The association of the secondary DNA-binding site of linker histone H5 in a nucleosome

Gerrit Koorsen

SUBMITTED IN FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN THE DEPARTMENT OF MOLECULAR AND CELL BIOLOGY
FACULTY OF SCIENCE
UNIVERSITY OF CAPE TOWN

2001

CONTENTS

Abstract

In order to understand the role of linker histones in the formation of the 30-nm chromatin fibre as well as their role in transcriptional repression, it is essential to know their location on the nucleosome. In this study, we have modelled the location of the globular domain of chicken linker histone H5 (GH5) on the nucleosome. The primary DNA-binding site of GH5 was modelled by homology to the co-crystal structure of the *E. coli* CAP-DNA complex. In this model, GH5 recognises DNA mainly through contacts of conserved residues located in helix III with the DNA-backbone. The secondary DNA-binding site of GH5 was modelled by examining the possible geometry of linker DNA at the entry/exit points of the nucleosome. We found that the secondary DNA-binding site of GH5 is in contact with nucleosomal DNA in the vicinity of the nucleosomal dyad. In addition, we showed that our model of both the primary and secondary DNA-binding site is in agreement with most experimental observations. In order to allow the future testing of the model by site-directed metal-affinity cleavage, we have designed 3 GH5 mutants, R37C, S41C and A96C. Substituted residues are located in the putative secondary DNA-binding site of GH5. Coding sequences for these mutants were generated by template mismatch PCR mutagenesis, and subcloned into pET20b(+) expression vectors. Both R37C and S41C were stably expressed in *E. coli* BL21(DE3)pLysS cells. However, A96C could not be expressed in these cells. Two native GH5 domains were also sub-cloned into pET20(b)+ expression vectors. We found that the domain spanning residues 20 – 109 was not expressed in *E. coli* BL21(DE3)pLysS cells. In contrast, the domain spanning residues 23 – 106, preceded by an N-terminal MTE-motif, was expressed at high levels. The GC content immediately downstream of the ATG start codon of the expression constructs may offer an explanation for these observations. In addition, we have developed a novel approach to the synthesis of an EDTA-derivative, cyst-EDTA-NPS, to modify GH5 peptides with suitably substituted cysteine residues in order to map the location of the secondary DNA-binding site on the nucleosome. Through this method, pure cyst-EDTA-NPS is produced without the need for chromatography. The mapping of the secondary DNA-binding site of GH5 on the nucleosome by site-directed metal-affinity cleavage is currently underway, although not reported for the purpose of this dissertation.

Abbreviations

°C	Degrees centigrade
Å	angstrom
A96C	GH5 mutant with alanine 96 substituted with cysteine
AAA/TTT	Three consecutive adenine/thymine DNA base steps
ADP	Adenosinediphosphate
AGGA	DNA sequence containing adenine-guanine-guanine-adenine
Ala	L-Alanine
Arg	L-Arginine
Asn	L-Asparagine
ATP	Adenosinetriphosphate
AT-rich	DNA rich in adenine and thymine
bp	base pair
br s	broad singlet
CAP	Catabolite Activator Protein
Cdc2	Cell-cycle dependent cyclin 2
$CDCl_3$	Deuterated chloroform
αCHC	α-cyano-4-hydroxycinnamic acid
cm	centimetre
CpG	5'-cytosyl-guanidine-dinucleotide
C-terminus	Carboxyterminus
Cys	L-Cysteine
cyst-EDTA-NPS	*S*-(nitrophenylsulphenyl)-cysteaminyl-EDTA
CZ	*ipso* carbon atom
δ	chemical shift
D_2O	Deuterated water
Da	Dalton
dd	Doublet of doublets
ddd	Doublet of doublet of doublets
DALI	3D Alignment

DCC	Dicyclohexylcarbodiimide
DHB	2,5-dihydrobenzoic acid
DMF	*N,N*-dimethylformamide
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
DNAse I	Deoxyribonuclease I
rDNA	DNA coding for ribosomal RNA
E. coli	*Escherichia coli*
EDIA	*N*-ethyldiisopropylamine
EDTA	Ethylene-diamine-tetra-acetic acid
EPD	*S*-(2-pyridylthio)cysteaminyl-EDTA
F^{254}	Fluorescent under 254 nm UV-radiation
Fe^{2+}	Ferrous ion
Fe^{3+}	Ferric ion
g	gram or gravity constant
G1 phase	Gap phase 1
GC box	Short regulatory DNA-sequence rich in guanidine and cytosine
GC-rich	DNA rich in guanine and cytosine
GGA-box	DNA sequence consisting of guanine-guanine-adenine
GH1	Globular domain of linker histone H1
GH5	Globular domain of linker histone H5
Gln	L-Glutamine
Gly	Glycine
h	hour
^1H	Proton
H1-CS	Cleavage stage H1
H1-GFP	H1 fused to Green Fluorescent Protein
H1t	Testes-specific H1
H_2O	water
H4-box	Regulatory element originally identified in *H4* gene promoter
H4TF2	H4-box binding transcription factor 2

H-bond	Hydrogen bond
HBT	1-hydroxybenzotriazole
HCl	Hydrochloric acid
His	L-Histidine
His-tag	Carboxy-terminal tract of 6 consecutive histidine residues
HNF-3	Hepatocyte nuclear factor 3γ
HPLC	High Performance Liquid Chromatography
HTH	Helix-turn-helix
IMAC	Immobilised Metal Affinity Chromatography
IPTG	isopropyl-β-D-thiogalactoside
Ile	L-Isoleucine
kb	kilobases
kcal	kilocalories
K_2CO_3	Potassium carbonate
kDa	kilodalton
K[S/T]PXK	Lysine-[serine/threonine]-proline-[arginine/lysine]-lysine
K[S/T]PK	Lysine-[serine/threonine]-proline-lysine
ℓ	litre
LB	Luria Broth
Lys	L-Lysine
m	multiplet
MALDI-TOF	Matrix Assisted Laser Desorption Ionisation/ Time Of Flight
MeOH	Methanol
mg	milligram
Mg^{2+}	Magnesium ion
$MgSO_4$	Magnesium sulphate
MHz	Megahertz
min	minute
μg	microgram
μℓ	microlitre

μm	micrometer
μM	micromolar
μmol	micromole
mℓ	millilitre
mm	millimetre
mM	millimolar
mmol	millimol
MNase	Micrococcal nuclease
mol	mole
MS	Mass Spectrometry
MTE-motif	Methionine–threonine-glutamate motif
Na^+	Sodium ion
NaCl	Sodium chloride
$NaHCO_3$	Sodium bicarbonate
NGGR	Deoxyribo(nucleotide-guanidine-guanidine-purine)
NH_4HCO_3	Ammonium bicarbonate
$NiSO_4$	Nickel sulphate
ng	nanogram
nm	nanometer
NMR	Nuclear magnetic resonance
NPS	2-nitrophenylsulphenyl
NPS·Cl	2-nitrophenylsulphenylchloride
N-terminus	Aminoterminus
OD_{600}	Optical Density at 600 nm
•OH	Hydroxyl radical
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PLE	Pig Liver Esterase
poly(dA-dT)	DNA sequence containing consecutive adenine-thymine base steps
ppm	parts per million

