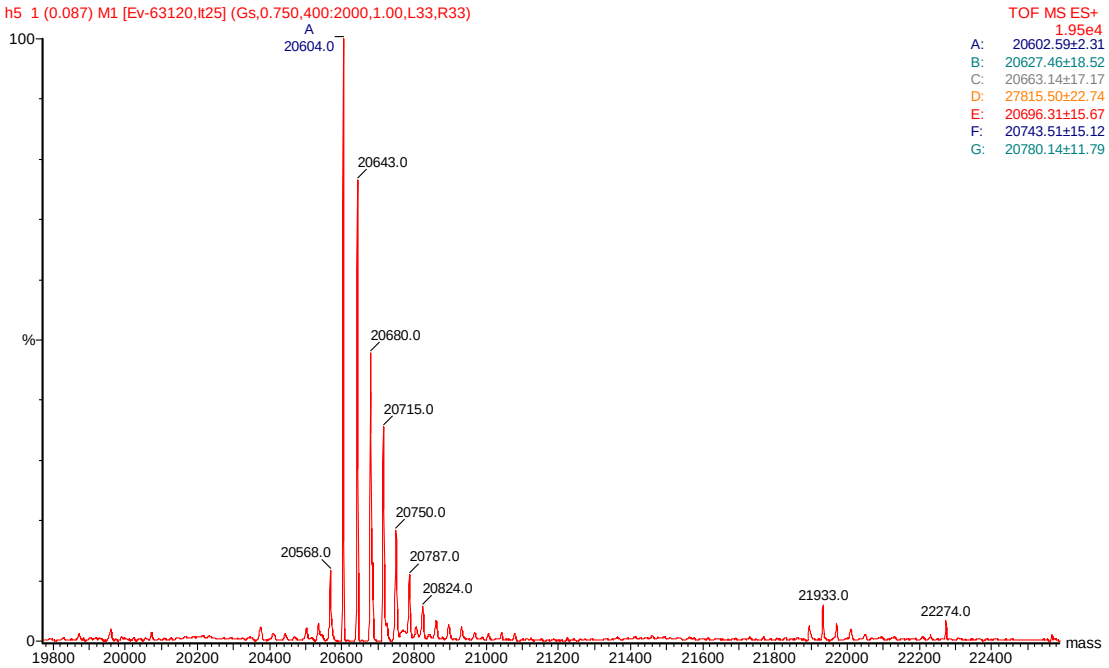

Figure 4.7 MaxEnt deconvolution of ESI-MS spectrum for Chicken erythrocyte H5 observed masses; 20604, 20643, 20680, 20715, 20750, 20787, 20824, calculated mass for the protein is 20474 Da (average mass calculated 20602).

The identification of H5 digest products was conducted by several means; SDS/PAGE gel analysis, MALDI analysis of digest products, chromatographic separation of digestion and MALDI analysis of collected fractions, in gel digestion and extraction of H5 digest products and elution from gel of H5 digest products. The results of these experiments are found here.

4.3 Trypsin digestion

The composition of H5 makes it highly compatible with mass spectrometry as an analysis tool. The rich lysine and arginine C-terminal tail lends itself to mass spectrometric analysis for identification. The aim was for mass spectrometry to identify the lysines and arginines in the C-terminal tail and assess the relative sensitivity of each to trypsin in varying digestion conditions.

Trypsin is a robust protein maintaining activity in protein denaturants such as SDS (0.1%), urea (1M), guanidine hydrochloride (1M) and acetonitrile (10%) (*Proteolytic Enzymes* 2001). Typical means of halting trypsin digestion include the addition of, but not limited to, PMSF, soya bean trypsin inhibitor (STI) (Rawlings et al. 2006) and formic acid. It is necessary to stop the action of trypsin quickly so as to identify

the sites of cleavage and determine how these cleavage sites differ in varying concentrations of salt in the presence of chromatin.

In the case of this study, STI was initially employed to stop the time course digestions, but in order to identify peptides from the H5 digestion by SDS-PAGE analysis, it was necessary to use an alternate means of stopping the digestion as the STI shows as two distinct bands on the gel which interfere with the analysis of the H5 digest products. (See figure 4.9) So, PMSF was employed to halt the digestion. In this study PMSF proved inconsistent at halting the digestion process. (See figures 4.10 and 4.11) The protein, after digestion, is acetone precipitated overnight and this acetone was employed to stop the digestion. This resulted in a time course containing a certain amount of variability. (See figure 4.12) As can be seen in figure 4.13, the H5 at T2 has been digested to a greater extent than at time point T30. An additional problem encountered was the failure of the digested protein to go completely into solution in the loading buffer. The result, in this case, was incomplete representation at a specific time point. (See figure 4.12) It was necessary to optimise the conditions of the experiment to ensure proper digestion time points in order to establish the relevant digest products.

PMSF is a known serine protease inhibitor, it is the least selective and therefore most useful as a general inhibitor. PMSF while an irreversible inhibitor of serine proteases reacts slowly; at a concentration of 1mM PMSF takes about 4 minutes to inhibit the action of trypsin (*Proteolytic Enzymes* 2001). The PMSF is dissolved in isopropanol and stored in the 37˚C water bath. PMSF undergoes spontaneous hydrolysis in aqueous solutions causing it to become inactive (25˚C, pH 7.5, 55min). When problems using PMSF as an inhibitor occurred, fresh PMSF was made up to eliminate the possibility of hydrolysis causing it to become inactive (James 1978). The concentration of PMSF used to stop the digestion was varied from 10mM to 250mM. The effect of an increase in PMSF concentration was not evident from the digestions and in this work the PMSF was inconsistent for the establishment of a time course (see figures 4.10 and 4.11 as examples). This is in comparison to the STI where the digestion time course was consistent, with the number and intensity of the digest products increasing over time.

The third means of halting the trypsin was to use acetone precipitation. As the acetone is already used at 6 times the aliquot volume for the precipitation of the proteins from the buffer it was also therefore used to stop the digestion. Halting the digestion was successful but again as with the PMSF not as consistent as halting the digestion with the STI. It appeared that a possible problem could be that the trypsin was able to regain activity once redissolved in the gel loading buffer or that the cessation of action was not as effective as with PMSF or STI (figure 4.12). See time point T2 in figure 4.9, here the acetone is ineffective for halting the digestion as this time point shows. Stopping the digestion with 6 times the aliquot volume with acetone, as needed for the protein precipitation, was unsuccessful and believed to be due to the fact that the trypsin 'regains' activity once the digest products are dissolved in the gel-loading buffer.

In the case of this study, trypsin was employed for digestion because of the compatibility with mass spectrometry.

TESLVLSPAPAKPKRVKASRR
SASHPTYSEMIAAAIRAEKSRGGSSRQSIQKYIKSHYKVGHNA
DLQIKLSIRRLLAAGVLKQTKGVGASGSFRLAKSDK
AKRSPGKKKKAVRRSTSPKKAARPRKARSPAKKPKATARKA
RKKSRASPKKAKKPKTVKAKSRKASKAKKVKRSKPRAKSGA
RKSPKKK

Figure 4.14: Sequence of linker histone H5 color coded to show regions of structural (globular domain and unstructured C- and N-terminal tails) and chemical (tryptic digestion sites) interest. Amino acid color coding: Blue-Globular domain, Red-Lysine and Arginine Green-Lysine, Arginine and Proline, as proline inhibits cleavage by trypsin when neighbor to lysine and arginine. Calculated average mass of the protein is 20602 Da.

The sequence of H5 (see figure 4.14) lends itself to the use of trypsin for this study. With a total of 66 lysines (color coded red in figure 4.14) and arginines (color coded green in figure 4.14) present in the protein, H5 is well suited for the use of trypsin in one-hit kinetics with the majority of these amino acids located in the N- and C-terminal tails of the protein. The high frequency of lysines and arginines should allow for better opportunities to identify potential binding sites. If there were fewer basic residues, the peptides would be larger and identification could prove more complicated. In the case of H5, the basic residues allow for greater coverage of the N- and C-terminal tails where potential binding sites are located. Work from (Allan et al. 1980) show the globular domain of the linker histones located at the dyad axis.

The cross-linking work, (Thomas et al1986), shows the lysines which are interacting with the DNA. This work does not tell us what the linker histone tails are interacting with. The assumption is the linker histone tails are interacting with the linker DNA for the compression of the fibre. Given the number of basic residues present in the linker histone tails in conjunction with the cross-linking work from (Thomas et al. 1986) and (Goytisolo et al. 1996) it is a possibility that the basic residues are interacting with the linker DNA. (Bharath et al. 2002) showed, using site-directed mutagenesis, that a 34 amino acid stretch (residues 145-178) of the C-terminal tail of H1 was responsible for a nearly 90% decrease in condensation of DNA when this stretch of the protein was deleted. Work from (Roque et al. 2004) shows the C-terminal domain of H1 maintains strong and preferential binding to scaffold-associated regions (SAR) of DNA while the N-terminal and globular domain show weak binding and poor specificity. The salt concentration has shown a pronounced effect on the structure the C-terminal tails of the linker histone adopt as well as how that structure, as a result of the salt concentration, affects the binding to the linker DNA.

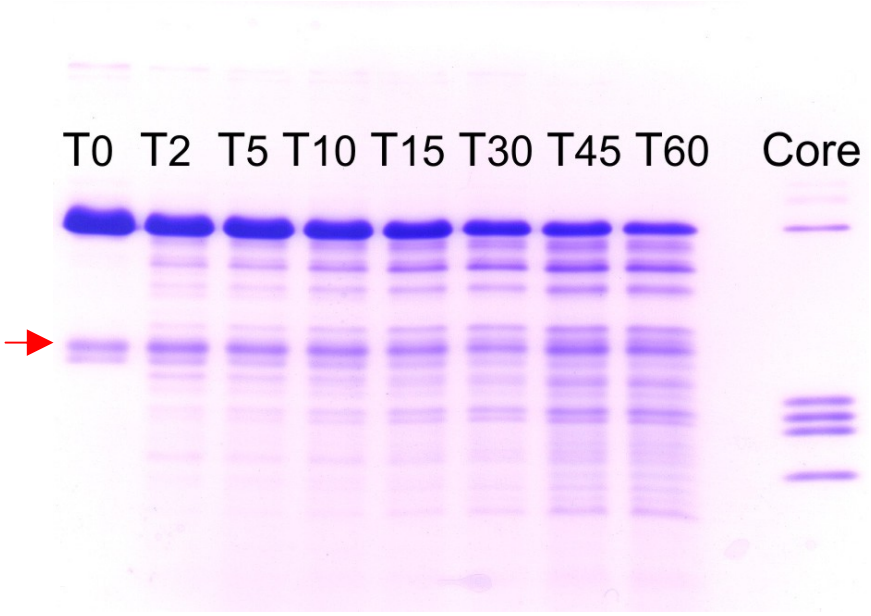

Figure 4.9 Time course digestion of H5 with soybean trypsin inhibitor (STI) to stop digestion, STI marked with red arrow. STI appears on the gel in the same location as a major fragment from the digestion and was consequently replaced by PMSF. T0=time point zero minutes, before the addition of trypsin to the solution; T2=time point 2, two minutes after the addition of trypsin to the solution; T5=time point 5, five minutes after the addition of trypsin to the solution; T10=time point 10, 10 minutes after the addition of trypsin to the solution; Core histones and H5 are used as markers. Reading down the gel, the top band is H5 followed by H2B, H2A, H3 and the bottom band is H4.

Figure 4.10 Short time course digestion (including only time points 0, 2, 5 and 15 minutes) halted with PMSF. Reading the gel from left to right, the first marker is H5, the second lane the core histones (H2B, H2A, H3 and H4), the third lane is the globular domain of H5.

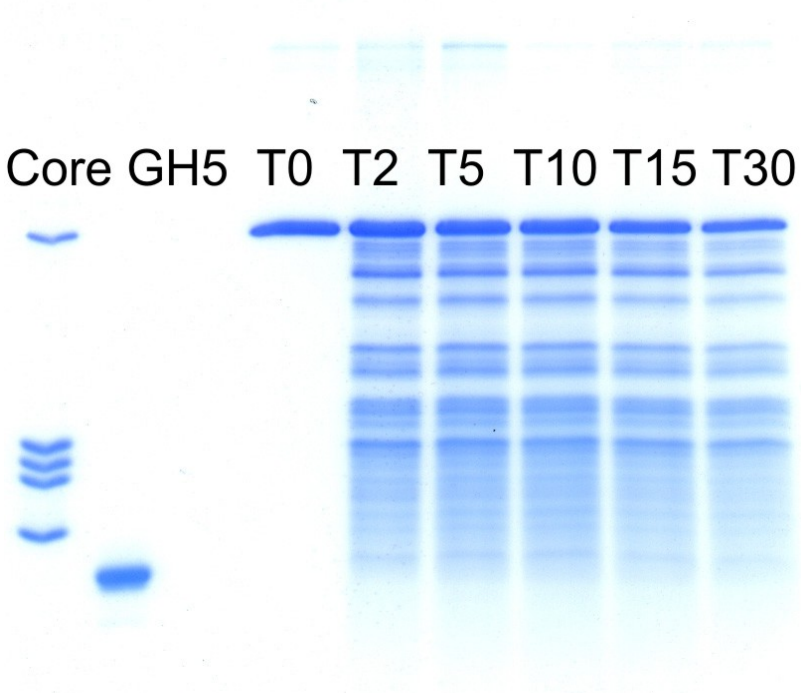

Figure 4.11 H5 time course digestion (time points 0, 2, 5, 10, 15 and 30 minutes) with PMSF used to stop the digestion. The markers are in the far left lane, reading down the gel, H5, H2B, H2A, H3 and H4, the following lane is the globular domain of H5. H5 digested in TE_{80} with trypsin at 1:50, digested halted with acetone, gel visualised with coomassie.

Figure 4.12: The above gel shows where acetone was used to stop the H5 time course digestion (time points 0, 2, 5, 10, 15, and 30 minutes). In this case the T2 sample has less H5 present than the following time points, this should not be the case as the peptide indicated by the red arrow is more intensely visualized then in the following time points again indicating over digestion at that time point. The markers for this gel are the core histones with H5 and the globular domain of H5.

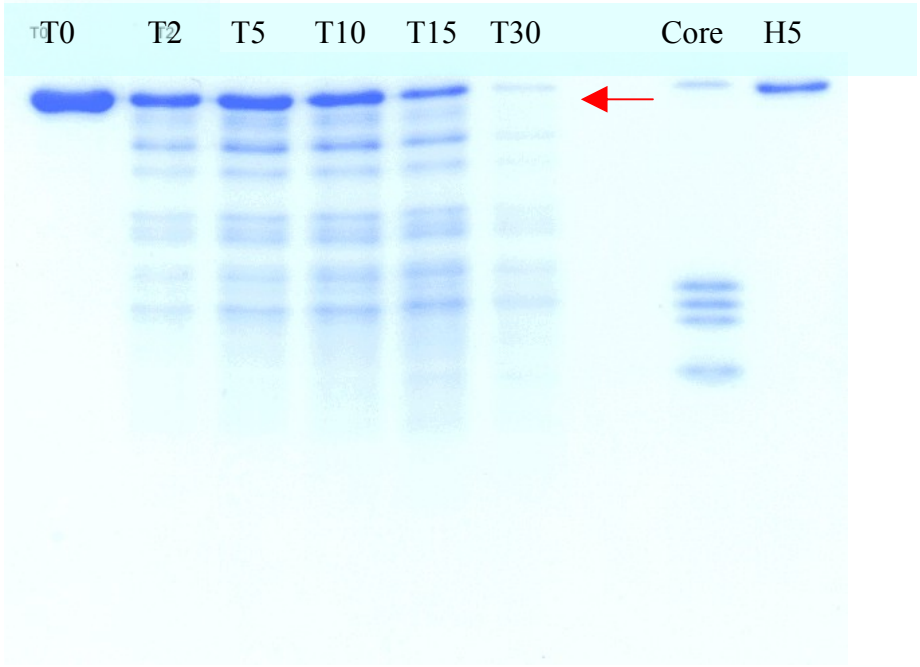

Figure 4.13 H5 digestion demonstrating failure of the protein/peptides to completely dissolve in the loading buffer as indicated with the red arrow at time point T30. The H5 protein and its digest products are barely visible on the gel. The digestion was halted with PMSF. The time points are T= 0, 2, 5, 10, 15 and 30 minutes followed by the markers of the core histones with H5 and H5 alone.