R37C	GH5 mutant with arginine 37 substituted with cysteine
hRFX-DBD	DNA-Binding Domain of human Recognition Factor of X-box
RMS	Root Mean Square
RNA	Ribonucleic acid
mRNA	Messenger ribonucleic acid
rRNA	Ribosomal ribonucleic acid
tRNA	Transfer ribonucleic acid
$tRNA^{Lys}$	Aminoacyl transfer ribonucleic acid for Lysine
rpm	Revolutions per minute
s	second (unit of time) or singlet (NMR description)
S41C	GH5 mutant with serine 41 substituted with cysteine
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
S-phase	Synthesis phase
SPKK	Serine-proline-lysine-lysine
SV40	Simian Virus 40
td	Triplet of doublets
TFA	Trifluoroacetic acid
TLC	Thin Layer Chromatography
Thr	L-Threonine
THR	Thyroid Hormone Response element
Tris	Tris(hydroxymethyl)aminomethane
Tyr	L-Tyrosine
U	unit
UNE	Upstream negative element
UPE	Upstream proximal element
UV	Ultraviolet
V	volt
Val	L-Valine
v/v	volume per volume
w/v	weight per volume

X-box	DNA-recognition sequence of hRFX-DBD
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

List of Tables

List of Figures

Acknowledgements

First and foremost, I would like to thank my supervisor, Hugh Patterton for his support. I would also like to thank Eugene Sickle, for his advice and help with the chemical syntheses as well as Kerstin Bysticki, for helpful personal communications. I am grateful towards all the technicians in the department of Molecular and Cell Biology at UCT for practical help, in particular Faezah Davids, Theresa Anderson and Anke Fiedler. With respect to the modelling part of this project, I wish to thank Gerrit Vriend and Celia Gelder at CMBI, Nijmegen, as well as Derek Litthauer at UOFS, Bloemfontein, for stimulating discussions and valuable advice. In addition, I wish to thank Venki Ramakrishnan for his gift of pET9-H5 that served as PCR template. Furthermore, I wish to acknowledge the Wellcome Trust and the National Research Foundation for funding. Lastly, I would like to thank my parents for support, as well as my colleagues, especially Lieschen, and my friends.

CHAPTER 1: INTRODUCTION

1.1. Background

The nucleosome is the fundamental structural unit of chromatin, the packaged form of DNA in a eukaryotic cell nucleus. It is composed of two copies of each of the core histones H2A, H2B, H3 and H4. These eight proteins form the histone octamer onto which approximately 168 bp of genomic DNA is coiled. A fifth histone, the linker histone, binds to linker DNA, which is the length of DNA connecting adjacent nucleosomes. The globular domain of the linker histone facilitates the binding of the protein to the nucleosome, and directs the basic tail domains to the linker DNA (Simpson, 1978). The association of a linker histone with a nucleosome gives rise to a chromatosome, the packing unit of heterochromatin (Wolffe, 1998).

Linker histones partially neutralise the negative charges on the linker DNA, allowing adjacent nucleosomes to come into close proximity in space. Functionally, linker histones are responsible for stabilizing the 30-nm chromatin fibre (Thoma *et al.*, 1979), a highly dense conformation of chromatin.

1.2. Evolutionary origin of linker histones

Contrary to the archaebacterial origin of the core histones (Van Holde, 1988), linker histones evolved independently from eubacterial precursors (Kasinsky *et al.*, 2001). The

first lysine-rich DNA-condensing proteins arose in *Chlamidia* with striking compositional similarity to the C-terminal tails of metazoan linker histones. However, the globular domain evolved only later in the protista, directing neutralisation to linker regions by anchoring the protein to the nucleosomal core particle. This domain co-evolved with the more variable N-terminal tail (Kasinsky *et al.*, 2001).

1.3. Linker histone variants

Linker histones function primarily by partially neutralising negative charges on the sugar-phosphate backbone of linker DNA and it is conceivable that this process might occur via the action of several possible structures. Therefore, the variability of linker histones comes as no surprise. Since we now know that linker histones have a bacterial origin (Kasinsky *et al.*, 2001), it is possible that the variety of linker histone subtypes arose in response to the increasing functional complexity of evolving chromatin domains. Originally, four classes were proposed based on both the biochemical and biophysical behaviour of linker histones (Van Holde, 1989). These included H1, H1°/H5, H1t and H1-CS. Alternatively (Khochbin, 2001), these groups could be designated replication-dependent linker histones (H1), differentiation dependent linker histones (H1° and H5), cleavage stage linker histones (H1-CS) and testes-specific linker histones (H1t), respectively. It should be noted that a single linker histone could functionally bridge more than one of these classes. This is observed for the replication-dependent H1c and H1e in mouse, which responds to cell-cycle signals (Wang *et al.*, 1997). The strong link between the progression of embryogenesis in invertebrates and differential deposition of

linker histone subtypes (Trieschmann *et al.,* 1997; Flenniken and Newrock, 1997) suggested that differences in primary structure of subtypes encode functional differences. The switching of linker histone subtypes during invertebrate development seems to be less dramatic in mammals (Franke *et al.,* 1998), where somatic subtypes are also present in the developing embryo. However, the synthesis of the mouse testes-specific linker histone, H1t, strongly coincides with the pachytene meiotic prophase (Bucci *et al.,* 1982).

The hypothesis that the need for H1 variants lies in the function of different primary structures was recently challenged by the lack of observable phenotypes in cases where linker histones were knocked out in various organisms. The most startling observations were made in mouse, where several groups reported normal spermatogenesis in mice lacking the H1t gene (Lin *et al.,* 2000; Drabent *et al.,* 2000). In addition, these mice exhibited normal H1:nucleosome ratios in spermatogonia, suggesting that other H1 subtypes could assume the function of H1t. In the light of these and similar observations, Khochbin (2001) proposed that the timing of the expression of linker histones might be functionally more important than differences in primary structure and that there is a certain redundancy in the primary structure of linker histones. However, the observation that as H5 replaces H1 in developing avian erythrocytes, the linker length changes (Schlegel *et al.,* 1980), suggests that the effect of different H1 subtypes on chromatin compaction is not necessarily trivial. Also, as the epigenetic histone code (Allis *et al.,* 2001) is elucidated, the post-translational modification potential of linker histone variants may be expected to carry as much functional weight as the timing of their deposition.

1.4. Linker histone H5

H5 forms part of the class of differentiation-dependent linker histones, but it has been suggested that H5 should be considered an erythrocyte-specific sub-type of H1° (Khochbin, 2001). Indeed, H5 is only found in the erythrocytes of birds, fish, amphibia (with the exception of *Rana catesbeiana*; Shimada *et al.,* 1981) and reptiles, where it accumulates during erythrocyte maturation. In addition, expression of the *H5* gene is not limited to S-phase, and coincides with inhibition of replication and the arrest of the cell-cycle in G1 phase (Grunstein, 1990). Recently, the inhibition of histone acetylation by H5 was reported (Herrera *et al.,* 2000).

The *H5* gene promoter contains three major *cis*-acting regulatory elements: an upstream negative element (UNE; -115 to -95), GC box (-83 to -74) and an upstream proximal element (UPE; -54 to -38) or H4-box (Rousseau *et al.,* 1989). The UPE appears to respond to the differentiation state of the cell through interaction with the H4-box specific factor, H4TF2 (Zlatanova and Doenecke, 1994). Additional activation is caused by the action of three enhancers, located both upstream and downstream of the coding sequence (Rousseau *et al.,* 1993). These enhancers appear to recruit erythrocyte-specific transcription factors.