The work needed to achieve the one-hit kinetic information on the H5 required that the digestion was consistent and reliable and that the gel analysis was the same. In establishing the criteria for the digestion, (this did not include the variation in salt concentrations), the assumption was if the conditions work for the 80mM TRIS they would also work for the 5mM TRIS and the 20mM TRIS. The same applied when approaching the digestion in the presence of DNA or with chromatin. As it turned out, the difficulties encountered establishing the protocol for the digestion were more time consuming than perhaps they should have been; as a result, it was not possible to address the digestions in all the varying conditions. Therefore the results as shown are an indication of the information that could be obtained from the experiment. The purpose of the work was to identify sites in the C-terminal tail of linker histone H5 which are accessible to trypsin, implying that these sections of the protein are not interacting or have less affinity with the chromatin. From that information it would perhaps be possible to determine how the H5 was interacting with the chromatin and that information would give an idea of the form the 30nm fibre was taking.

An examination of the three gels in figures 4.9, 4.11 and 4.12 shows the means used to halt the digestion process; all three of the inhibitors are successful in halting the action of trypsin. The STI was the most effective at halting the digestion of the protein and creating a clear digest time course, figure 4.9.

Due to the potential for delay and inconsistency in activity of the PMSF it is apparent that the STI needs to be used to stop the digestions on the time scale required for the experiments. This work was preparatory for the digestion of the linker H5 with DNA and H5 reconstituted with chromatin.

Figure 4.14 shows the Ferguson plot (a plot of molecular mass against electrophoretic mobility) for the histones. Figure 4.14 shows that the H5, as compared to the remaining histones, has the greatest divergence from expected observed mass by gel. The globular domain of H5, GH5, contains just 18 basic amino acids; 9 lysines, 6 arginines and all of the 3 histidines, and the assignment of mass by gel electrophoresis is more accurate as compared to the calculated mass for the amino acid sequence. It is apparent from this plot that the identification of the digest products of H5 by gel alone will be unsuccessful at identifying specific cleavage sites. The implications of this graph on the work at hand is that even as the

mass of the linker histone H5 decreases due to digestion and moves closer in mass to the core histones, there is still variation from the molecular weight markers, due to the persistent content of basic residues. The mobility of the protein is affected by the high content of basic residues present in the protein. The resulting high positive charge, even in the presence of the anionic detergent SDS, means that the histones do not have equal charge per unit length and therefore migrate anomalously on a protein gel. The implications of the difference between observed mass to actual mass of H5 in SDS gel means that even as the H5 is digested, we will still be unable to determine the mass of the peptide. Thus the possibility of identifying the site of cleavage in the C-terminal or N-terminal tails of the H5 will still be elusive.

4.4 Gel analysis

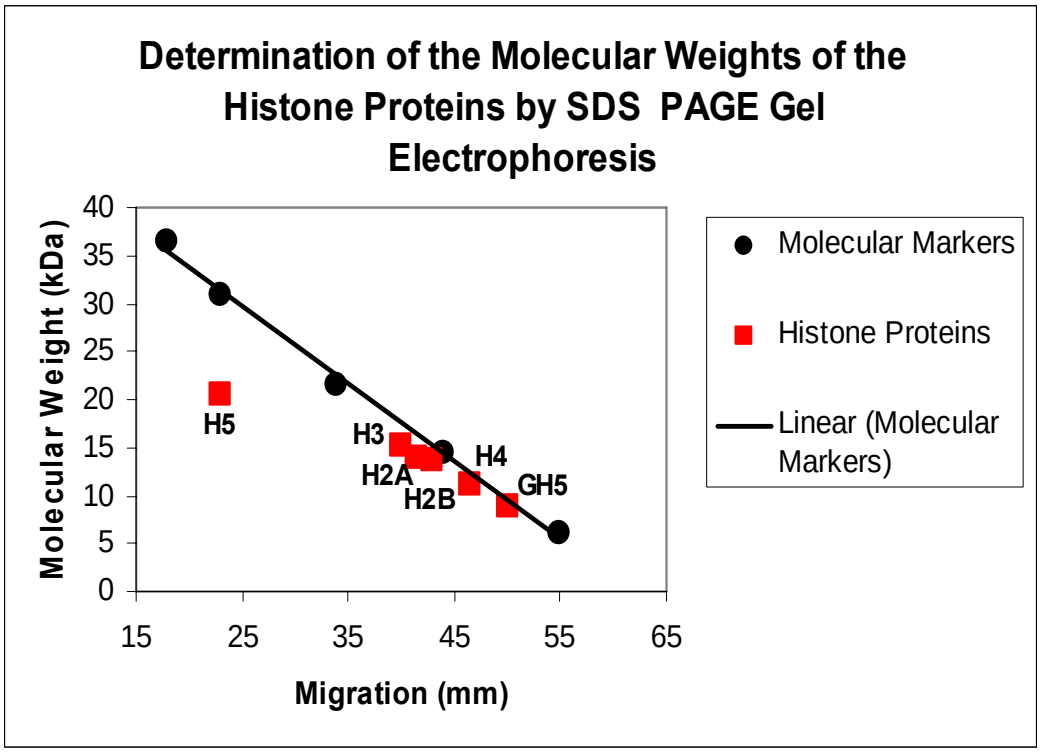

Figure 4.14 A Ferguson plot of the mobility of histones in gel electrophoresis as compared to molecular markers, the relative molecular weight of the histones and H5 globular domain was determined by electrophoresis (Mr) and compared with the calculated mass. A standard curve of distance migrated vs. log10MW, for known samples, was plotted, and the logMr of the sample was read off after measuring distance migrated on the same gel. Accuracy in determining the molecular weight of the histone proteins by SDS/PAGE was very, low therefore preventing identification of bands.

The H5 digestion was halted at specified time points, the digest products acetone precipitated and analysed on SDS/PAGE. There were a number of issues that came

up through this process. There appear to be problems with the resuspension of the H5 digest products in the loading buffer (this is not a consistent problem – see Figures 4.10 and 4.13). In some cases all time points resuspended efficiently and in others the earlier time points have not or the trypsin was not made inactive with acetone or PMSF. There were many problems to be overcome while establishing the digestion time course. The reason for not using the soya bean trypsin inhibitor was to avoid the overlay of this protein in the gel, so that all bands in the gel were from the H5 alone. With the gels where soya bean trypsin inhibitor was used to halt the digestion, there is no ambiguity in the digestion. When acetone or PMSF are used there is variability to the digestion. It is possible that some of the inconsistencies in the gels are due to inefficient resuspension of the H5 digest products in the loading buffer. There were persistent errors but they were not consistent. As a result it was difficult to isolate the problem.

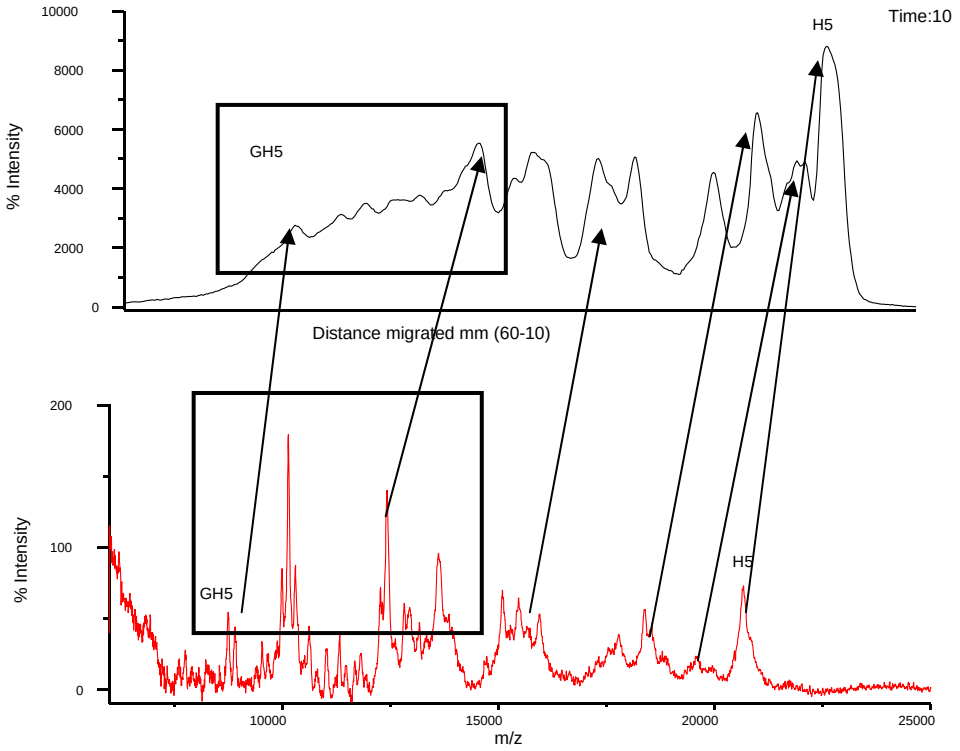

Figure 4.15: The top panel is the gel scan output for the 10 minute time point indicating the intensity of staining. The lower panel is the MALDI spectrum for the same 10 minute time point. 1mg/ml H5 digested with 50µg/ml trypsin digestion stopped with PMSF. The distance migrated in the top panel has been reversed to allow for a better visual match to the mass spectrum.

Figure 4.15 shows a comparison of migration through a gel vs. mass measurement. The lack of a quantitative correspondence between the two sets of data (protein gel vs. mass spectrum) makes a direct comparison impossible. There appear to be more digest products in the mass spectrum than in the gel scan, examining the region within the box in both the gel scan and the mass spectrum, see box region in figure 4.15, the gel itself shows the staining spreading in the lower mass region, this compares to the region boxed out in the mass spectrum. The MALDI data in figure 4.15 is the result not of gel extraction but of analysis of the digest products of H5 at time 10minutes from solution. These two means of analysis do not correlate to one another. The mass identification of the linker histone H5 by gel is not an accurate assessment of the masses. As a result, the data observed in the MALDI spectrum shows no correlation between the MALDI analysis of the digest solution and the scan of SDS/PAGE gel. The gel is scanned and the AIDA output file is plotted for distance migrated through the gel vs. % intensity of stain. Because the proteins migrate through the gel further the lower in mass, the information in Figure 4.14 is reversed from observed so that the mass spectrum and the gel scan may be roughly aligned for mass.

Taking the gels and comparing the migration distance of the digest products and the intensity of the stain increasing, indicating an increase in the amount of peptide in that band. The gel in figure 4.9 was scanned and the output file from AIDA was plotted in Origin, the graphs of the output for time points 2, 15, 30 and GH5 follow. The distance of the protein digest products is measured in millimetres (mm) and the STI is at 32 mm. Note the arrival of the peak at 22 mm which is not present at T2 (figure 4.16), becomes a shoulder on the 21 mm peak at time 15 (figure 4.17) and appears more prominently at time 30 (figure 4.18). Figure 4.19 is the plot of the GH5, the distance migrated here is 50 mm. It is for this reason that mass spectrometry was employed to give specific mass identification of the digest products.

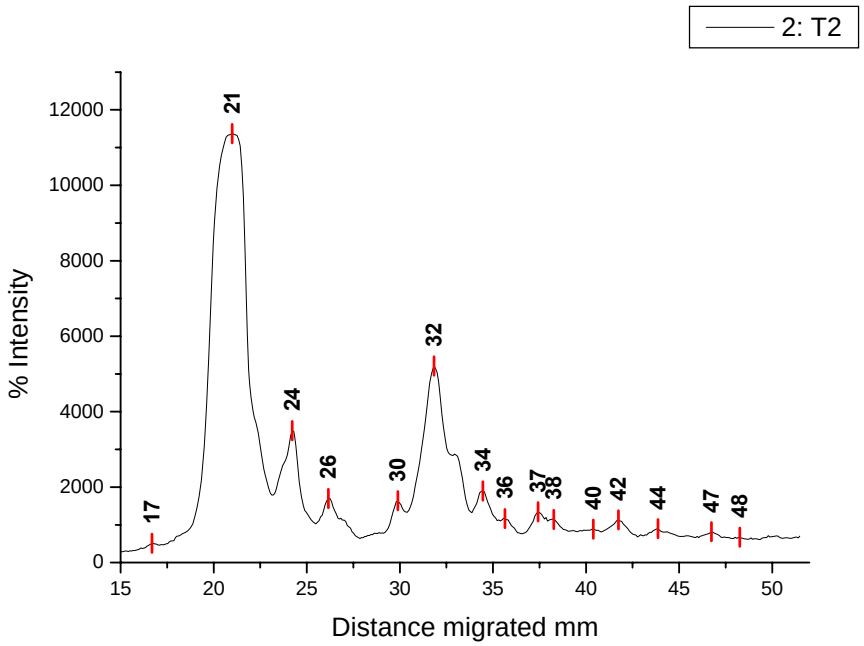

Figure 4.16 Gel output scan for time point T2, H5 has migrated 21mm and STI has migrated 32mm.

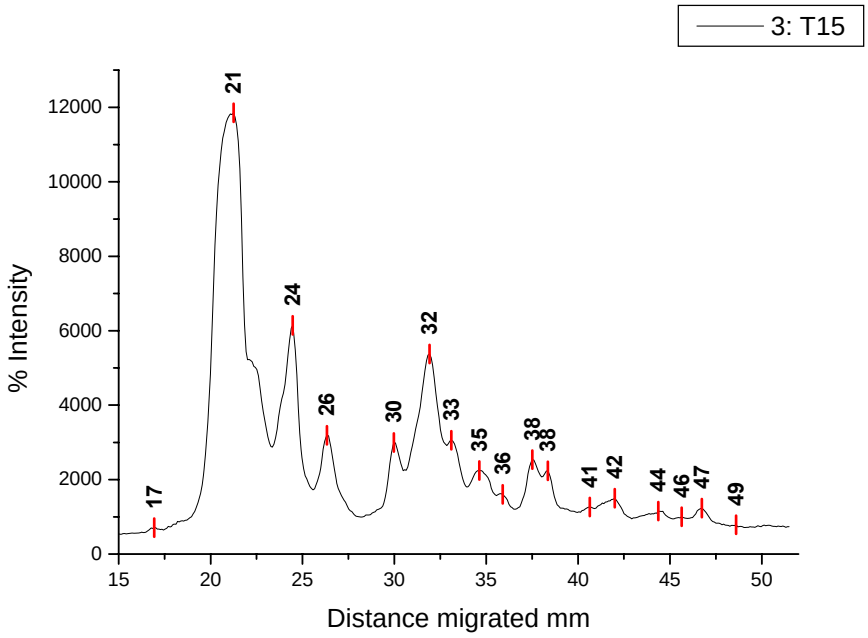

Figure 4.17 Gel output scan of time point T15, H5 has migrated 21mm and STI has migrated 32mm.

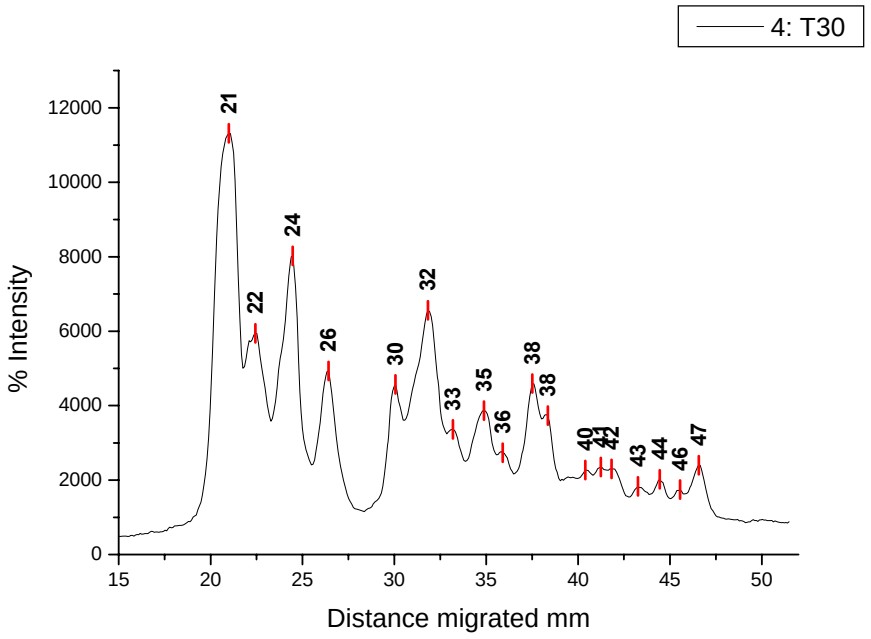

Figure 4.18 Gel output scan for time point T30, H5 has migrated 21mm and STI has migrated 32mm.

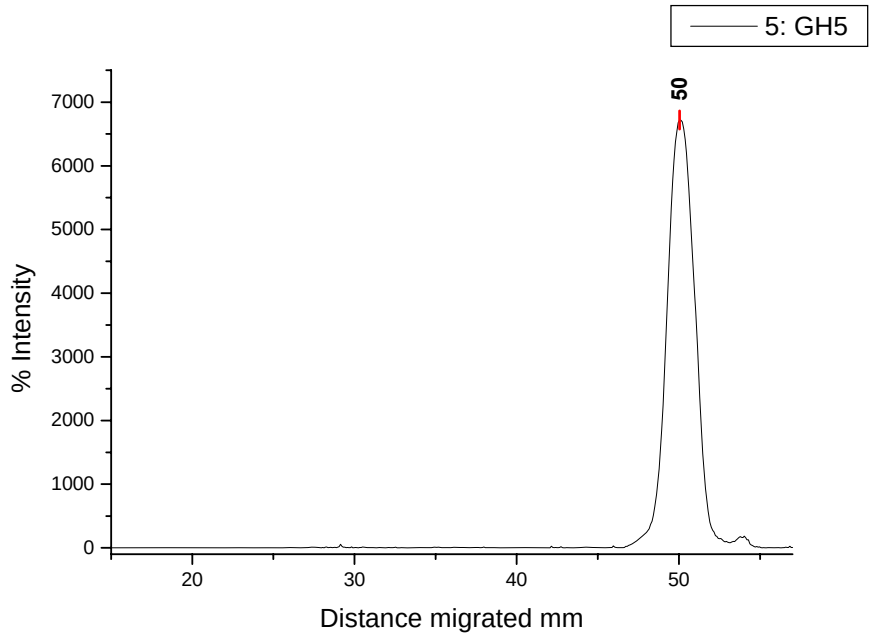

Figure 4.19 AIDA output scan for globular domain linker histone H5 a reference compound on the gels for purposes of molecular marker. Output for GH5 obtained from separate gel; see figure 4.8. GH5 was not added as a marker to gel in figure 4.9.

The above figures show the comparison of the digest time points. The data for the globular domain, figure 4.19 containing only GH5, does not contain STI as in figures 4.16, 4.17 and 4.18 because it was used as a marker for the gel analysis and was not

subjected to the digestion conditions as the other samples were. But, in order to compare one gel to another, the size of the gel, the movement of the protein and peptides through the gel have all to be consistent.

4.5 In-gel digestion

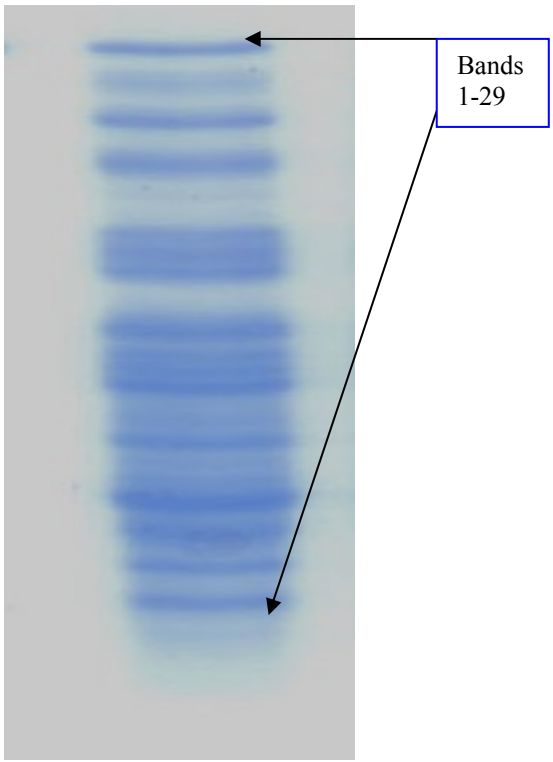

Figure 4.20 One lane from gel used for in-gel trypsin digestion. 29 bands were excised from the gel, trypsin digested and MALDI analysed.

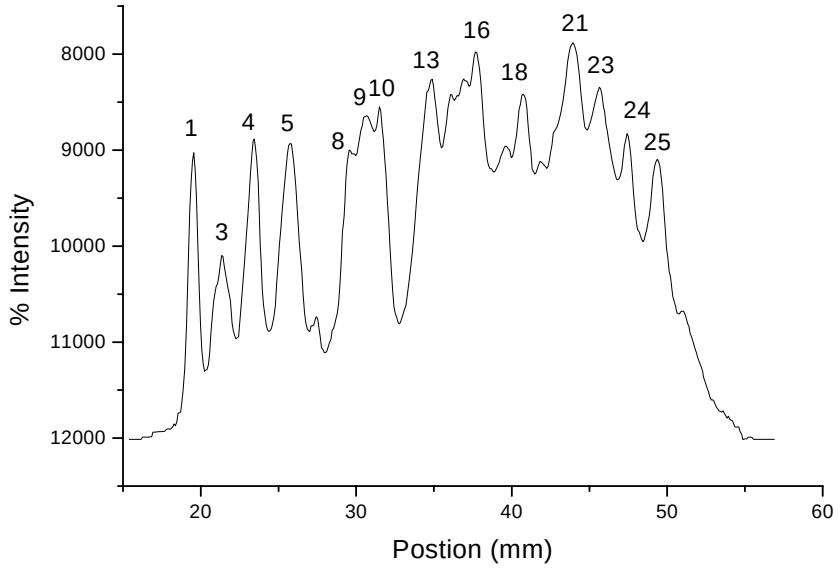

Figure 4.21 graph of gel lane, see figure 4.23, excised and submitted for in-gel trypsin digestion. Peak labels indicate band number for digestion and analysis, for mass identifications see table 4.1.

Band #	Identity
1	H5 peaks observed
2	H5 peaks observed
3	no H5 peaks
4	no H5 peaks
5	Possible H5 peaks
6	no H5 peaks
7	no H5 peaks
8	no H5 peaks
9	no H5 peaks
10	H5 peaks observed
11	no H5 peaks
12	no H5 peaks
13	Possible H5 peaks
14	no H5 peaks
15	no H5 peaks
16	no H5 peaks
17	no H5 peaks
18	no H5 peaks
19	no H5 peaks
20	no H5 peaks
21	no H5 peaks
22	no H5 peaks
23	no H5 peaks
24	no H5 peaks
25	no H5 peaks
26	no H5 peaks
27	no H5 peaks
28	no H5 peaks
29	no H5 peaks

Table 4.1 29 bands were cut out of the gel see figure 4.20 and submitted for in gel trypsin digestion MALDI mass spectrometric analysis. Only 3 of the 29 bands had confirmed H5 peptides, 2 bands had possible H5 peptides and the remaining bands showed no peptides associated with H5. This experiment was repeated twice, the results were as inconclusive. Band 1 would be the whole protein.

These 29 bands (see figure 4.20) were submitted for in-gel digestion analysis. Figure 4.21 shows the AIDA scan of the gel in figure 4.20, shown here is the graphical representation of the gel the 29 bands were excised from for MALDI analysis. Standard in-gel digestion protocol was undertaken and the resultant analyzed by MALDI/MS. The objective for this analysis was to identify the peptides by identifying as much of the c-terminal tail as possible allowing for an identity of the cleavage site from the original digestion. Three separate samples sets were submitted

for in-gel digestion analysis and the above were the best results obtained. The problem again relates to the length of time staining with coomassie, the time for destaining and the time the peptides had been in the gel before being submitted for analysis, personal communication McIntyre, S. data unpublished.

4.6 SDS/PAGE gel analysis

In the identification of digest products by gel analysis a distinct pattern is observed; see figures 4.11 and 4.20. As discussed previously, see Ferguson Plot figure 4.14 and discussion, histones, particularly H5, run anomalously by SDS/PAGE. MALDI analysis of the digestion products was complicated by a number of factors; the first of which was the presence of salts which hindered the assignment of the mass and the complexity of the sample which did not allow for a calibration of the gel.

In gel digestion results: The idea is to identify the cleavage sites from in-gel digestion. One lane in the gel was cut into 29 slices, which were submitted to an in-house service for in-gel digestion and the results are shown in table 1. The bands where H5 peptides were identified were inconclusive for determining where along the tails the digestion may have occurred. A possible reason for the lack of identifications could be due to the time between the running of the gel and the extraction of the digest products from the gel. Work from (McIntyre 2005) show that gel age affects peptide recovery, using BSA sequence coverage from in-gel digestion resulted in 15-30% sequence coverage, 7 days of storage resulted in sequence coverage less than 10%. The use of SDS/PAGE the presence of SDS detergent and protein visualization dye severely compromises mass spectrometric response to proteins (Cohen 1997).

The objective of the in-gel digestion work was to determine if segments of the tails could be identified as in tact and through inference give information of the binding to linker DNA. The experiment was repeated 3 times with consistent results in each case, sufficient information for protein identification was obtained, but this was insufficient for determining the specific lysine or arginine cleavage points in the C-terminal tail.

The resulting MALDI data were searched in Mascot (Perkins et al 1999) for peptides that would have resulted from tryptic digestion of H5. There were only five bands, which showed any evidence of H5 (see table 4.1), and identification of specified

regions of the protein was not possible as the peptides identified were not indicative of the C-terminal tail region of the protein.

This experiment was repeated 3 times with similar results from each occasion. The in-gel digestion procedure was no more effective at identifying the digest products than the technique for protein extraction from gels.

There is some evidence, personal communication McIntyre, S. data unpublished, to show that the longer the gel is stained, in this case the gels were always stained overnight, and the longer the protein/peptide remains in the gel, the more difficult it is to extract from the SDS/PAGE gel. The type of stain used for visualising the gel also has an impact on the efficiency recovery, personal communication McIntyre, S. data unpublished.

The extraction process is also inefficient, Hart S. personal communication (2005), in work with P^{32} labelled phosphopeptides, measuring radioactivity before and after in-gel digestion and extraction indicated significant residual radioactivity after extraction, implying inefficient extraction of the peptides from the gel. There was also the observation that these results were related to the sequence of the peptide. The radioactivity persisted after two extractions of the gel slices. (Zhou et al. 2005) show loss of radioactively labelled proteins from 2-D gel analysis. The losses were cumulative and dependent on the amount of protein loaded. The greater the amount of protein, the greater the percentage lost, 100μg protein loaded resulted in at 69% loss, 1mg of protein loaded resulted in 79% loss.

The work for peptide identification from gel extraction was successful only in the first instance. All repeated experiments, using a variation of the protocol from (Cohen et al.1997), failed to yield any conclusive data. The data from the successful extraction resulted in poor resolution of the observed peaks, see figures 4.26-4.42. This means that it is only possible to infer a decrease in the size of the protein, it is not possible to identify the sites of cleavage.

4.7 Chromatography

Reversed-phase high-pressure liquid chromatography, HPLC was employed as an addition means of identifying the H5 digest products. This technique separates the analytes by using a nonpolar stationary phase, C7 (packing material of the column is comprised of a 7 carbon chain, C7), in conjunction with a polar mobile phase that is

largely aqueous. HPLC requires the use of elevated pressures to force the liquid through a packed bed of the stationary phase. Neue (1997) Gradient chromatography, where there is a continuous variation of the mobile phase composition from low elution strength to high elution strength, was employed for this work. The purpose of this experiment was to perform the digestion, remove an aliquot and analyze by MALDI the remaining sample would be run on SDS/PAGE. But as the fractions collected from the column were 500µl and as can be seen in figure 4.24 were still comprised of multiple components.

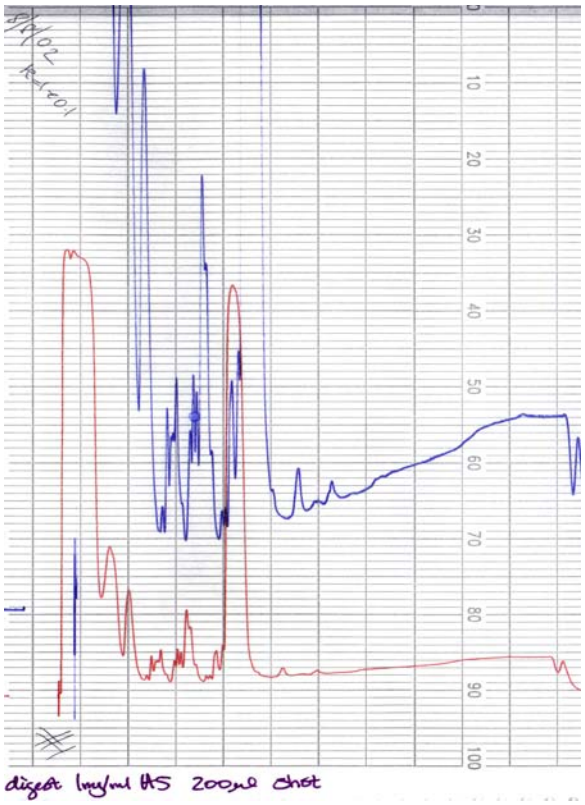

Figure 4.22 Chromatogram from 1mg/ml digest of H5 (48.5µM H5). 200µl of the digest was loaded onto a C7 column and run with a linear gradient from 10% acetonitrile (ACN), 0.1% triflouroacetic acid (TFA) to 100% ACN/0.1% TFA over 60 minutes. 500µl fractions were manually collected. The chromatogram in blue is the full scale, in red the output has been shown at 1:10 decrease of intensity, bringing the peaks into scale.

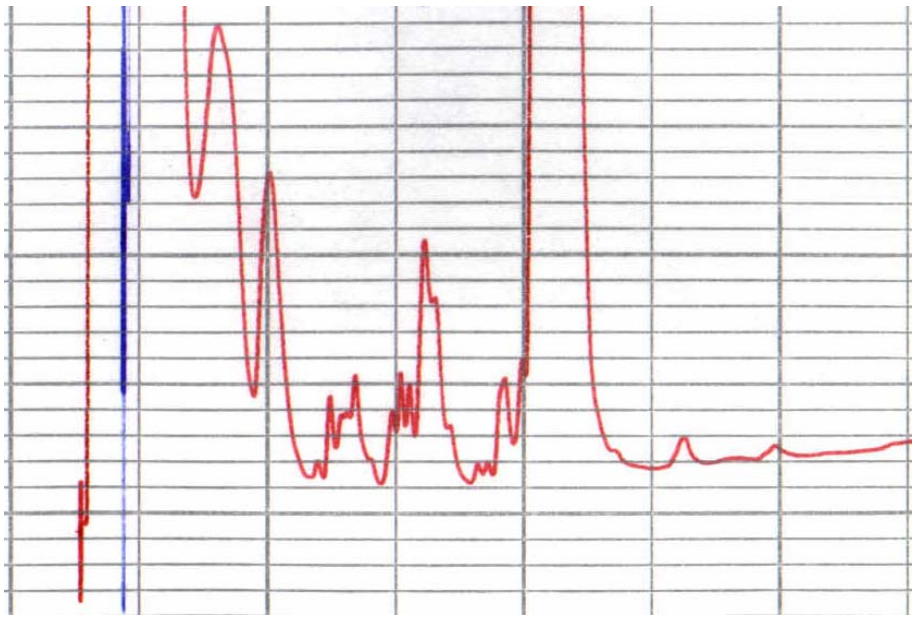

Figure 4.23 expansion of chromatogram in figure 4.25.

The fractions were collected and the sample dried down under vacuum and reconstituted in a 50/50 water/acetonitrile solution with 0.1% formic acid. This was then analyzed by MALDI. Spectra from four fractions are shown in figure 4.24 below. As is seen in figure 4.24, there was insufficient separation of the peptides; there is still a complex mixture of components in the individual fractions.