1.5. Post-translational modification of linker histones

Although acetylation of free amino groups of N-terminal serine residues of linker histones (Ruiz-Carrillo *et al.,* 1975) as well as ADP-ribosylation of linker histones (Tanigawa *et al.,* 1983) have been reported, the best studied example of post-translational modification of linker histones and core histones alike is phosphorylation. Linker histones are phosphorylated by the cell cycle dependent Cdc2 kinase at specific K[S/T]PXK or K[S/T]PK motifs residing in the unstructured domains. Cell-cycle independent phosphorylation of linker histones by protein kinase A was also reported by Sweet and Allis (1993).

Hill and colleagues made the intriguing observations that phosphorylation of N-terminal sites on H1 inhibited association of these linker histones with DNA *in vitro*, while phosphorylation of C-terminal sites seemed to have little effect on DNA-binding (Hill *et al.,* 1991). Surprisingly, when both N- and C-terminal sites were phosphorylated, DNA-binding was unaffected, yet these proteins had a lower affinity for chromatin. This study pointed towards a domain-dependent phosphorylation of linker histones, a process that seems to be further regulated by the presence of phosphorylation sites of two independent protein kinases.

Pioneering studies on the cell-cycle dependent phosphorylation of linker histones in *Physarum* (Bradbury, 1973) and Chinese Hamster Ovary (CHO) cells (D'Anna *et al.,* 1985), led to the popular hypothesis that H1 hyperphosphorylation served as a trigger for

mitotic chromatin condensation. This conclusion seems paradoxical, since hyperphosphorylation would decrease the interaction of H1 with linker DNA. The conclusion stems from the observation that levels of phosphorylated H1 increase in dividing cells, but decrease in quiescent cells (Bradbury *et al.,* 1973). This increase commences in G1 and proceeds through S phase to mitosis, where phosphorylation levels peak on or just before mitosis, followed by a sharp decline (D'Anna *et al.,* 1985). In addition, it was observed that extracts from mitotic cells containing H1 kinase activity, stimulates premature chromatin condensation when added to non-mitotic cells.

In the heterokaryon, *Tetrahymena*, macronuclear H1 is rapidly and completely phosphorylated during conjugation. The fact that H1 phosphorylation peaks under conditions of heat shock and conjugation (Roth *et al.,* 1988), which are not linked to cellular or macronuclear division, suggested that H1-phosphorylation is not causally linked to mitosis.

Another uncoupled system is that of avian erythropoiesis. Newly synthesized H5 is highly phosphorylated, but is dephosphorylated during the final stages of erythropoiesis, where chromatin becomes highly condensed (Sung *et al.*, 1977). Therefore, in this system dephosphorylation, rather than phosphorylation seems to be linked to chromatin compaction.

Sea urchin spermatogenesis and fertilisation presents a way of studying the reversible condensation of chromatin in response to phosphorylation signals. In agreement with

avian erythropoiesis, Green and Poccia (1985) observed high levels of phosphorylated H1 and H2B in spermatids, while these histones were in the unphosphorylated state in the condensed chromatin of mature sperm.

Bradbury *et al.* (1973) proposed that although H1 phosphorylation would weaken interaction with linker DNA, phosphorylated H1 could be in a conformation that would allow H1-H1 interactions driving chromatin compaction. Contrary to this model, Roth and Allis (1992) proposed that H1 phosphorylation does not play a direct physical role in the condensation process, but that H1 phosphorylation allows transient decondensation of chromatin domains, making specific DNA sequences accessible to trans-acting factors (e.g. topoisomerase II) involved in chromosomal compaction.

The role of ADP-ribosylation of linker histones is also unclear, although this modification is associated with DNA repair (Van Holde *et al.,* 1989).

1.6. Linker histone structure

Most linker histones exhibit a tripartite structure, comprising a short N-terminal "nose", a folded, globular domain, and a highly basic C-terminal tail. These domains are thought to act as independent modules, and to interact differently with DNA (Wolffe, 1998). There are at least two linker histones that differ from the canonical tripartite structure. Macronuclear H1 of the heterokaryon, *Tetrahymena*, lacks the central globular domain

(Hayashi *et al.,* 1987), while *Saccharomyces cerevisiae* H1 contains two putative globular domains (Patterton *et al.,* 1997).

The X-ray crystal structure of the globular domain of H5 (GH5) was solved by Ramakrishnan and colleagues. The domain has the topology of a three-helix-bundle with a β-hairpin or "wing" at the C-terminal end. The protein is not a true helix-turn-helix protein (HTH), since the characteristic four-residue turn connecting HTH helices II and III is replaced by a seven-residue loop. GH5 crystallized as a dimer, and the conformation of the "wing" of individual monomers was quite distinct due to crystal packing effects (Ramakrishnan *et al.,* 1993). This observation points to the flexibility of this motif in solution, which was also inferred from NMR studies on the globular domain of histone H1 (GH1; Cerf *et al.,* 1994). In fact, the only difference between the structure of GH5 and GH1 is the conformation of the loop connecting helices II and III. His 62 is present in GH5, while this residue is absent in GH1, where the whole loop has a negatively charged character. This difference may result in the different affinities of these domains for DNA.

Tarkka *et al.* (1997) suggested that folding of the globular domain may not be as reported (Ramakrishnan *et al.,* 1993). Using circular dichroism to measure α-helicity, the authors demonstrated that ATP promoted H1 folding at sub-physiological concentrations. In addition, ATP-dependent conformational changes were observed that resulted in distortion of poly(dA-dT) tracts complexed with H1. This observation is of particular

importance, since poly(dA-dT) sequences occur frequently at the ends of nucleosomes or in linker DNA (Nelson *et al.,* 1987) as well as in regulatory areas of genes (Lewin, 1996).

Linker histone tail domains are essentially unstructured in solution (Aviles *et al.,* 1978). The N-terminal tail is required for correct positioning on the nucleosome, while the C-terminal tail is required for formation of the 30-nm fibre (Allan *et al.,* 1986). It is this tail domain that interacts strongly with linker DNA. It has been shown that the C-terminal tail assumes a segmented α-helical conformation under conditions mimicking phosphate groups on the DNA backbone (Clark *et al.,* 1988). Helical segments, interrupted by proline residues, were proposed to follow the grooves of linker DNA. Indeed, in an electronmicroscopic study (Hamische *et al.,* 1996), it was inferred that the C-terminal tail was not in a fully extended conformation in association with DNA. Suzuki (1989) proposed that SPKK phosphorylation motifs in tail domains constitute a stable β-hairpin motif known to preferentially bind to the major groove of DNA.

1.7. Interaction of linker histones with DNA

1.7.1. Interaction with linear DNA

Since cross-linked H1 homopolymers were isolated from chromatin (Thomas and Khabaza, 1980) and the same complexes were observed for H1 molecules cross-linked to linear DNA (Clark and Thomas, 1986), the interaction of linker histones with linear, double-stranded DNA might offer insight into the binding of H1 to DNA on the interior of the 30-nm fibre. This premise is supported by the observation that the ionic strength at

which nucleosomal arrays compact, correlates with the ionic strength required for the formation of aggregates of H1-DNA complexes (Zlatanova and Van Holde, 1996).