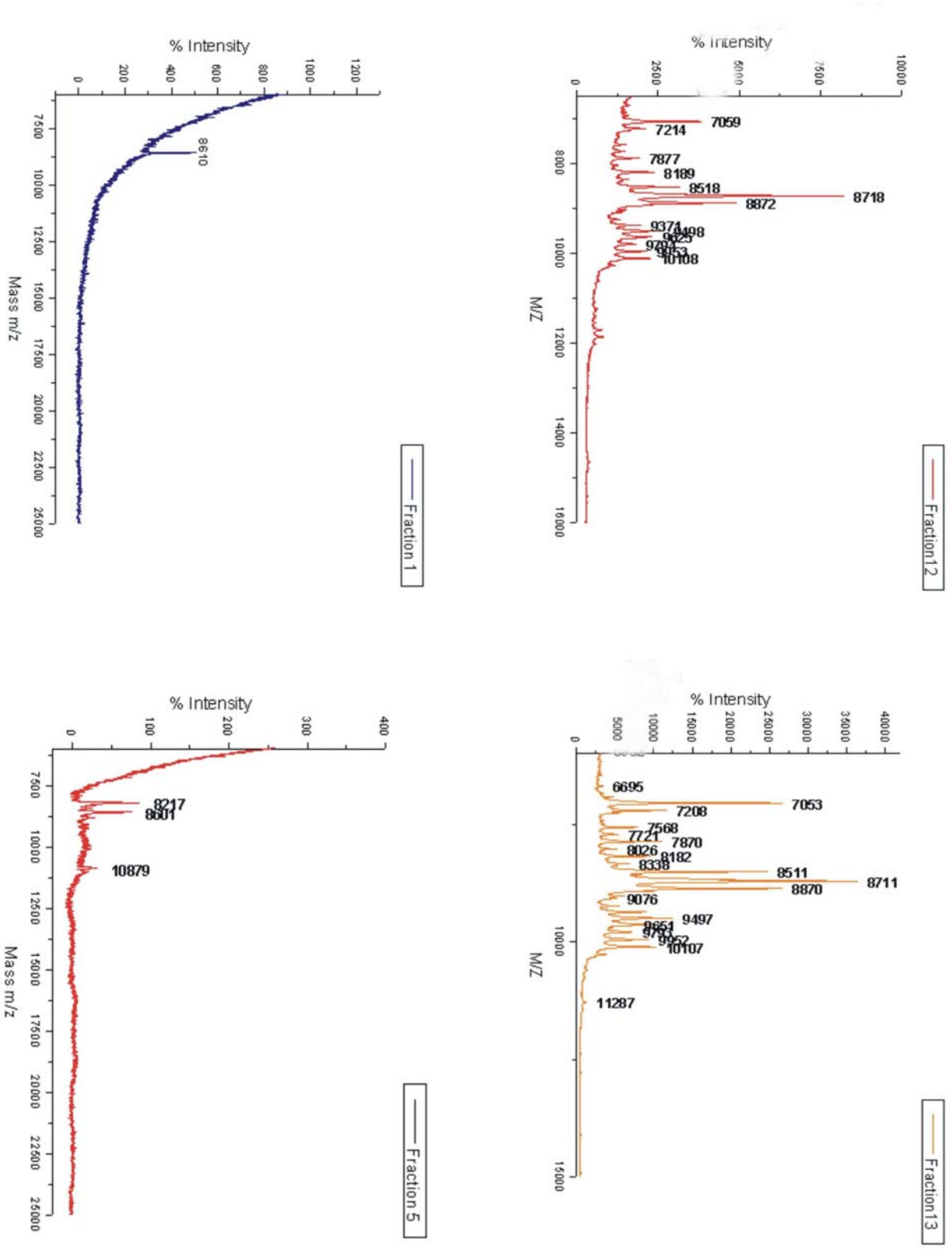

Figure 4.24 Selected fractions; 1, 5, 12 and 13, from separation of 1mg/ml H5 digest. Fractions collected from separation on C7 column were dried down and reconstituted with 50/50/0.1% ACN/H_2O/TFA. 2µl of each reconstituted fraction was mixed with 2µl sinapinic acid.

4.8 Elution of protein products from SDS/PAGE gels

The following samples are from a preparation which is a variation of the method set out by Cohen et al. (1997). This work was intended to be the crux of the material for identification of the peptides and as a result was repeated numerous times, but the elution from the gel was successful in only the following instance. The calculated average mass of the protein H5 is 20602 (accurate mass is calculated to be 20590).As can be seen in the mass spectrum, see figures 4.26 – 4.42, accurate mass assignment of the peptide was not possible. But, a rough assignment has been given; see tables 4.2 and 4.3, as it is possible to distinguish a trend of decreasing mass in the digestion products of the gel bands as you move down the gel. For these data it was not possible to identify possible specific cleavage sites in the H5 tails, the identification of the masses is not accurate as the width of the peak is in many cases greater than 1000 m/z wide. MALDI analysis conditions are outlined in the Methods and Materials section 3.16.

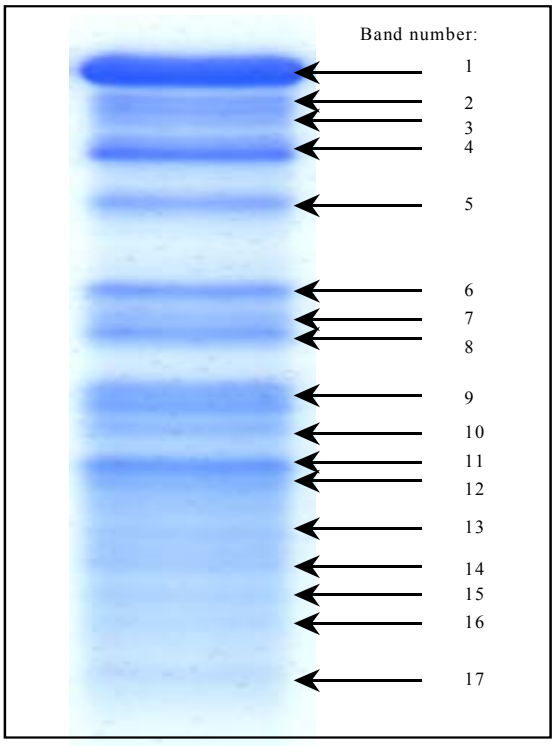

Figure 4.25 SDS/PAGE gel of 1mg/ml digestion of H5, numbers along side indicate band numbers excised for mass analysis. Mass spectra follow in figure 4.26 – figure 4.42 and tables 4.1 and 4.2 summarise the mass information identified in the mass spectra.

Band #	Mass observed
1	22894
2	22045
3	21182
4	19807
5	19265
6	None observed
7	18684
8	None observed
9	16505
10	None observed
11	15922
12	15355
13	None observed
14	14345
15	13133
16	See table 4.5
17	12647

Table 4.2 Masses observed from gel extraction MALDI spectra, figures 4.26 – figure 4.42, bands 1 through 17 of gel in figure 4.25 with band 16 (from figure 4.25) occupying a separate table, table 4.2.

Band 16	Masses observed
	15549
	15546
	14722
	14300
	13483
	13257
	12802
	12723
	11191
	11108
	11089
	11070
	11032
	10862

Table 4.3 Masses observed for Band 16 from figure 4.25. Despite the lack of a distinct band present on the gel, this sample identified many more digestion products than the more obvious bands on the gel.

From table 4.2 we can see that while band 1 from the gel in figure 4.25 should be H5 where the observed mass by MALDI should be 20603 $(M + H)^+$, instead the mass observed is approximately 22894. Given that the width of the peak is greater than 5,000 m/z making an accurate assessment of the mass of the elution product from this data would be impossible. This could be due to the protein having taken on

adducts of the polyacrylamide from the gel. The remaining elution products also fail to identify specific masses due either to the width of the peaks making an accurate assessment of the mass impossible (see figures 4.27, 4.28, 4.29, 4.30, 4.32 4.34, 4.37, 4.39, 4.40 and 4.42) or because there is no observed peak in the spectrum (see figures 4.31, 4.33, 4.35 and 4.38). The exception is the spectrum for band 16 as shown in table 4.3 and figure 4.41. There are multiple identifications to be made from this spectrum despite what appears in the gel, see figure 4.25, to be a faintly stained indistinct band.

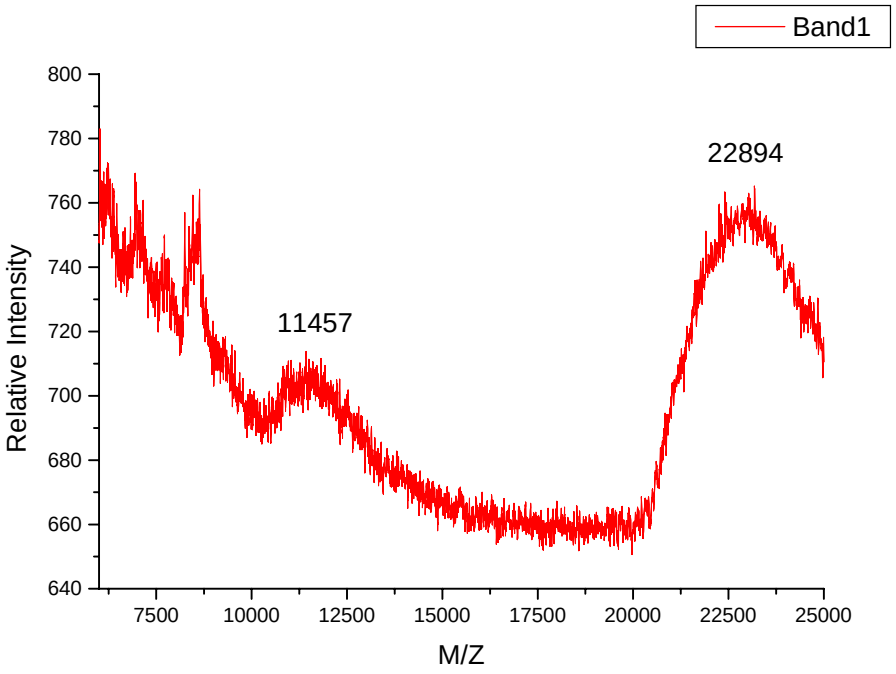

Figure 4.26 Mass spectrum of band 1 excised and eluted from gel in figure 4.28 and analyzed by MALDI.

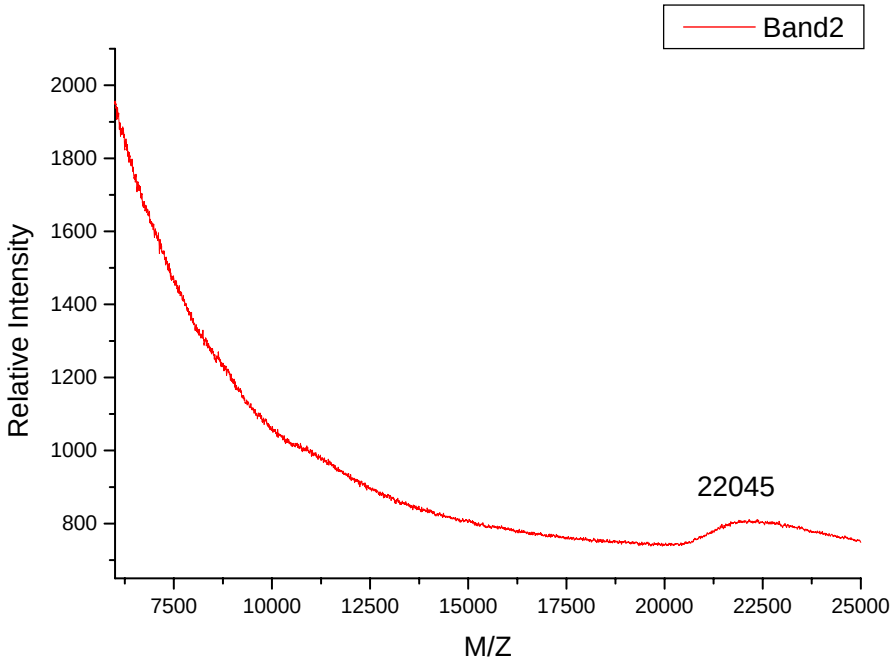

Figure 4.27 Mass spectrum of band 2 excised and eluted from gel in figure 4.28 and analyzed by MALDI.

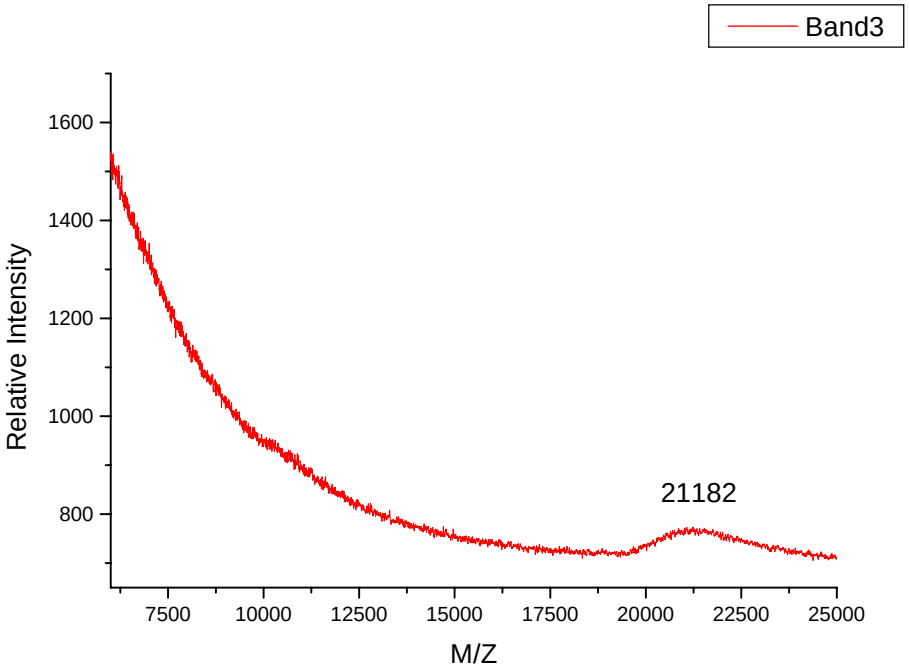

Figure 4.28 Mass spectrum of band 3 excised and eluted from gel in figure 4.28 and analyzed by MALDI.

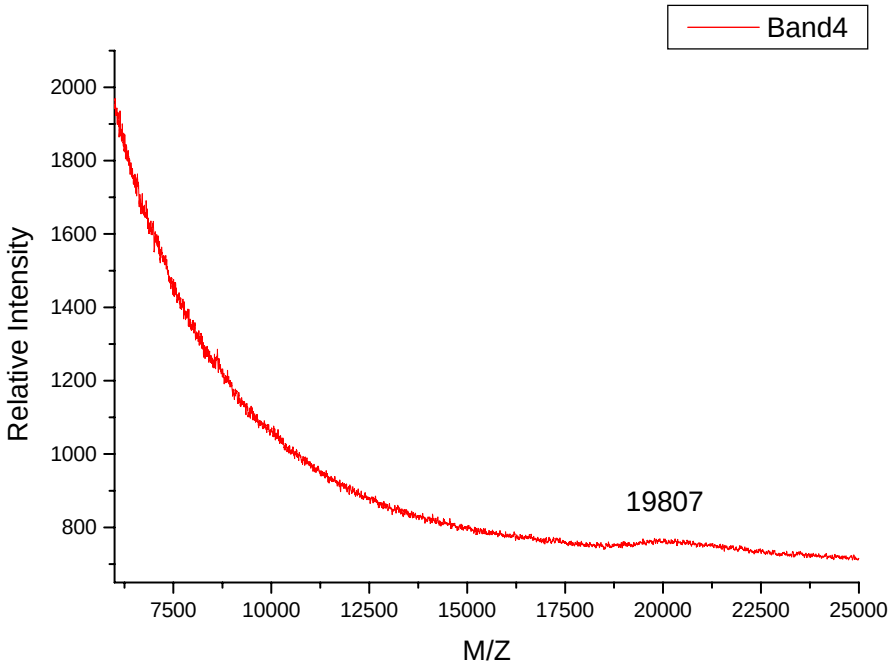

Figure 4.29 Mass spectrum of band 4 excised and eluted from gel in figure 4.28 and analyzed by MALDI.

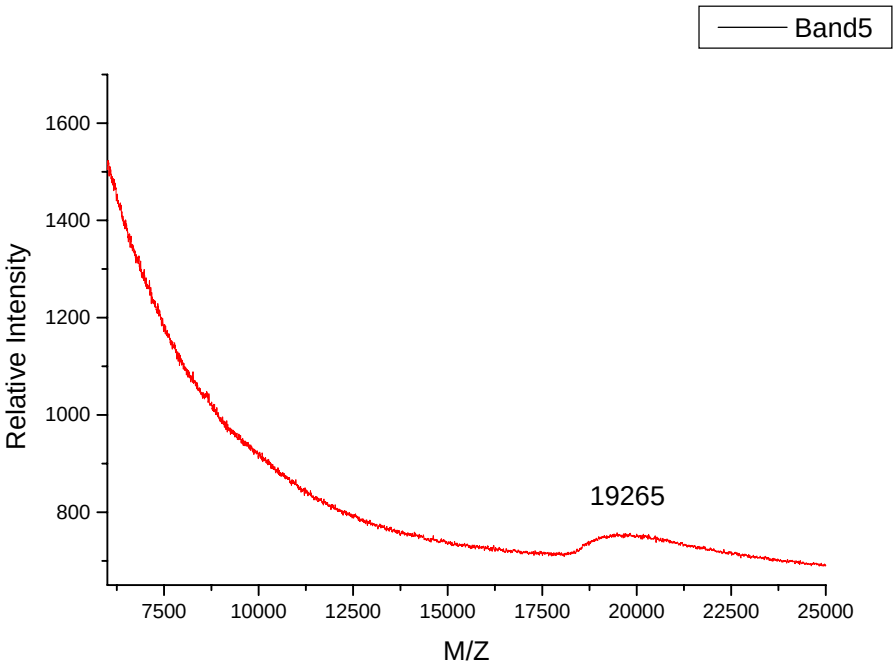

Figure 4.30 Mass spectrum of band 5 excised and eluted from the gel in figure 4.28 and analyzed by MALDI.

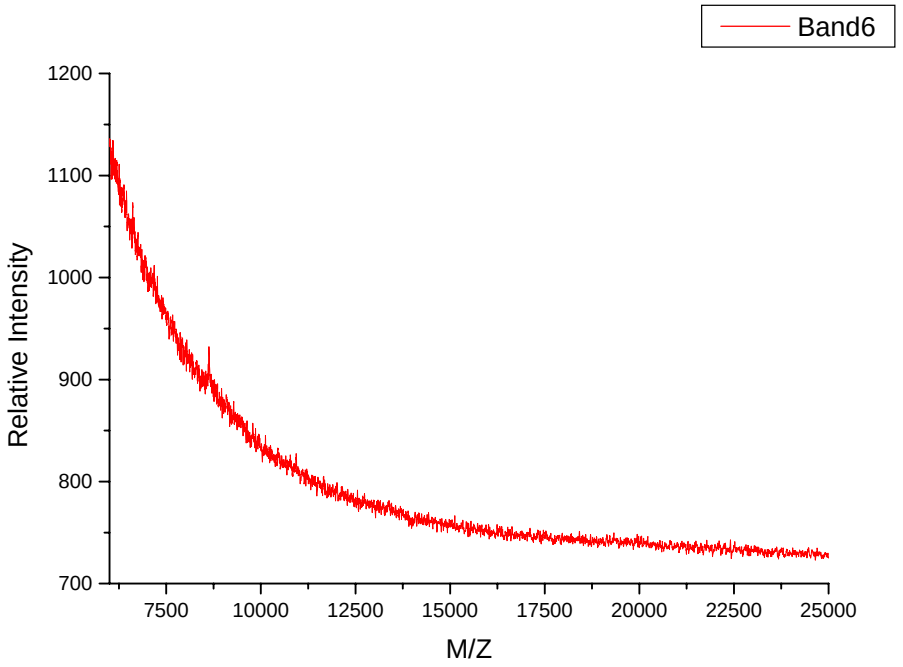

Figure 4.31 Mass spectrum of band 6 excised and eluted from the gel in figure 4.28 and analyzed by MALDI.

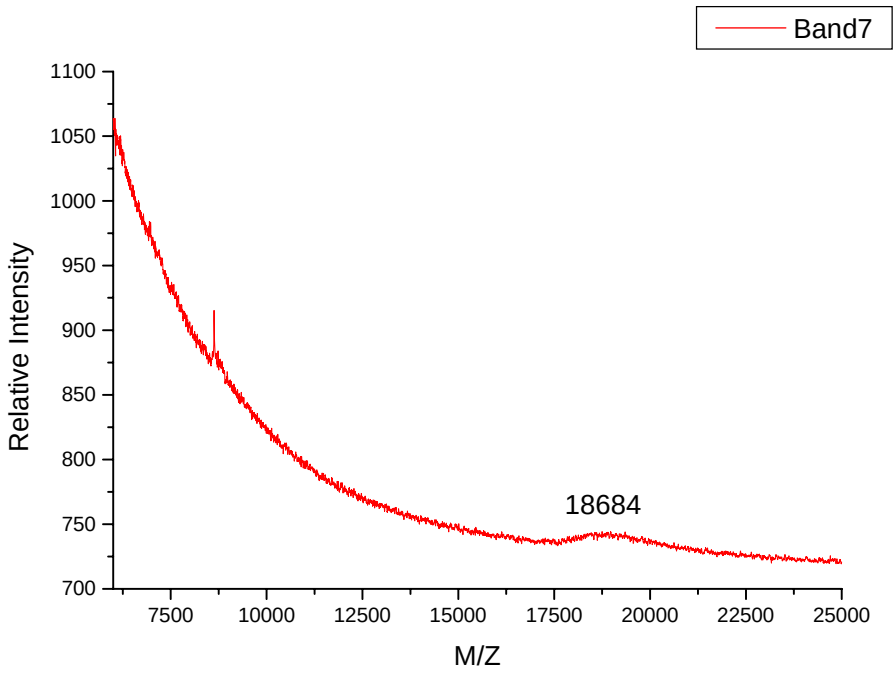

Figure 4.32 Mass spectrum of band 7 excised and eluted from the gel in figure 4.28 and analyzed by MALDI.

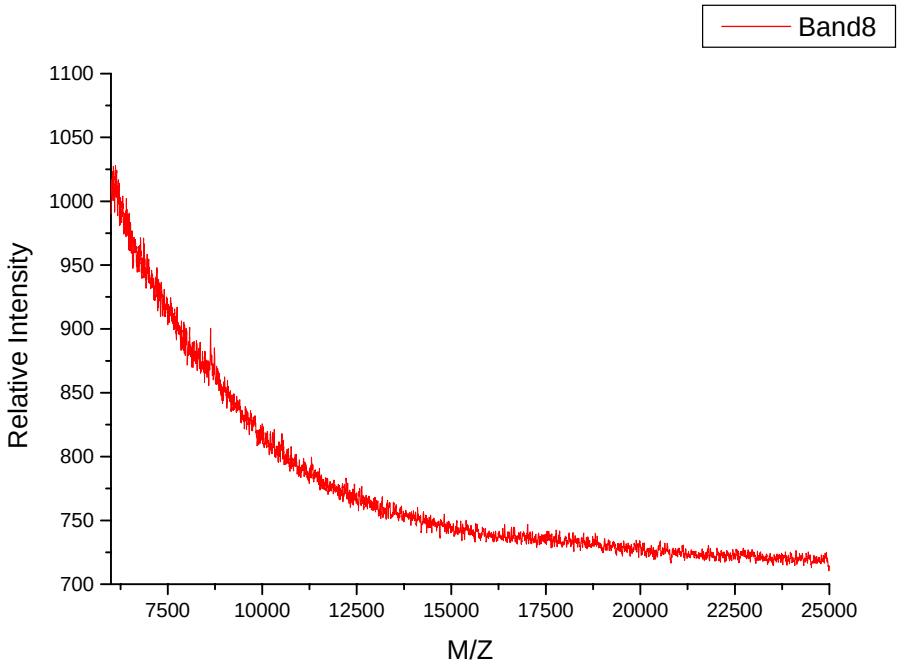

Figure 4.33 Mass spectrum of band 8 excised and eluted from the gel in figure 4.28 and analyzed by MALDI.

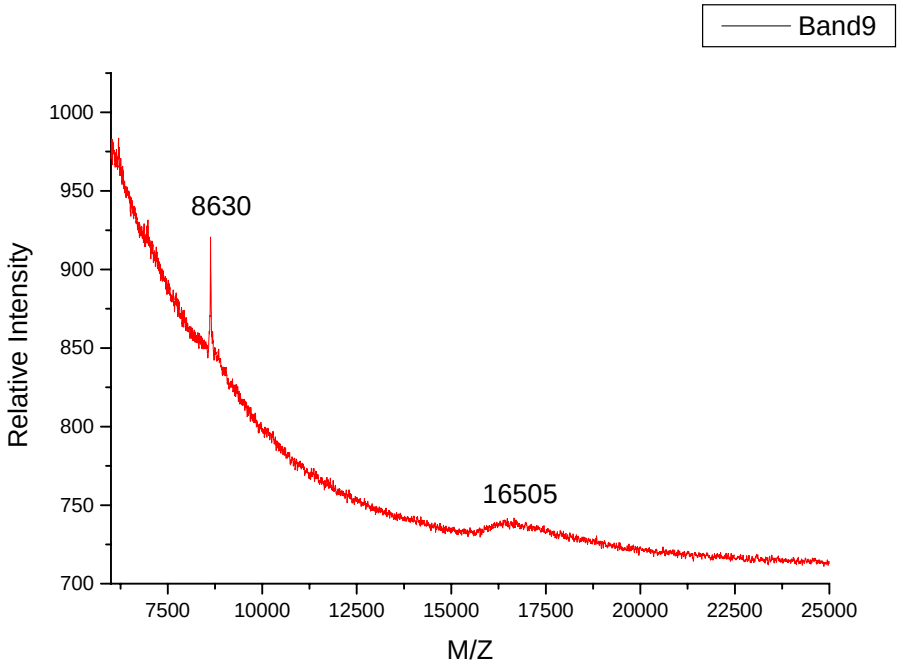

Figure 4.34 Mass spectrum of band 9 excised and eluted from the gel in figure 4.28 and analyzed by MALDI.

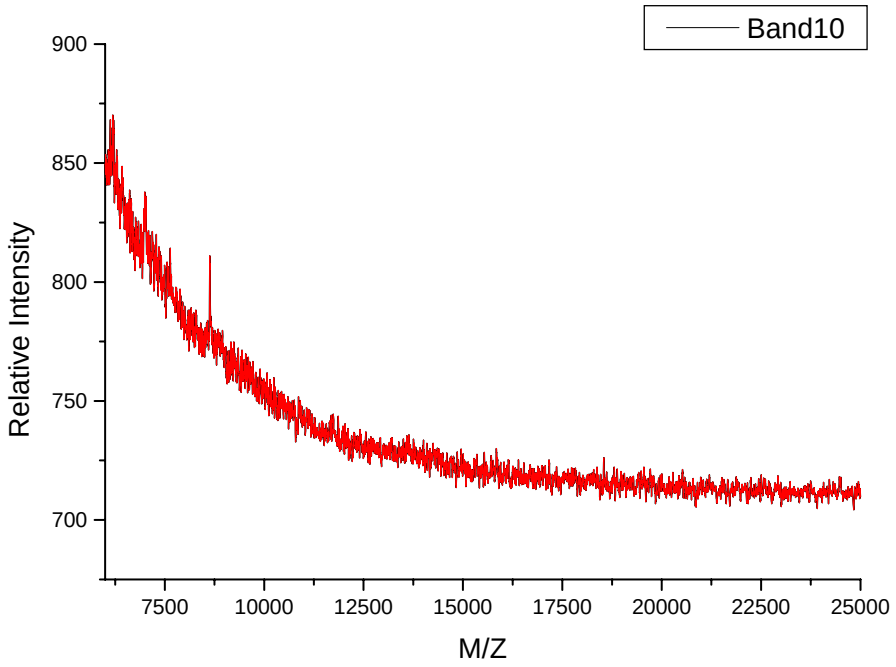

Figure 4.35 Mass spectrum of band 10 excised and eluted from the gel in figure 4.28 and analyzed by MALDI.

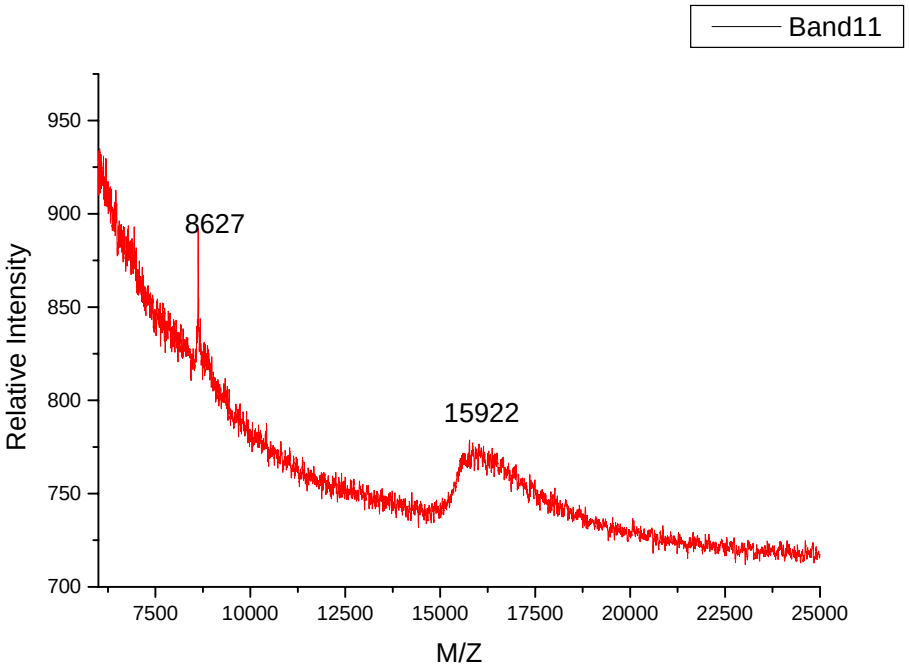

Figure 4.36 Mass spectrum of band 11 excised and eluted from the gel in figure 4.28 and analyzed by MALDI.

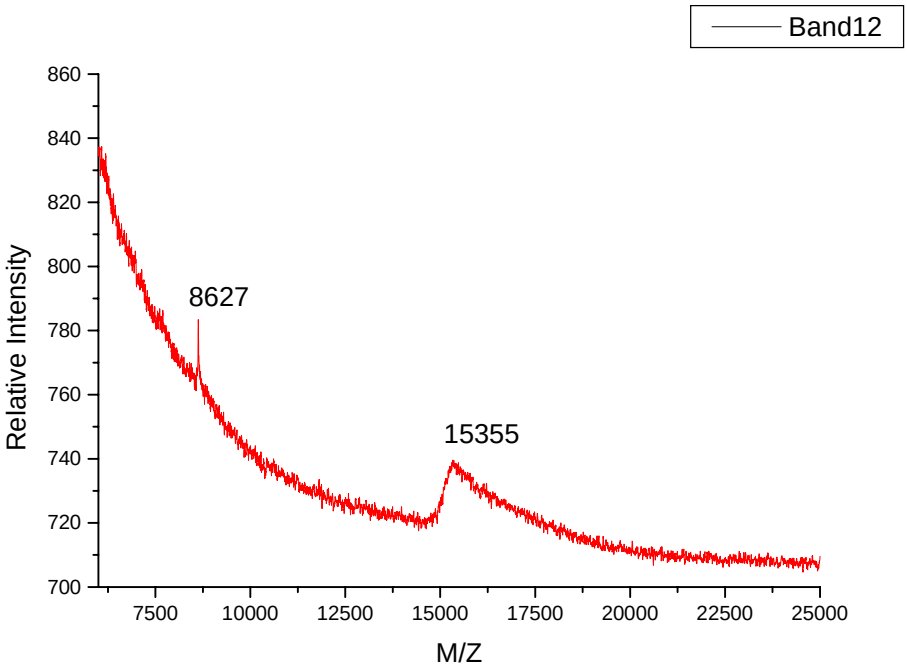

Figure 4.37 Mass spectrum of band 12 excised and eluted from the gel in figure 4.28 and analyzed by MALDI.

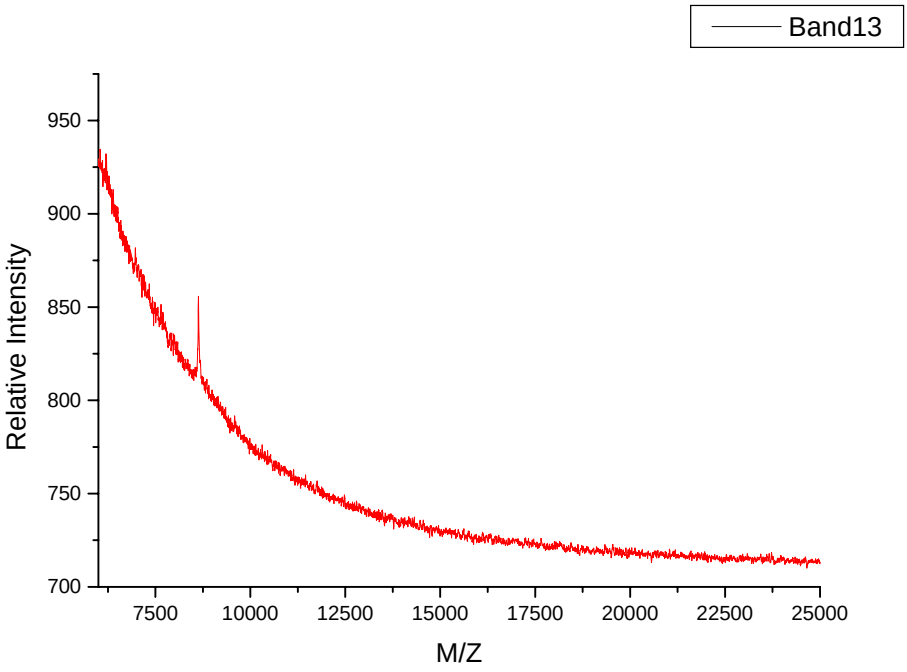

Figure 4.38 Mass spectrum of band 13 excised and eluted from the gel in figure 4.28 and analyzed by MALDI.