It was shown that co-operative binding of H1 to linear DNA is salt-dependent and occurs above a critical salt concentration of 20 – 50 mM NaCl (Clark and Thomas, 1986). The globular domain of H5 also appears to interact cooperatively with linear, double-stranded DNA, although the cooperativity seems independent of ionic strength (Draves *et al.,* 1992). In the same study, GH5 molecules were reported to be in close proximity in such GH5-DNA complexes, as judged by cross-linking.

Electronmicroscopy revealed several structures for H1-DNA complexes, depending on the ionic strength and the H1:DNA ratios in these complexes. At low salt concentrations, thin filaments are observed, consisting of two DNA duplexes bridged by linker histones, while rodlike and cablelike structures are observed at higher ionic strengths. Doughnuts and circles have also been observed (Zlatanova and Van Holde, 1996). From this observation, it seems evident that linker histones have the capacity for strong DNA-bending, but the physiological significance of the above structures remains unclear.

1.7.2. Interaction with DNA exhibiting pronounced secondary structures

The resemblance of the nucleosomal entry/exit DNA to synthetic DNA cross-overs or four-way junctions, motivated studies on the binding of linker histones to such DNA structures *in vitro*.

It was shown that linker histones prefer binding to superhelical rather than linear DNA (Van Holde, 1989). Krylov *et al.* (1993) examined the binding of H1 and H5 to supercoiled pBR322 plasmid containing 0 – 5 superhelical turns. It was observed that both H1 and H5 exhibit binding preference to topoisomers with the highest number of superhelical turns, suggesting a preference for linker histones to bind to DNA cross-overs.

The binding of chicken erythrocyte linker histones H1 and H5 to synthetic four-way junction DNA was studied by Varga-Weisz *et al.* (1994). The authors demonstrated a higher affinity of these linker histones, as well as their globular domains, to four-way junction DNA compared to linear, double-stranded DNA. Interestingly, H5 compacted four-way junctions to a larger extent than H1. The inhibition of binding by cations (Mg^{2+}, Na^{+}) and spermidine was correlated to a change in four-way junction structure and led to the conclusion that linker histones preferentially bind to square planar DNA conformations. This is of particular importance, since it is conceivable that the nucleosomal entry/exit DNA exists in this conformation.

1.7.3. Sequence-specificity and preference for methylated DNA

It is widely accepted that linker histones show a general preference for AT-rich DNA-sequences (Zlatanova and Van Holde, 1996).

The protection of SV40 DNA restriction sites complexed with H1 was compared by Hendrickson and Cole (1994). In addition, this group generated several SV40 DNA fragments and measured the ability of each fragment to compete for limiting amounts of H1. Preferential binding of H1 to AT-rich sequences was observed. Although the AT-preference was observed even for prokaryotic DNA, the significance of this study is questionable, given that the DNA was not in a chromatin context.

Wellman *et al.* (1994) studied the binding of the H1e linker histone variant to homogeneous DNA-fragments by monitoring the effects of H1e on melting of DNA and by gel-mobility-shift assay. Surprisingly, it was observed that H1e shows a preference for GC-rich sequences and that the preference resides in the C-terminal tail. This report is in agreement with the identification of a number of strong H1-binding sites, H1-hypersites, that are not necessarily AT-rich (Zlatanova and Van Holde, 1996). A clue towards the understanding of hypersites came from Tomaszewski and Jerzmanowski (1997), who showed that AT-rich sequences flanking the oocyte-type *Xenopus borealis* 5S rRNA gene were capable of realigning nucleosomal cores, densely packed on this well-defined *in vitro* chromatin template. This was not observed for GC-rich sequences, flanking the somatic 5S rRNA gene repeat. Since it is known that the differential expression of these two types of 5S rRNA genes are dependent on the linker histone (Wolffe, 1998), the preference of the linker histone for oocyte chromatin serves as an explanation for this phenomenon.

In an attempt to investigate the hypothesis that the repressive effect of DNA methylation on gene activity is due to the selective association of H1 to methylated DNA, Nightingale and Wolffe (1995) demonstrated that methylation of CpG dinucleotides within a GC-rich fragment containing a *Xenopus borealis* 5S rRNA gene does not influence H1 binding to either naked or nucleosomal DNA. However, McArthur and Thomas (1996) identified a clear preference of H1 for CpG-methylated dinucleotides, irrespective of sequence. In addition, since it was observed that H1 variants (H5 and spH1) isolated from transcriptionally repressed nuclei show a stronger preference for methylated DNA than H1 from transcriptionally competent nuclei, it was proposed that this effect might well be relevant to transcriptional regulation.

1.7.4. Sequence preference of the chromatosome

By analysing 280 chromatosomal DNA-sequences, Travers and Muyldermans (1996) identified a rotational signal (NGGR), preferentially located at one, but not both chromatosomal termini. This signal was shown to inhibit rotational positioning of the nucleosome core within the chromatosome, and may impose a structural asymmetry on the chromatosome. Two translational signals, a GGA-box and AAA/TTT-rich region were also identified, separated by almost a full nucleosomal supercoil. It was proposed that these signals could constitute strong positioning signals for the globular domain of the linker histone. Wolffe and Hayes (Hayes *et al.,* 1994; Pruss *et al.,* 1996; Hayes, 1996) as well as Belikov and Karpov (1998), however, proposed that the globular domain of the linker histone is located within the borders of the core particle. If this were the

case, it is possible that H3, or the tails of H2A and H2B, dislocated upon linker histone binding, could bind to these markers (Travers and Muyldermans, 1996).

1.7.5. GH5 residues implicated in DNA-binding

Even before the availability of the crystal structure of the globular domain of GH5, several studies implicated basic residues as important in DNA-binding. Thomas and Wilson (1986) observed that most lysines in the globular domain, especially Lys 85 of GH5, are protected from methylation when bound to DNA in chromatin, suggesting close association of this residue with the nucleosomal DNA. Mirzabekov *et al.* (1990) reported that His 25 and His 62 could be chemically cross-linked to chromatin fragments. Using photo-affinity labelling, His 62 was also implicated in DNA-binding by Goytisolo *et al.* (1996). In addition, Buckle *et al.* (1991) demonstrated that mutants of recombinant GH5 lacking Lys 69 did not confer the same chromatosome protection as the native peptide. Armed with the crystal structure of the globular domain of H5 (Ramakrishnan *et al.*, 1993), the involvement of conserved residues (Arg 73, Arg 74, Lys 69 and Lys 85) in DNA-binding was predicted. The residues mentioned above are shown in Figure 1.1.