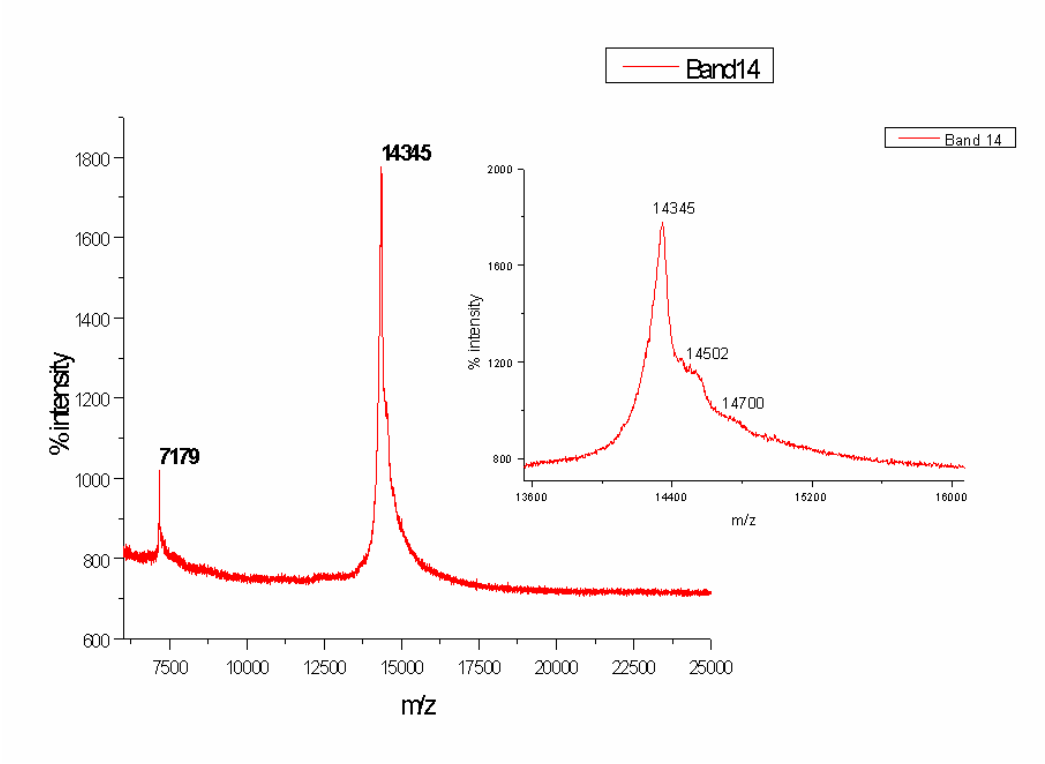

Figure 4.39 Mass spectrum of band 14 excised and eluted from the gel in figure 4.28 and analyzed by MALDI.

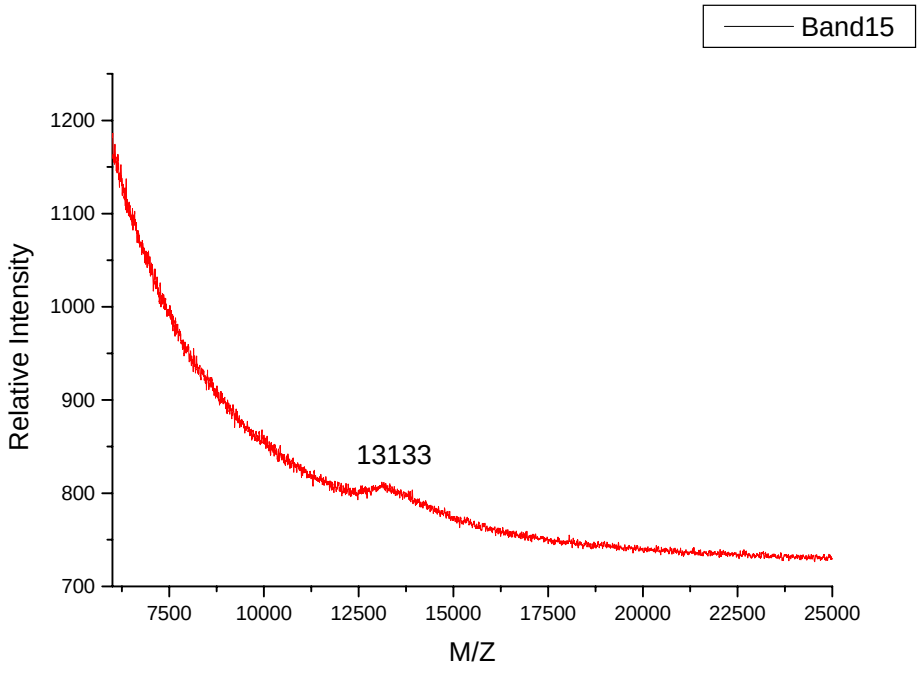

Figure 4.40 Mass spectrum of band 15 excised and eluted from the gel in figure 4.28 and analyzed by MALDI.

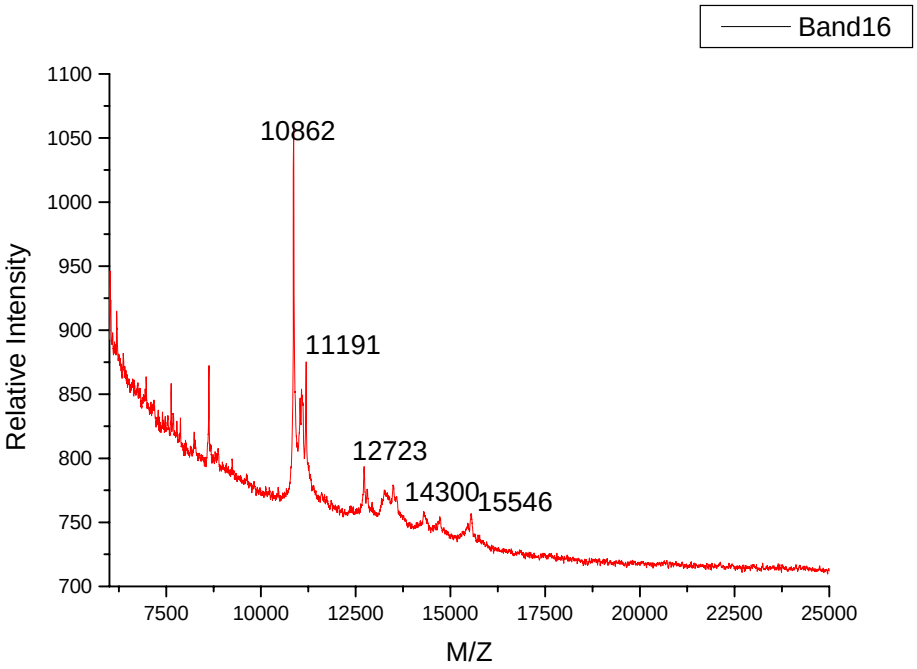

Figure 4.41 Mass spectrum of band 16 excised and eluted from the gel in figure 4.28 and analyzed by MALDI.

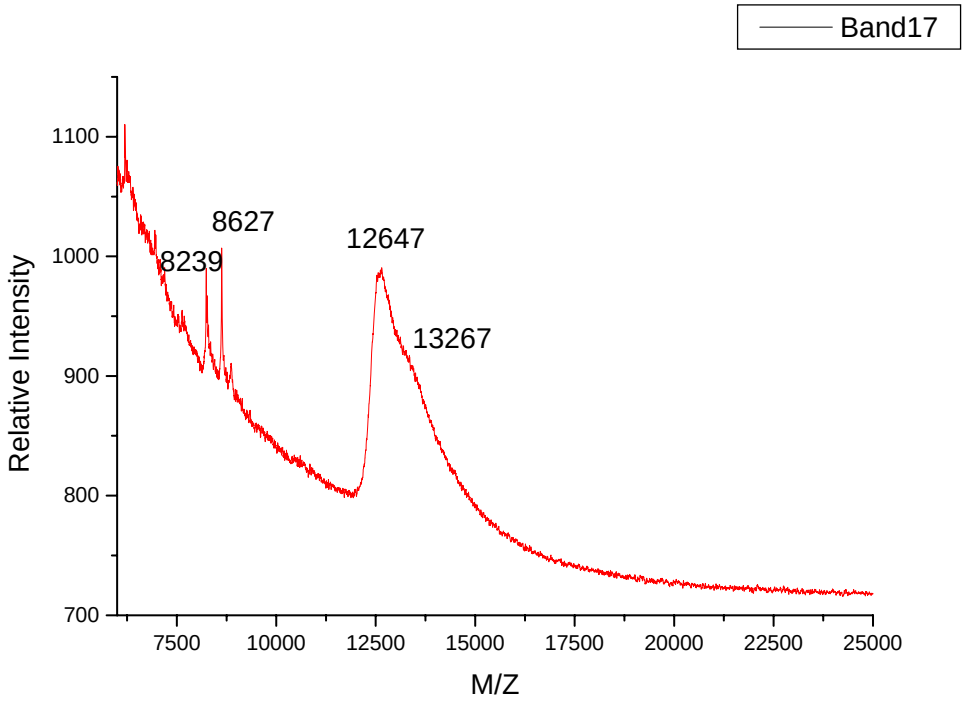

Figure 4.42 Mass spectrum of band 17 excised and eluted from the gel in figure 4.28 and analyzed by MALDI.

The above spectra, figures 4.26-4.42, represent the data obtained from the extraction of the peptides from the gel. With the resolution obtained in these data it would not be possible to identify specifically what the mass is of the peptide and therefore specifically where the cleavage has occurred. But the information obtained does allow for the inference of cleavage sites in the N- and C-terminal tails of the H5. Had it been possible to repeat the experiment and obtain more accurate mass assignments, the SDS/PAGE gels could have been calibrated for the H5 digest products. This would as well have allowed for a better understanding of the relative sensitivies of the lysines and arginines in the C-terminal tail of the H5.

4.9 MALDI Analysis of in-solution digest

In order to identify the digest products MALDI mass spectrometry was used to analyse aliquots from individual time points. The aliquots were mixed with matrix and spotted onto the MALDI target plate. The following figures, 4.43, 4.44, 4.45 and 4.46, are the MALDI spectra of three selected digest time points. The conditions for analysis are given in the Methods and Materials section 3.16.

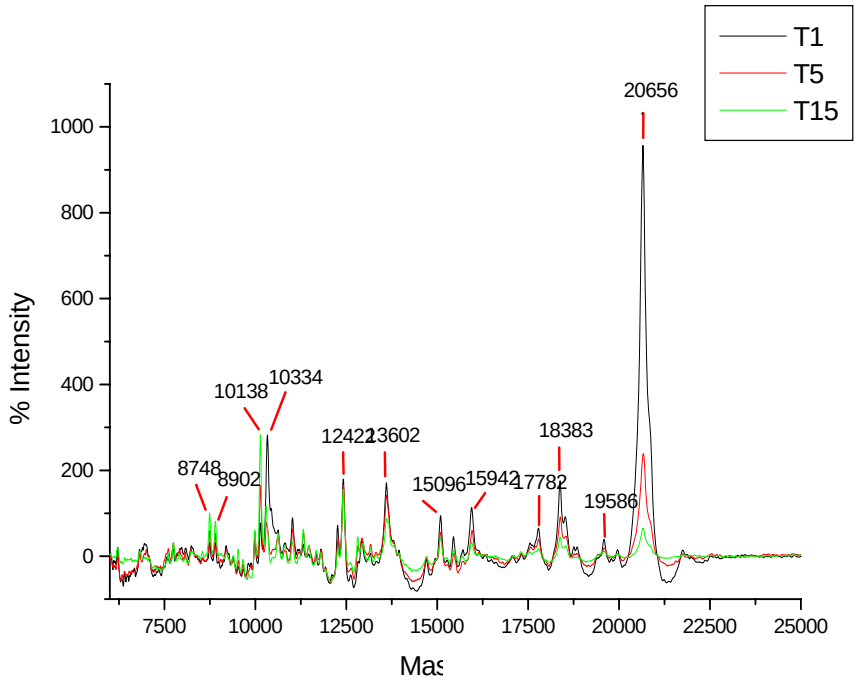

Figure 4.43 Complete mass range analyzed by MALDI for three time points aliquots.

Figure 4.46 shows the time points T1 (black), T5 (red) and T15 (green) together on one spectrum in order to demonstrate the digestion pattern of the H5 over time. The

following figures show expansions of selected mass ranges in order to view the individual peaks more closely. Table 4.4 takes the information from the spectra and assigns sequence from the H5 to the masses observed.. MALDI analysis of the digest products at this stage was used to attempt identification of the digest products. While this method of mass analysis identifies the digest fragments it is still necessary to identify when in the digestion process the fragments are generated. With these data we can infer some information regarding when they are generated from the changes in intensity of the peaks, but it is not possible to take these mass assignments back to the gel and identify the masses of the individual bands.

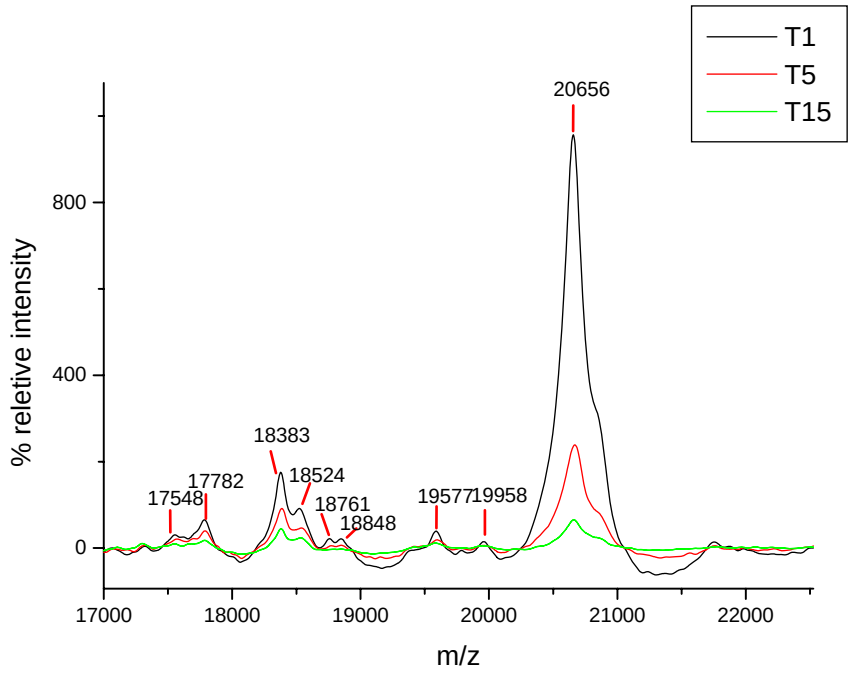

Figure 4.44 Expansion of figure 4.46 upper mass range 17000m/z to 22500 with masses identified.

In figure 4.47 the upper mass range of the spectrum is expanded and the peaks identified. In this figure the change in the intensity of a peak is demonstrated. If we consider the peak at m/z 20656, at T1 it has a relative intensity of ~1,000 but by T5 its relative intensity has dropped to ~200. This peak represents the H5 as a complete protein and its decrease in the intensity is best described as being caused by the digestion of the H5. By T15 the relative intensity of the H5 peak has decreased to less than 100. As the ion at m/z 20656 decreases in relative intensity the peaks in the lower mass range, see figure 3, grow in relative intensity.

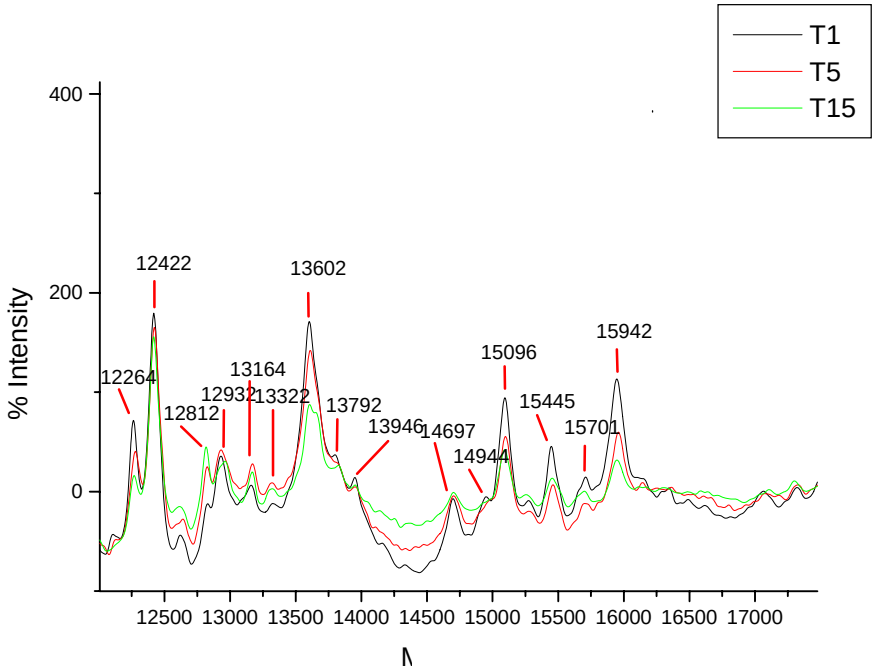

Figure 4.45 Expansion of figure 4.46 mid range of mass analyzed; 12000 m/z to 17300 m/z.