1.7.6. Primary DNA-binding site

Clore *et al.* (1987) suggested that GH5 might be quite similar, topologically, to the DNA-binding domain of *E. coli* CAP, which also exists as a three-helix bundle with a hairpin at

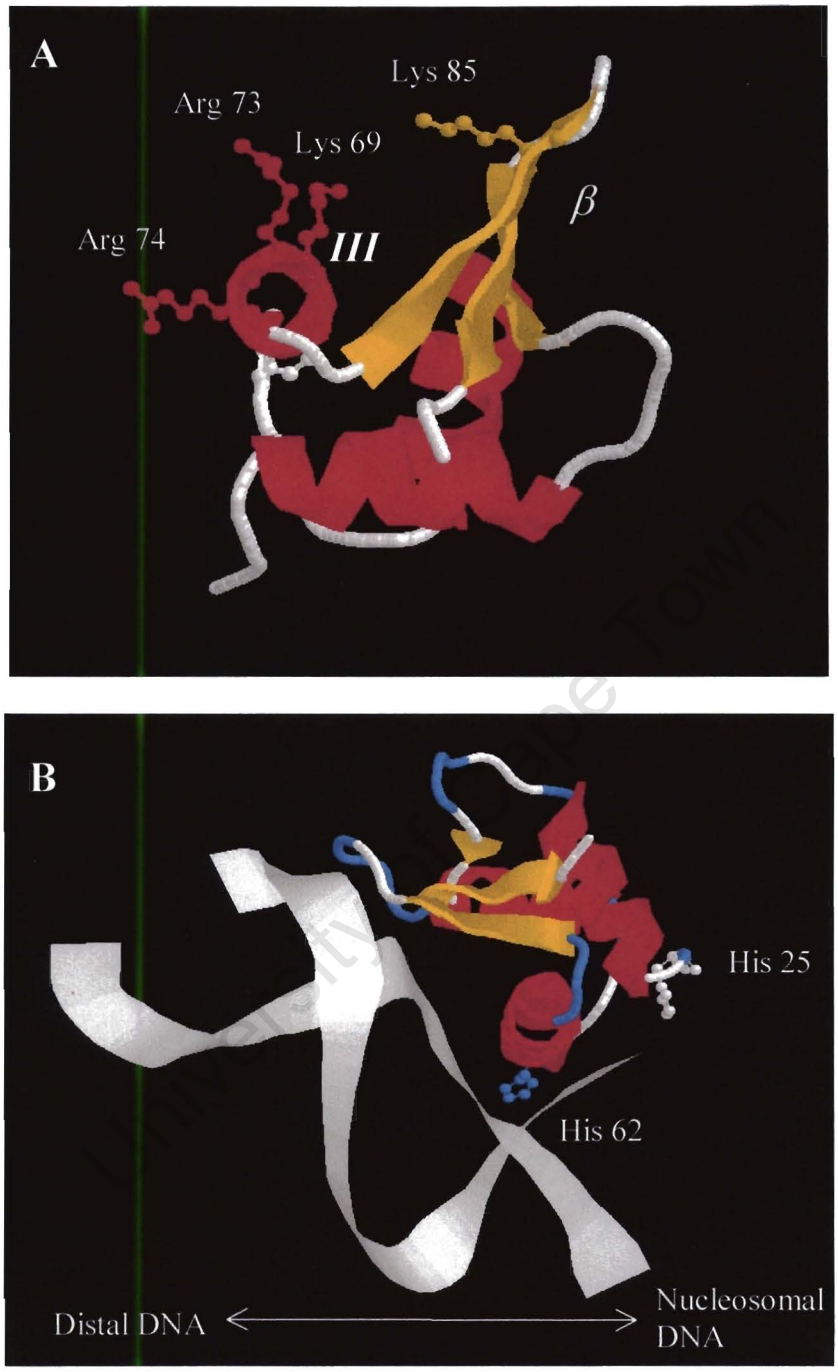


Figure 1.1. GH5 residues implicated in DNA-binding. GH5 is represented as a ribbon, with selected residues in ball-and-stick format. **A.** Lys 85 and Lys 69, implicated in DNA-binding by Thomas and Wilson (1983) and Buckle *et al.* (1992). Arg 73 and 74 are conserved and were implicated by Ramakrishnan *et al.* (1993). Helix III (*III*) and the β-hairpin (*β*) are indicated. **B.** Mirzabekov *et al.* (1993) implicated His 25 and His 62 in DNA-binding and suggested an orientation as shown where His 62 cross-links more distal than His 25, which cross-links closer to nucleosomal DNA.

the C-terminal end. This observation suggests a mode of DNA-binding, where helix III (helix F of CAP) recognizes DNA and binds in the major groove.

Similarly, Clarke *et al.* (1993) also suggested a clear primary DNA-binding site by pointing to the structural similarity of GH5 to HNF-3γ, also a member of the winged helix sub-family of helix-loop-helix (HTH) transcription factors. It was suggested that this site consists of helix III, the C-terminal β-hairpin and loop between helices II and III.

Recently, Gajiwala *et al.* (2000) solved the crystal structure of hRFX-DBD in complex with its cognate X-box DNA binding-site and revealed a novel mode of DNA-binding of a member of the winged helix protein family. According to this model, the major groove of DNA is recognised by the β-wing motif instead of helix III. Therefore, two modes of binding of a winged helix helix-loop-helix protein exist: recognition of the major groove of DNA via the C-terminal β-wing, or via helix III. This observation has interesting consequences for DNA-binding by GH5, also a winged helix protein. The authors proposed that the type of DNA-recognition exhibited by a winged helix fold, depends on the difference in basicity between the β-hairpin (wing) and helix III. This hypothesis was based on the observation that helix III of HNF-3γ as well as other HTH transcription factors is more basic than the β-wing, but that the β-wing of hRFX-DBD is more basic than helix III of the same protein. According to this hypothesis, GH5 should also bind to DNA through recognition of the major groove by the C-terminal β-wing. To support their statement, Gajiwala *et al.* (2000) constructed a model of GH5 complexed to the X-box by homology to hRFX-DBD binding, where the β-wing motif is involved in DNA-

recognition. In such a setting, Lys 85 of GH5 forms two strong H-bonds to the DNA backbone. Moreover, Lys 69 and His 62 also exhibit strong interaction with the X-box DNA backbone. Since this hypothesis is based on a limited number of crystal structures, a stronger argument is required to imply β-wing recognition in DNA-binding by GH5.

1.7.7. Nucleosomal binding

1.7.7.1. The chromatosome

Digestion of chromatin in the absence of the linker histone by micrococcal nuclease (MNase) results in the formation of the nucleosomal core particle, consisting of the histone octamer onto which 146 bp of DNA is wrapped (Finch *et al.,* 1977). However, in the presence of the linker histone, a metastable intermediate, consisting of the histone octamer, 168 bp of DNA and the linker histone, is formed upon MNase digestion (Simpson, 1978). This structure is referred to as the chromatosome, and is the fundamental structural unit of higher-order chromatin.

1.7.7.2. The involvement of more than one DNA-biding site

Since the first model of the placement of the globular domain of the linker histone on the nucleosome (Allan *et al.*, 1980; discussed below), it was suspected that more than one DNA-binding surface is present on the face of the globular domain. The hypothesis of three DNA-binding sites on the globular domain (Allan *et al.,* 1980; Crane-Robinson and

Ptitsyn, 1989) was only contradicted once the crystal structure of the globular domain of H5 was solved (Ramakrishnan *et al.,* 1993).

The observations that linker histones prefer cross-overs in DNA (section 1.8.2.), as well as the detection of GH5-bridged DNA-structures by electron microscopy (Thomas *et al.* 1992), suggested the presence of more that one DNA-binding site on the surface of GH5. In addition to a primary DNA-binding site formed by helix III of GH5, Ramakrishnan *et al.* (1993) identified a second basic cluster of residues, 25-30 Å away from the primary site, which could potentially interact with DNA. The existence of such a site was unambiguously demonstrated by Goytisolo *et al.* (1996), who replaced several basic GH5 residues (Lys and Arg) in either the primary or secondary DNA-binding site with alanine or glutamic acid. In this way, the net charge of the primary site was changed from +3 to 0, while the net charge of the putative secondary DNA-binding site was changed from +4 to –2. The authors indicated mutations in either site impaired the formation of cooperative tramline DNA complexes, abolished binding to synthetic four-way junctions and that although these mutants were able to bind chromatin, the ability to confer 166 bp chromatosome protection was lost. Although the nature of this binding site remains unclear, it has been shown that the site has a diffuse nature, and that at least some residues involved act as a cluster (Duggan and Thomas, 2000).

1.7.7.3.Models for nucleosomal binding

The location of the linker histone on the nucleosome has been a controversial subject for a number of years. In order to understand the nature of the dispute, a brief account of the models proposed for nucleosomal binding is discussed below.

Noll and Kornberg (1977) were the first to note the location of the linker histone on the DNA connecting nucleosomes. Carbodiimide cross-linking of the globular domain of H1 in intact nuclei, chromatin and nucleosomes, revealed an intimate contact between the globular domain of H1 and the histone octamer, most notably H2A (Boulikas *et al.*, 1980). The same group suggested that the C-terminal tail of the globular domain of H1 is in close contact with linker DNA.

Allan *et al.* (1980) proposed a completely symmetrical setting for GH5. It was observed in this study that the trypsin generated globular domain of H1 alone can protect an additional 20 bp of DNA present in the chromatosome, but cannot facilitate complete chromatin compaction. The authors employed intact H1 and GH1 reconstituted on linker-histone depleted oligonucleosomes as an experimental system. A model was therefore proposed where GH1 is "caged" between the nucleosomal entry/exit DNA and the DNA at the nucleosomal dyad.

Armed with an understanding of the fold of the globular domain and by examining the conservation of basic residues in various linker histone amino acid sequences, Crane-

Robinson and Ptitsyn (1989) identified three DNA-binding surfaces on the face of the globular domain. It was proposed that helices II and III constitute the primary DNA-binding site, and that the remaining two subsidiary DNA-binding sites occur orthagonally to the primary site. This prediction was consistent with the model of Allan *et al.* (1980). In addition, the authors proposed that the primary DNA-binding site binds to the nucleosomal dyad, while the subsidiary sites contacts the linker DNA at the entry and exit region of the nucleosome, respectively. The anchoring of the globular domain to the dyad axis of the nucleosome was consistent with the protection of nucleosomal DNA at the dyad axis of the core particle against DNAse I cleavage (Staynov and Crane-Robinson, 1988).

Ramakrishnan *et al.* (1993), proposed only two DNA-binding sites on opposite faces of GH5 from the X-ray crystal structure. These two sites were calculated to be separated by between 25 and 30 Å. Helix III was reported to be the primary recognition helix, while a positive cluster on the opposite side of the molecule (comprising of Lys 40, Arg 42, Lys 52 and Arg 94) was identified as the putative secondary DNA-binding site. In this study, the mode of binding of GH5 to the nucleosome as proposed by Crane-Robinson and Ptitsyn (1989) was criticized, since it presupposed three DNA-binding sites and an unprecedented mode of binding of helix III to a minor groove, the solvent-exposed area of DNA at the pseudodyad axis of the nucleosomal core particle (Luger *et al.*, 1997).

Lambert *et al.* (1991) was the first to question the symmetry of the chromatosome. Neutron scattering studies suggested that the globular domain should be located closer to

the centre of mass of the structure and therefore between two central gyres of nucleosomal DNA, slightly away from the dyad axis. In this setting, the globular domain of the linker histone bridges two superhelical turns of DNA between one terminus and the dyad.

A radically different, off-axis model was proposed by the Wolffe group (Hayes *et al.*, 1994; Pruss *et al.*, 1996; Hayes, 1996) for homogeneous chromatosomes positioned on the *Xenopus borealis* 5S rRNA gene sequence. The results of micrococcal nuclease (MNase) mapping, site-directed metal-affinity cleavage and cross-linking, were interpreted as showing that the globular domain of H5 as well as its parental variant H1° is located at a single site on the inside of the nucleosomal gyre, approximately 67 bp away from the dyad axis of the core particle. In addition, the putative secondary DNA-binding site was proposed to be in intimate contact with the H2A-H2B dimer. Thomas (1999) pointed out that the lack of effect of mutating the putative secondary binding site of H1° on 5S nucleosomal binding, might be due to the masking of impaired binding by the extended basic C-terminal tail of H1°. It was later shown (Duggan and Thomas, 2000) that mutation of the secondary DNA-binding site of GH5 had the same effect on both bulk nucleosomes and nucleosomes reconstituted on the *Xenopus* 5S sequence.

Hayes (1996) noted that the chromatosome stop observed for the 5S nucleosome (Pruss *et al.*, 1996) was most likely due to core histone rearrangements upon H1-binding and that the asymmetry of binding observed might be a GH5/H1°-specific effect. Indeed, GH5 and the globular domain of *Xenopus* H1° exhibit the same location with respect to a

nucleosome core positioned on a DNA sequence containing a *Xenopus* thyroid hormone response (THR) element (Guschin *et al.,* 1998). The same study also revealed core histone rearrangement upon linker histone association.

Recently, an asymmetrical model for the binding of the globular domain of H5 to bulk chromatin was proposed by Zhou *et al.* (1998). The locations of specific Ser to Cys substitutions were determined by chemical cross-linking. It was concluded that GH5 bridges DNA at one terminus of the chromatosome and DNA in the vicinity of the dyad axis. Contrary to the premise of Crane-Robinson and Ptitsyn (1989), it was shown that helix III recognises a major groove in DNA corresponding to the first helical turn of the chromatosome, while the secondary DNA-binding site contacts nucleosomal DNA close to the dyad axis of symmetry of the octamer. Moreover, it was concluded that helix I faces away from the nucleosome, while helix II faces towards the nucleosome and that the C-terminal tail binds linker DNA exiting or entering the nucleosome.

The main difference between the 5S model and the models of Lambert *et al.* (1991) and Zhou *et al.* (1998), is that the C-terminal tail of GH5 is directed along the upper DNA gyre of the 5S nucleosome towards the nucleosome dyad, invoking a U-turn in the C-terminal tail (Travers, 1999). This highly unlikely distortion is not required in the bridging model. An *et al.* (1998) demonstrated that the *Xenopus* 5S rRNA gene fragment is prone to artefacts in nuclease digestion. Moreover, Panetta *et al.* (1998) showed that a population of core particles and chromatosomes formed on 5S rDNA from *Xenopus borealis* exhibits different individual translational settings. Travers (1999)

pointed out that the contacts deduced from cross-linking (Pruss *et al.,* 1996) and site directed DNA-cleavage (Hayes, 1996) data used to deduce the location of the linker histone in the 5S chromatosome is ambiguous due to the presence of multiple translational frames of the octamer on the 5S fragment.

Belikov and Karpov (1998) examined the location of histone H1 in Alu-repeats in the *Drosophila melanogaster* ribosomal gene spacer by UV-induced cross-linking. Surprisingly, and consistent with the 5S rDNA nucleosome, an asymmetric placement was observed. Two main clusters of cross-links were observed, corresponding to the primary and secondary DNA-binding site, respectively. The primary site (facing the nucleosome) cross-links occurred at +4.5, +5.5 and +6.5, while the secondary site cross-links were observed at –4, -3, -2 and between –0.5 and –1. However, this study was not done at high resolution and did not receive attention in subsequent commentaries on the location of the globular domain on the nucleosome (Thomas, 1999; Travers, 1999).

1.7.7.4. The symmetry of the chromatosome

Another long-standing question is whether the chromatosome is inherently symmetrical or asymmetrical.