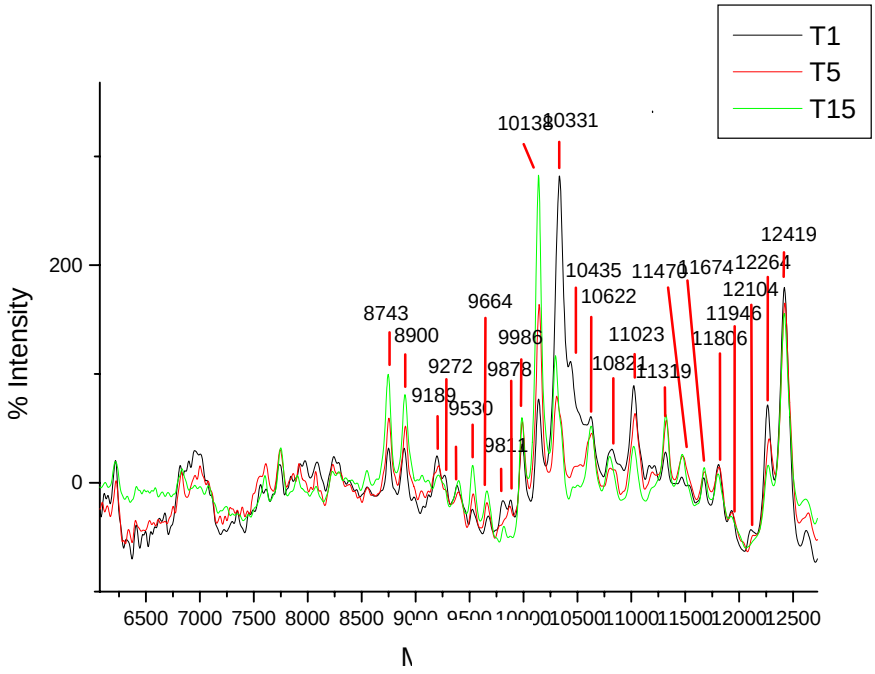

Figure 4.46 Expansion of figure 4.46 lower mass range 6000 m/z to 12600 m/z.

Figures 4.43, 4.44, 4.45 and 4.46 MALDI mass spectra of H5 time course digest. Three time points, T1, T5 and T15, are overlaid to show changes in intensity of

digest products. Figure 4.46 is the full mass range of selected time points, figure 4.47 is an expansion of mass range 8500-12500 and figure 4.48 is an expansion of mass range 14700-19960. In figure 1 the intensity of m/z 20667 (H5) decreases as the digestion progresses. In Figure 4.47 two ions m/z 10138 and 11022, exemplify the change of intensity in the peaks as the digestion progresses. Some of the masses appear to maintain the same intensity across these time points, see figure 4.48 mass 12419 and figure 4.47 mass 12422.

The masses observed in the spectrum are listed in table 4.4 and a sequence which most closely correspond to sequence from the H5 protein. The numbers in column 2 are the amino acid number from the H5 sequence; 1:183, for example, indicates that mass 19958 is assigned to amino acids 1 to 183 of the H5 sequence.

Observed mass	Amino acids	Calculated mass	Δm	Sequence of suggested (calculated) peptide
20656	H5	20602	54	For full sequence see figure 4.5
19958	1:183	19893	65	TESLVLSPAPAKPKRVKASRRSASHPTYSE MIAAAIRAEKSRGGSSRQSIQKYIKSHYKV GHNADLQIKLSIRRLLAAGVLKQTKGVGAS GSFRLAKSDKAKRSPGKKKKAVRRSTSPKK AARPRKARSPAKKPKATARKARKKSRASPK KAKKPKTVKAKSRKASKAKKVKRSKPRAKS GAR
19577	1:179	19522	55	TESLVLSPAPAKPKRVKASRRSASHPTYSE MIAAAIRAEKSRGGSSRQSIQKYIKSHYKV GHNADLQIKLSIRRLLAAGVLKQTKGVGAS GSFRLAKSDKAKRSPGKKKKAVRRSTSPKK AARPRKARSPAKKPKATARKARKKSRASPK KAKKPKTVKAKSRKASKAKKVKRSKPRAK
18848	17:189	18788	60	KASRRSASHPTYSEMIAAAIRAEKSRGGSS RQSIQKYIKSHYKVGHNADLQIKLSIRRLL AAGVLKQTKGVGASGSFRLAKSDKAKRSPG KKKKAVRRSTSPKKAARPRKARSPAKKPKA TARKARKKSRASPKKAKKPKTVKAKSRKAS KAKKVKRSKPRAKSGARKSPKKK
18761				No specified sequence
18524	1:170	18471	53	TESLVLSPAPAKPKRVKASRRSASHPTYSE MIAAAIRAEKSRGGSSRQSIQKYIKSHYKV GHNADLQIKLSIRRLLAAGVLKQTKGVGAS GSFRLAKSDKAKRSPGKKKKAVRRSTSPKK AARPRKARSPAKKPKATARKARKKSRASPK KAKKPKTVKAKSRKASKAKK
18383	1:169	18343	40	TESLVLSPAPAKPKRVKASRRSASHPTYSE MIAAAIRAEKSRGGSSRQSIQKYIKSHYKV

				GHNADLQIKLSIRRLLAAGVLKQTKGVGAS GSFRLAKSDKAKRSPGKKKKAVRRSTSPKK AARPRKARSPAKKPKATARKARKKSRASPK KAKKPKTVKAKSRKASKAK
17782	1:163	17858	-76	TESLVLSPAPAKPKRVKASRRSASHPTYSE MIAAAIRAEKSRGGSSRQSIQKYIKSHYKV GHNADLQIKLSIRRLLAAGVLKQTKGVGAS GSFRLAKSDKAKRSPGKKKKAVRRSTSPKK AARPRKARSPAKKPKATARKARKKSRASPK KAKKPKTVKAKSR
17548	1:161	17487	61	TESLVLSPAPAKPKRVKASRRSASHPTYSE MIAAAIRAEKSRGGSSRQSIQKYIKSHYKV GHNADLQIKLSIRRLLAAGVLKQTKGVGAS GSFRLAKSDKAKRSPGKKKKAVRRSTSPKK AARPRKARSPAKKPKATARKARKKSRASPK KAKKPKTVKAK
15942	1:146	15896	46	TESLVLSPAPAKPKRVKASRRSASHPTYSE MIAAAIRAEKSRGGSSRQSIQKYIKSHYKV GHNADLQIKLSIRRLLAAGVLKQTKGVGAS GSFRLAKSDKAKRSPGKKKKAVRRSTSPKK AARPRKARSPAKKPKATARKARKKSR
15701	1:144	15653	48	TESLVLSPAPAKPKRVKASRRSASHPTYSE MIAAAIRAEKSRGGSSRQSIQKYIKSHYKV GHNADLQIKLSIRRLLAAGVLKQTKGVGAS GSFRLAKSDKAKRSPGKKKKAVRRSTSPKK AARPRKARSPAKKPKATARKARKK
15445	1:142	15396	49	TESLVLSPAPAKPKRVKASRRSASHPTYSE MIAAAIRAEKSRGGSSRQSIQKYIKSHYKV GHNADLQIKLSIRRLLAAGVLKQTKGVGAS GSFRLAKSDKAKRSPGKKKKAVRRSTSPKK AARPRKARSPAKKPKATARKAR
15096	1:139	15041	55	TESLVLSPAPAKPKRVKASRRSASHPTYSE MIAAAIRAEKSRGGSSRQSIQKYIKSHYKV GHNADLQIKLSIRRLLAAGVLKQTKGVGAS GSFRLAKSDKAKRSPGKKKKAVRRSTSPKK AARPRKARSPAKKPKATAR
14944	Small signal			Very small signal
14697	13:146	14702	-5	PKRVKASRRSASHPTYSEMIAAAIRAEKSR GGSSRQSIQKYIKSHYKVGHNADLQIKLSI RRLLAAGVLKQTKGVGASGSFRLAKSDKAK RSPGKKKKAVRRSTSPKKAARPRKARSPAK KPKATARKARKKSR
13946	1:128	13905	41	TESLVLSPAPAKPKRVKASRRSASHPTYSE MIAAAIRAEKSRGGSSRQSIQKYIKSHYKV GHNADLQIKLSIRRLLAAGVLKQTKGVGAS GSFRLAKSDKAKRSPGKKKKAVRRSTSPKK AARPRKAR

13792				No specified sequence
13602	1:125	13550	52	TESLVLSPAPAKPKRVKASRRSASHPTYSE MIAAAIRAEKSRGGSSRQSIQKYIKSHYKV GHNADLQIKLSIRRLLAAGVLKQTKGVGAS GSFRLAKSDKAKRSPGKKKKAVRRSTSPKK AARPR
13321	1:123	13297		TESLVLSPAPAKPKRVKASRRSASHPTYSE MIAAAIRAEKSRGGSSRQSIQKYIKSHYKV GHNADLQIKLSIRRLLAAGVLKQTKGVGAS GSFRLAKSDKAKRSPGKKKKAVRRSTSPKK AAR
13164	22:142	13124		RSASHPTYSEMIAAAIRAEKSRGGSSRQSI QKYIKSHYKVGHNADLQIKLSIRRLLAAGV LKQTKGVGASGSFRLAKSDKAKRSPGKKKK AVRRSTSPKKAARPRKARSPAKKPKATARK ARKK
12932	22:140	12897		RSASHPTYSEMIAAAIRAEKSRGGSSRQSI QKYIKSHYKVGHNADLQIKLSIRRLLAAGV LKQTKGVGASGSFRLAKSDKAKRSPGKKKK AVRRSTSPKKAARPRKARSPAKKPKATARK
12812	22:139	12769		RSASHPTYSEMIAAAIRAEKSRGGSSRQSI QKYIKSHYKVGHNADLQIKLSIRRLLAAGV LKQTKGVGASGSFRLAKSDKAKRSPGKKKK AVRRSTSPKKAARPRKARSPAKKPKATAR
12419	1:114	12371		TESLVLSPAPAKPKRVKASRRSASHPTYSE MIAAAIRAEKSRGGSSRQSIQKYIKSHYKV GHNADLQIKLSIRRLLAAGVLKQTKGVGAS GSFRLAKSDKAKRSPGKKKKAVRR
12264	1:113	12214		TESLVLSPAPAKPKRVKASRRSASHPTYSE MIAAAIRAEKSRGGSSRQSIQKYIKSHYKV GHNADLQIKLSIRRLLAAGVLKQTKGVGAS GSFRLAKSDKAKRSPGKKKKAVR
12104	21:132	12172		RSASHPTYSEMIAAAIRAEKSRGGSSRQSI QKYIKSHYKVGHNADLQIKLSIRRLLAAGV LKQTKGVGASGSFRLAKSDKAKRSPGKKKK AVRRSTSPKKAARPRKARSPAK
11946	1:110	11888		TESLVLSPAPAKPKRVKASRRSASHPTYSE MIAAAIRAEKSRGGSSRQSIQKYIKSHYKV GHNADLQIKLSIRRLLAAGVLKQTKGVGAS GSFRLAKSDKAKRSPGKKKK
11806				No specified sequence
11674	1:108	11632		TESLVLSPAPAKPKRVKASRRSASHPTYSE MIAAAIRAEKSRGGSSRQSIQKYIKSHYKV GHNADLQIKLSIRRLLAAGVLKQTKGVGAS GSFRLAKSDKAKRSPGKK
11470				No specified sequence
11319				No specified sequence
11023	1:102	10979		TESLVLSPAPAKPKRVKASRRSASHPTYSE

				MIAAAIRAEKSRGGSSRQSIQKYIKSHYKV GHNADLQIKLSIRRLLAAGVLKQTKGVGAS GSFRLAKSDKAK
10821	1:100	10779		TESLVLSPAPAKPKRVKASRRSASHPTYSE MIAAAIRAEKSRGGSSRQSIQKYIKSHYKV GHNADLQIKLSIRRLLAAGVLKQTKGVGAS GSFRLAKSDK
10622				No specified sequence
10435				No specified sequence
10331				No specified sequence
10138	21:113	10098		RSASHPTYSEMIAAAIRAEKSRGGSSRQSI QKYIKSHYKVGHNADLQIKLSIRRLLAAGV LKQTKGVGASGSFRLAKSDKAKRSPGKKKK AVR
9986				No specified sequence
9878				No specified sequence
9811				No specified sequence
9664				No specified sequence

Table 4.4: column one lists the masses observed in the MALDI analysis, column 2 gives the sequences which are believed to be related to the mass observed and column 3 are the calculated average masses for the specified sequence of H5 stated in column 2 column 3 indicates the difference in mass between the observed and the calculated and the sequence is shown in column 4. The difference between the calculated and the observed can be attributed to a number of factors including that the chosen sequence is incorrect for the mass observed. For this experiment the sequence with the closest mass to the observed was chosen with the intentions of indicating a trend in the digestion pattern.

From direct MALDI analysis of the individual aliquots, these masses were identified as sequences from the digestion. The overlap of the three time points is to show the changes in the intensity of the peaks allowing us to decide when, in the course of the digestion, the peak occurs. The sites of sensitivity, those with the amino acids lysine and/or arginines which are early sites of cleavage, are accessible to the trypsin for digestion. As these are located predominately on the tails of the H5 this information should help to identify how the H5 tails are interacting with the linker DNA in the 30nM fiber.

4.10 H5 in presence of DNA H5 in the presence of Chromatin

H1 has been shown to interact with linear DNA, Thomas et al. (1986), at salt concentrations above 50mM NaCl to form rods of a uniform shape (11-15nm). This depends on both the salt concentration and the ratio of H1/DNA. Therefore a control digest of H5 in the presence of DNA at a ratio of 1 molecule of H5/200bp DNA in

80mM NaCl was run to determine if there was a significant change in the H5 digestion products. Comparing the H5 digest products from H5 alone with those from H5/DNA it was shown that there was not a significant change in the digestion products as observed by protein gel electrophoresis (compare figure 4.6 to figure 4.50).

The results presented here provide preliminary evidence that the accessibility of H5 tail domains to trypsin alters in the 30nm fiber, compared with H5 reconstituted on DNA or in solution.

In order to determine the cutting pattern of the tails, it is necessary to acquire quantitative information on the digestion products. SDS/PAGE would allow for quantitative determination, but the basic nature of the histones means that they run anomalously by SDS/PAGE. Figure 4.12 shows a comparison of the histone masses with known protein masses as run on a 15% polyacrylamide gel. This information confirms the necessity for an alternative means of determining the masses of the peptide fragments.

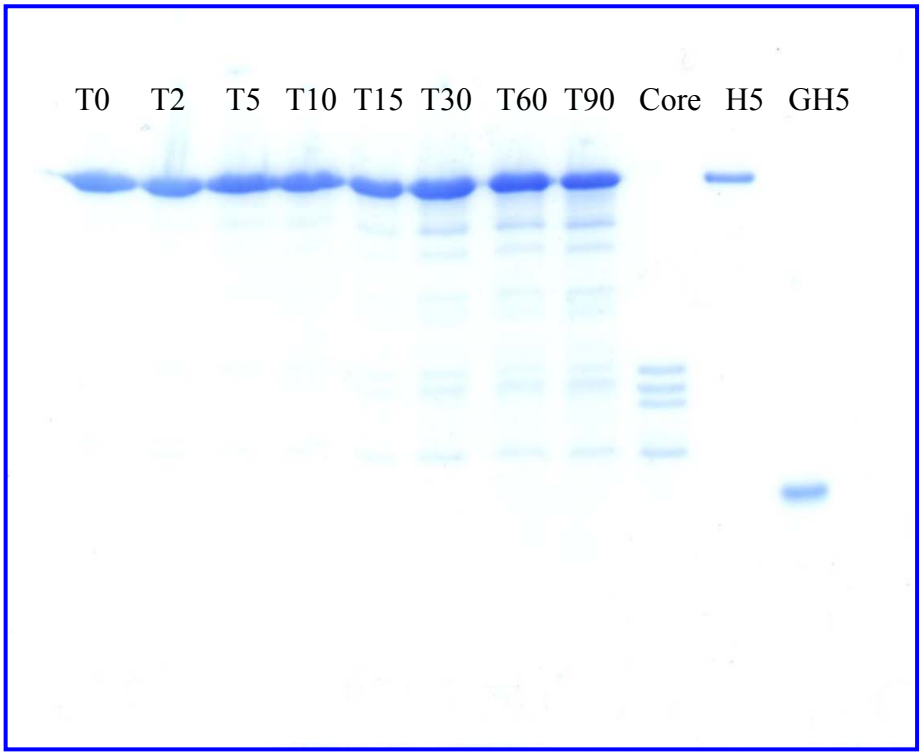

Figure 4.47 H5 time course digest with 1μg/ml trypsin in the presence of DNA. H5/DNA ratio was 1 molecule H5/200bp DNA at 80mM NaCl. H5 was allowed to settle with DNA for 1 hour prior to digestion. The digestion was stopped with 5mM PMSF and the H5 was acetone precipitated, run on protein gel and stained with coomassie brilliant blue.