Simpson (1978) observed the symmetrical extension of DNA protection (10 bp at both termini of the core particle) against MNase digestion upon linker histone binding. Further support for the symmetry of the chromatosome came from the DNase I protection

experiments on chicken oligonucleosomes by Staynov and Crane-Robinson (1988). In this study, the authors observed a broad 'footprint' with complete protection of the DNA at the core particle dyad. Decreasing partial protection at symmetrical sites on either side of the dyad were also observed, making it clear that chromatosomes from bulk chromatin exhibit symmetrical extensions of 10 bp from both core particle termini. However, the possibility that the protection pattern observed represents the mean of a population of equally preferred off-axis binding sites could not be excluded (Crane-Robinson, 1997), since native chromatin was used. In addition, the asymmetry of the linker histone (Van Holde, 1989) as well as the possible asymmetry of the nucleosomal core particle (Harp *et al.,* 2000) disqualifies the premise that the chromatosome is completely symmetrical. Moreover, a symmetrical model for the chromatosome would not explain the directional binding of linker histones in chromatin (Lennard and Thomas, 1985).

Contrary to the bi-directional 10 bp extension of DNA protection observed in bulk chromatin, Hayes and Wolffe (1993) observed asymmetric extensions of 5 bp and 15 bp from both core particle termini, without protection of DNA at the core particle dyad. This observation was echoed by Belikov and Karpov (1998) for homogeneous chromatosomes positioned on *Drosophila* Alu repeats, while Guschin *et al.* (1998) reported an extension of 20 bp from one terminus and none from the other in a nucleosome positioned on a DNA sequence containing a *Xenopus* thyroid hormone response (THR) element. The recent bulk model of Zhou *et al.* (1998) assumes a symmetrical extension of DNA protection against MNase upon GH5 binding, although the authors mention that the data is also consistent with an asymmetrical extension of

protection. However, Belikov and Karpov (1998) warned that the asymmetrical interpretation observed at least in the case of the model of Wolffe and co-workers might be a consequence of the sequence specificity of MNase.

Recently, DNase I was used to probe compact chromatin in whole nuclei (Staynov, 2000). The protection from digestion of entry/exit DNA in the nucleosome was indirectly attributed to the linker histone tails, via the compact structure sustained by these proteins. In addition, DNase I digestion fragments over several nucleosome repeat lengths were analysed and compared with computer simulations using combinations of the experimentally obtained rate constants. In this way, alternating, asymmetrically protected nucleosomes were observed.

Another argument for the asymmetry of the core particle is the asymmetric sequence organization of the chromatosome discussed in section 1.8.4. (Travers and Muyldermans, 1996). It was shown that an AAA/TTT translational positioning sequence occurs both 95 bp from the chromatosomal terminus (Travers and Muyldermans, 1996) as well as in the vicinity of the dyad of core particle DNA, 75 bp from the core particle terminus (Satchwell and Travers, 1989). Although Travers and Muyldermans did not suggest a possible explanation for the conservation of the AAA/TTT sequence at these two locations, Belikov and Karpov (1998) suggested an interpretation where the AAA/TTT triplet is fixed at the dyad position, in both isolated nucleosomes and chromatosomes, and that the observed shift in the location of this sequence is an artefact of the asymmetry of the chromatosome.

1.7.7.5. A single binding site for the linker histone to the nucleosome

Bates and Thomas (1981) showed that the molar ratio of H1 in lymphocyte and glial nuclei is 1.0 per histone octamer, while liver nuclei contained 0.8 per histone octamer. A value of 0.9-linker histone per octamer was reported for H5 in chicken erythrocytes. Although linker histones and octamer appear to occur at equimolar amounts in compact chromatin, it is not clear, excluding the possibility of steric exclusion (Travers, 1999), why linker histones prefer only one of two seemingly identical binding sites on the nucleosome.

Thomas (1999) suggested the possibility that the octamer might undergo a conformational change upon linker histone binding leading to a polarized nucleosomal array, where linker histones bind directionally. Such nucleosomal arrays were observed by Lennard and Thomas (1985). Further support for this premise comes from the core histone rearrangements observed by Guschin *et al.* (1998) upon linker histone binding. Recently, Harp *et al.* (2000) reported an asymmetrical crystal structure of a nucleosomal core particle, constructed from perfectly two-fold symmetrical chicken histone octamers and a palindromic DNA-sequence. It is therefore also possible that the core particle may be inherently asymmetrical, independent of linker histone binding.

A further possibility is that polarity of nucleosomal arrays may be due to the AGGA-sequence found at the termini of bulk chromatosomes (Thomas, 1999; Travers and Muyldermans, 1996).

1.8. The linker histone and higher order chromatin structure

1.8.1. The 30-nm filament

The linker histone facilitates the compaction of nucleosomal arrays into a dense nucleoprotein filament, the 30-nm fibre (Thoma *et al.,* 1979). In this fibre, nucleosomes are spaced at regular intervals, a feature that has been attributed to the linker histone (Van Holde, 1989). H1/H5 can space nucleosomes regularly *in vitro*, where the spacing is dependent on the stoichiometry of linker histone binding as well as the linker histone isotype (Lauderdale and Stein, 1993). However, this effect is not observed *in vivo* (Linder and Thoma, 1994) where the situation seems to be more complex.

Thoma *et al.* (1979) showed the requirement of H1 for the formation of the 30-nm fibre. Electron microscopy (Hamische *et al.,* 1996) revealed that H1 constrains the exit and entry DNA leaving the nucleosome, bringing adjacent nucleosomes together in space. The requirement of H1 for chromatin compaction was questioned after observations by Schwartz and Hansen (1994) that H1-depleted chromatin compacts in the presence of divalent cations, forming a fast-sedimenting structure. However, the neutron scattering studies of Graziano *et al.* (1988) indicated that there is no correlation between the data obtained for native chromatin at low ionic strength and H1-depleted chromatin at high ionic strength, suggesting that the compaction observed for H1-depleted chromatin at high ionic strength (Schwartz and Hansen, 1994) might not impose structural regularity on the fast-sedimenting species. In addition, Carruthers and Hansen (2000) indicated that the N-termini of the core histones function independently of the linker histone in

chromatin compaction using sedimentation analysis. However, the nucleosomal location of the linker histone is independent of the core histone N-termini (An *et al.,* 1999). It therefore seems likely that the linker histone plays a stabilizing rather than essential role in chromatin compaction.

1.8.2. The location of the linker histone in the 30-nm fibre

Biochemical approaches to determine the location of the linker histone in the 30-nm fibre involved measuring the salt-dependent accessibility of linker histones to either proteolytic cleavage (Leuba *et al.,* 1993) or recognition by antibodies (Banchev *et al.,* 1990). Such approaches yielded conflicting results. Thomas and Rees (1983) noted the exchange of linker histones H1 and H5 in soluble chromatin fragments. More recently, Misteli *et al.* (2000) elegantly demonstrated the dynamic binding of H1-GFP to chromatin *in vivo* and it is therefore possible that the determination of the accessibility of single linker histones in the chromatin fibre is a complex matter.