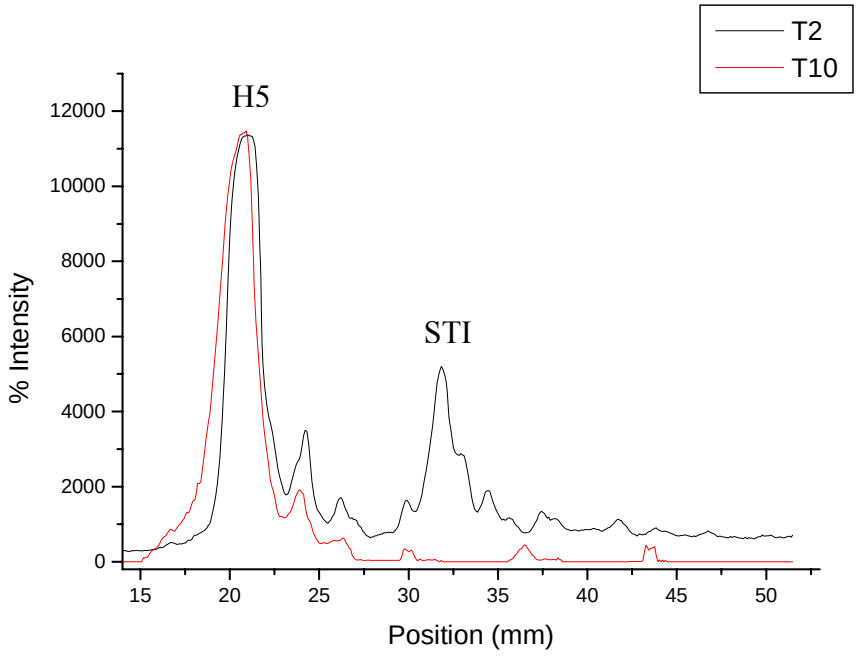

Figure 4.48 AIDA scans from T2 of H5 digest gel and T10 of H5/DNA gel. Both digestions were in TE_{80} the H5 digest is 50µg/ml H5 halted with STI and in the H5/DNA the concentration of H5 is 50µg/ml and the concentration of DNA is 325µg/ml halted with PMSF.

The data in figure 4.52 shows an overlap of different time points from different digestions and different gels. The significance of this data is in the presence of DNA the H5 digestion is similar to the digestion of H5 alone. The concentration of H5 is the same in both digestions and this is reflected in the value of the H5 peak.

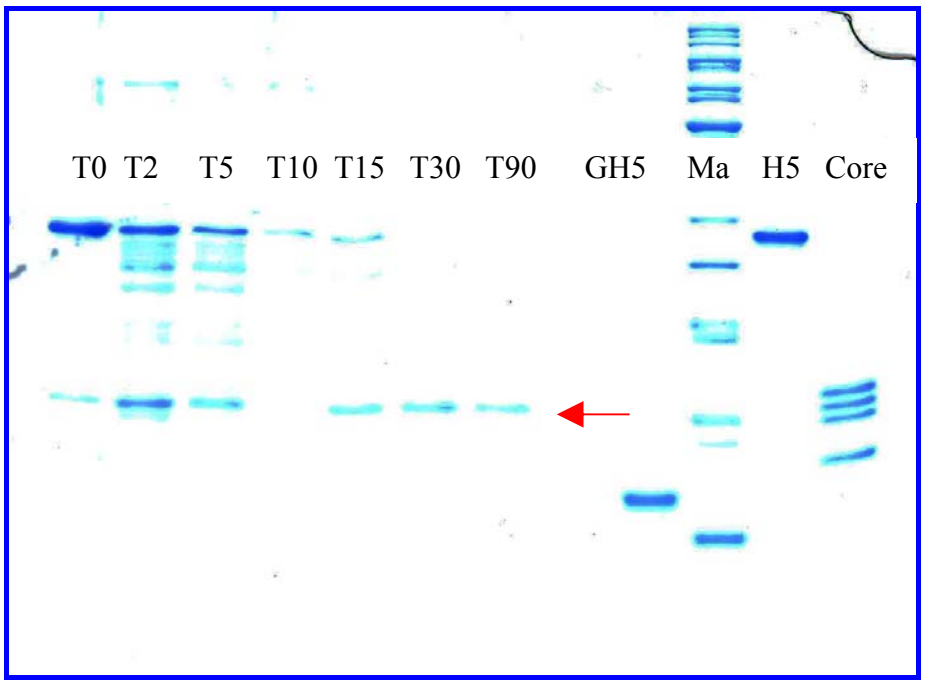

Figure 4.49 H5 time course digestion in the presence of stripped chromatin. To determine the appropriate concentration of chromatin to H5 the same ratio of H5/DNA was used as above, 1molecule/200bp DNA. The chromatin had been previously stripped of linker histones as stated above and sized. The digestion was stopped with STI for this experiment. STI bands indicated with the red arrow

Results references:

Allan, J., Hartman, PG, Crane-Robinson, C, Aviles, FX, Nature **288** (5792) 675-679 (1980) *The structure of histone-H1 and its location in chromatin*

Allan, J. Mitchell, T., Harbone, N. Bohm, L., Crane-Robinson, C. Journal of Molecular Biology **187** (4) 591-601 (1986) *Roles of H-1 domains in determining higher-order chromatin structure and H-1 location*

Bharath, MMS, Ramesh, S., Chandra, NR, Rao, MRS, Biochemistry **41** 7617-7627 (2002) *Identification of a 34 amino acid stretch within the C-terminus of histone H1 as a DNA-condensing domain by site-directed mutagenesis*

Beynon, R., Bond, J. Oxford University Press *Proteolytic Enzymes* 2nd Edition (2001)

Butler, P.J.G., Thomas, J.O., Journal of Molecular Biology, **144,** 89-93 (1980) *Size-dependence of a stable higher-order structure of chromatin*

Cohen, S., Chait, B., Analytical Biochemistry **247**, 257-267 (1997) *Mass spectrometry of whole proteins eluted from sodium dodecyl sulphate-polyacrylamide gel electrophoresis gels*

Fahrney, D., Gold, A., Journal of the American Chemical Society **85** (7) 997-1000 (1963) *Sulfonyl fluorides as inhibitors of esterases I. Rates of reaction with acetylcholinesterase, α-Chymotrypsin, and Trypsin*

Gold, A. Biochemistry **4** (5) (1965) *Sulfonyl fluorides as inhibitors of esterases 3. Identification of serine as site of sulfonylation in phenylmethanesulfonyl alpha-chymotrypsin*

Goytisolo, F.A., Gerchman, SE., Yu, X., Rees,C., Graziano, V., Ramakrishnan, V., Thomas, J.O., EMBO Journal **15** (13) 3421-3429 (1996) *Identification of two DNA-binding sites on the globular domain of histone H5*

Hart, S., Michael Barber Center for Mass Spectrometry, University of Manchester Personal communication 2005 data unpublished

James, G. Analytical Biochemistry, **86** (2) 574-579 (1978) *Inactivation of the protease inhibitor phenylmethylsulfonyl flouoride in buffers*

Jørgensen, C.S., Jagd, M., Sørensen, B.K., McGuire, J., Barkholt, V., Højrup, P., Houen, G. Analytical Biochemistry **330** 87-97 (2004) *Efficacy and compatibility with mass spectrometry of methods for elution of proteins from sodium dodecyl sulphate-polyacrylamide gels with polyvinyldifluoride membranes*

Meyer, T., Lamberts, B., Biochimica et Biophysica Acta **107** (1) 144-145 (1965) *Use of* coomassie *brilliant blue R250 for electrophoresis of microgram quantities of parotid saliva proteins on acrylamine-gel strips*

McIntyre, S., Michael Barber Center for Mass Spectrometry, University of Manchester Personal communication 2005 data unpublished

McIntyre, S., Sugden, D., Riba, I., Gaskell, S.J., ABRF poster presentation 2005: *From gel to MS: factors influencing peptide recovery from gels.*

Nomura, E., Katsuta, K., Ueda, T., Toriyama, M., Mori, T., Inagaki, N. Journal of Mass Spectrometry **39** 202-207 (2004) *Acid-labile surfactant improves in-sodium dodecyl sulphate polyacrylamide gel protein digestion for matrix-assisted laser desorption/ionization mass spectrometric peptide mapping*

Neue, U., Wiley-VCH, *HPLC Columns Theory, Technology, and Practice* (1997)

Orlando, V., Strutt, H., Paro, R., Methods: A Companion to Methods in Enzymology **11** 205-214 (1997) *Analysis of chromatin structure by in Vivo formaldehyde cross-linking*

Poncz, L., Dearborn, D.G., Journal of Biological Chemistry **258** (3) 1844-1850 (1983) *The resistance to tryptic hydrolysis of peptide bonds adjacent to N^{ϵ},N-dimethyllysyl residues*

Powers, J., Asgian, J., Ekici, O., James, K., Chemical Reviews **102** (12) 4639-4750 (2002) *Irreversible inhibitors of serine, cysteine and threonine proteases*

Prouty, W., Goldberg, A. Journal of Biological Chemistry, **247** (10) 3341-3352 (1972) *Effects of protease inhibitors on protein breakdown in Escherichia-coli*

Puchades, M., Westman, A., Blennow, K., Davidsson, P. Rapid Communications in Mass Spectrometry **13** 344-349 (1999) *Removal of sodium dodecyl sulphate from protein samples prior to matrix-assisted laster desorption/ionization mass spectrometry*

Roque, A., Orrego, M., Ponte, I., Suau, P. Nucleic Acids Research **32** (20) 6111-6119 (2004) *The preferential binding of histone H1 to DNA scaffold-associated regions is determined by its C-terminal domain*

Thoma, F., Koller, T.H., Klug, A. Journal of Cell Biology, **83**, 403-427 (1979) *Involvement of Histone H1 in the Organization of the Nucleosome and of the Salt-dependent Superstructures of Chromatin*

Zhou, S., Bailey, M., Dunn, M., Preedy, V., Emery, P. Proteomics **5**, 2739-2747 (2005) *A quantitative investigation into the losses of proteins at different stages of a two-dimensional gel electrophoresis procedure*

5. Summary/Discussion

The data presented here shows the need for further work on the calibration of the SDS/PAGE gels for a more accurate identification of the mass of the histones and their trypsin digest products.

The direct analysis of the digest products by MALDI was unsuccessful in generating the information needed to identify the sites of sensitivity although was successful in determining the masses of digest products. The direct analysis allows for mass identification but does not allow for taking that information back for structural determination.

Figure 4.47 and figure 4.49 are the gels for the digestion products of H5 in first the presence of DNA (figure 4.47) and second in the presence of chromatin (figure 4.49). These alas were the only two gels demonstrating these experiments and as can be seen in figure 4.49 data quality is poor and would be insufficient to draw distinct conclusions from. This was in part due to the work required to determine successful digestion conditions for the H5 as well as trying to calibrate the mass identification of the H5 and its digest products on polyacrylamide gels. There was a limited amount of time available for the work and as a result there are only the two gels (figures 4.47 and 4.49) to use for any determination of the digest products of H5 in the varying conditions in order to determine the sites on the H5 tails which are interacting with the chromatin.

The multiple routes of peptide identification used; elution from gel combined with MALDI analysis, in-gel digestion of peptides combined with MALDI analysis, chromatographic separation of the digest products followed by MALDI analysis and direct MALDI analysis of the digest products, allowed for the determination of some of the peptides generated by digesting the H5 with trypsin but there is little overlap and an unclear determination of the specific masses generated from the digestion. This makes comparison of the masses observed in the differing techniques impossible to make.

The work makes few conclusions but it was determined that the most successful means for halting the trypsin efficiently is with the use of soya bean trypsin inhibitor. The time course digestion is most clearly and consistently shown, see figure 4.9 using the STI. Work to determine an alternate was inclusive for consistency (see

figures 4.10, 4.11, 4.12 and 4.13 in generating a digest pattern consistent with the pattern observed using STI.

6. Future Work

It seems clear that, to build on the achievements made in this research and to obtain better results regarding information of the 30nm fiber chromatin structure, two important changes will have to be made to the methodology employed in this project:

1. staining of the gels that are used for identification will have to be avoided, and
2. the products of the digestion will have to be extracted from the gels immediately after they are formed.

While the extraction procedure will always present difficulties, these two key factors should improve the quality of the results obtained in the future. Other improvements that could be made to improve the elution of the peptides include making use of the methodology reported recently by (Nomura et al. 2004) and by (Jørgensen et al. 2004). The work by (Nomura et al. 2004) indicated that treatment of SDS/PAGE gels used for digestion with an acid-labile surfactant (3-[(2-methyl-2-undecyl-1,3-dioxolan-4-yl)methoxy]-1-propanesulfonate, ALS), improves the elution of peptides from gels and would therefore increase the number of peptides observed from the extraction. Jørgensen used electroblotting methods and various membranes to improve the elution efficiency of peptides from gels. These various procedures may help to improve the elution of the peptides from the gels, but there will continue to be a problem caused by the necessity to identify all of the C-terminal tail of H5 in order to identify the cleavage sites and by inference the binding with the linker DNA in the 30nm fiber.

12 nucleosomes are required for two turns of the solenoid to give the minimum for the 30nm fiber. But it is preferable to study larger fibers. This would require five turns of the solenoid or 30 nucleosomes equating to ~ 6,000 bp DNA giving a mass of roughly 5 megadalton (MDa). At this mass it would currently be a challenge to examine an intact fiber by mass spectrometry. This would also require a change in the approach to the project; looking at the whole of the fiber as opposed to looking at the tails of the linker histone.

The development of MALDI and electrospray ionization techniques opened the door to the analysis of large complex components by mass spectrometry; improvements on these techniques have continually expanded the upper mass range of the instrumentation. (Brown and Lennon 1995) showed the use of pulsed ion extraction in order to improve the resolution in MALDI time-of-flight analysis, and (Bahr et al. 1997) demonstrated that the use of delayed extraction in MALDI mass spectrometry improved the observation of proteins above 25,000 Da. We can now see examples of complexes in the megadalton range observed both by MALDI (Wenzel et al. 2005) and by electrospray (Videler et al. 2005). While it may not be possible at present to use mass spectrometry as the tool to unwind the structure of chromatin, it seems probable that it will soon be possible to do so.

Future Work References:

Bahr U, StahlZeng J, Gleitsmann E, Karas M, Journal of Mass Spectrometry, **32**, (10) 1111-1116 (1997) *Delayed extraction time-of-flight MALDI mass spectrometry of proteins above 25,000 Da*

Brown, RS, Lennon, JJ, Analytical Chemistry, **67**, (13) 1998-2003 (1995) *Mass resolution improvement by incorporation of pulsed ion extraction in a matrix-assisted laser-desorption ioinization linear time-of-flight mass-spectrometer*

Jorgensen CS, Jagd M, Sorensen BK, McGuire J, Barkholt V, Hojrup P, Houen G, Analytical Biochemistry, **330**, (1) 87-89 (2004) *Efficacy and compatibility with mass spectrometry of methods for elution of proteins from sodium dodecyl sulfate-polyacrylamide gels and polyvinyldifluoride membranes*

Nomura E, Katsuta K, Ueda T, Toriyama M, Mori T, Inagaki N, Journal of Mass Spectrometry, **39**, (2) 202-207 (2004) *Acid-labile surfactant improves in-sodium dodecyl sulfate polyacrylamide gel protein digestion for matrix-assisted laser desorption/ionization mass spectrometric peptide mapping*

Videler, H, Ilag, LL, McKay, ARC, Hanson, CL, Robinson, CV, FEBS Letters, **579**, (4) 943-947 (2005) *Mass spectrometry of intact ribosomes*

Wenzel, RJ, Matter, U, Schultheis, L, Zenobi, R, Analytical Chemistry, **77**, (14) (2005) *Analysis of megadalton ions using cryodetection MALDI time-of-flight mass spectrometry*