Neutron scattering offered a way of solving this dispute, since it measures time-averaged location and is not affected by linker histone exchange, or sample heterogeneity. Using this technique, Graziano *et al.* (1994) reported that the centre of mass of deuterated H1 is located at an interior position in the 30-nm fibre. Consistent with neutron scattering studies on the chromatosome (Lambert *et al.,* 1991), the data was interpreted as a location of H1 on the interior face of nucleosomes in the 30-nm fibre. However, this might not

reflect the location of the globular domain of the protein, since this domain constitutes only half of the mass of the protein (Hayes, 1996; Ramakrishnan, 1997).

To investigate whether the role of H1 in chromatin condensation might be due to H1-H1 contacts in the chromatin fibre, several workers examined the cross-linking potential of H1. Thomas and Khabaza (1980) reported cross-linking of H1 molecules in rat liver chromatin with bismidoesters. In addition, the authors demonstrated that cross-linking is highly dependent on a chromatin context. Ring and Cole (1983) purified H1-homodimer from nuclei treated with either long or short cross-linking reagents. Cross-links were analysed by cleavage with appropriate agents. In this way, they were able to show substantial contacts between C-terminal parts of neighbouring H1 molecules, as well as between the N-terminal part of one H1 and the C-terminal part of an adjacent H1. Carter and Van Holde (1998) explored DNA-dependent self-contacts of H5 and its globular domain in solution, and were able to explain their observations in terms of the GH5 crystal dimer, where GH5 molecules make contact between the C-terminal end of one protein and helix III of another. However, it is likely that no self-contacts between adjacent globular domains of H1 occur inside the 30-nm fibre, since the globular domain is located too far (60 – 65 Å) from the fibre axis (Graziano *et al.,* 1994).

Recently, Staynov (2000) found that linker DNA is located on the interior of the 30-nm fibre, while individual nucleosomes are positioned on either side of a helical path alternating their orientation towards the dyad axis.

1.8.3. Conformation of linker DNA in the 30-nm fibre

Mitra *et al.* (1984) provided constraints for linker DNA by photochemical dichroism studies on psoralen-bound chromatin. However, Zlatanova and Van Holde (1995) pointed out that these constraints can only be interpreted as such, assuming a regular chromatin superstructure and that the data of Mitra *et al.* (1984) could be indicative of an irregular 30-nm fibre. Ramakrishnan (1997) pointed out that studies of linker bending in dinucleosomes should be treated with caution, since the linker histone content of these molecules is often depleted. Nevertheless, Butler and Thomas (1998) observed clear linker bending in dinucleosomes by analytical ultracentrifugation. Contrary to this finding, evidence for straight linkers came from Pehrson (1996) who studied the patterns of pyrimidine dimer formation in intact nuclei. Clearly, the question of whether linker DNA is straight or bent remains unresolved.

1.8.4. Models for the structure of the 30-nm fibre

Several models have been proposed for the structure of the 30-nm fibre. These models differ with respect to the connectivity of adjacent nucleosomes, the mass per unit length of the fully compacted fibre, the relationship between the fibre diameter and linker length, the conformation of linker DNA connecting nucleosomes, as well as the location of the linker histone in the fibre (Ramakrishnan, 1997; Zlatanova and Van Holde, 1995). Since the latter two aspects were discussed above, models for the structure of the 30-nm fibre will be discussed in terms of these findings.

The solenoidal model of Finch and Klug (1976) proposed that the 30-nm fibre consists of a helical array of nucleosomes, with 6 to 8 nucleosomes per turn. Moreover, this model places both linker DNA, which is assumed to be bent, and the linker histone on the inside of the filament. A variant of this model, where linker DNA also assumes a bent conformation, is the continuously bent linker model of McGhee *et al.* (1983). According to this model, linker DNA continues the curvature exhibited by the exit DNA of one nucleosome into the entry point of an adjacent nucleosome and is not located at a fixed position in the fibre.

Straight linker DNA could also result in a regular packing of nucleosomes. Woodcock *et al.* (1984) concluded that nucleosomes form a zigzag with linker DNA parallel to the filament axis, while the inside of the filament is hollow. In other words, the fibre forms a helical ribbon with linker histones located at its periphery. Another possibility was suggested by Williams *et al.* (1986), who proposed a crossed-linker model where DNA criss-crosses between nucleosomes in a direction perpendicular to the filament axis. In this way, linker histones are forced to the interior of the filament. Both this model, and the original solenoidal model are consistent with the neutron scattering data of (Graziano *et al.*, 1994).

Woodcock *et al.* (1994) proposed an irregular model by observing chromatin fibres *in situ* on frozen hydrated sections. However, in this model, the native fibre diameter is not correlated with the nucleosomal repeat length. According to the model of Zentgraf and

Franke (1984), nucleosomal arrays do not resemble a filament, but clusters called 'super-beads' that contain different numbers of nucleosomes.

1.9. Concluding remarks

Although a lot of work has been published on the structure of the linker histone, a number of aspects need to be addressed in greater detail. Firstly, the structure and role of the linker histone tails are still unclear. In addition, since H1 might constitute an ATP-sensor (Tarkka *et al.,* 1997), an investigation into the allosteric properties of this protein might shed light on its biological role. The sequence specificity of H1 subtypes as well as the effect of sequence markers on the location of the globular domain on the nucleosome should also be addressed in greater detail, since these might have an effect on the role of H1 in the differential regulation of transcription. It is also not clear whether the primary mode of recognition of linker DNA by the globular domain is via helix III or rather via the β-hairpin motif. Moreover, since there seems to be clear distinction between the location of the globular domain of the linker histone on nucleosomes positioned on 5S rDNA from *Xenopus borealis* and nucleosomes from bulk chromatin, a high-resolution method to probe the location of the globular domain on other 5S sequences, as well as well-characterized DNA fragments seems to be required.

In terms of higher order chromatin structure, the asymmetry of the linker histone, as well as the related problem of a single binding site of this domain to the nucleosome, remains a point of dispute. Indeed, the effect of conformational changes in the core histone

octamer on the symmetry of the chromatosome upon H1-binding remains to be determined. In addition, the necessity of H1 in promoting chromatin compaction is also questionable. Although H1 seems to be able to bend DNA, the conformation of linker DNA in the compact chromatin fibre is unresolved.

1.10. Aim of this study

The aim of this study is to determine the location of the secondary DNA-binding site of the globular domain of linker histone H5 on the nucleosome. This problem will be addressed at two levels. Firstly, a theoretical model of the binding of GH5 to the nucleosome will be generated by homology to the co-crystal structure of *E. coli* CAP. Secondly, the model will be evaluated experimentally by employing site-directed metal affinity cleavage to determine the binding site of GH5 to nucleosomes positioned on the *Lytechinus variegatis* 5S gene sequence, as well as a histone gene spaces from *Psammechinus miliaris*. Both these sequences are known to contain strong nucleosomal phasing signals.

For the purpose of this M.Sc. dissertation, the following components of this project are reported. A model for the location of GH5 on the nucleosome is discussed (Chapter 3). In addition, the over-expression of mutant GH5 proteins, engineered to aid in the determination of the location of GH5 on the nucleosome is described (Chapter 4). Lastly, a novel route to the synthesis of *S*-(nitrophenylsulphenyl)cysteaminyl-EDTA, the cleavage reagent to be employed, is investigated (Chapter 5). Although not included as

part of this dissertation, the testing of the model proposed in this study by site-directed metal-affinity cleavage is currently underway.

CHAPTER 2: MATERIALS AND METHODS

2.1. Computer modelling

2.1.1. Crystal structures

The A-chain of the dimeric crystal structure of the globular domain of H5 (Ramakrishnan *et al.*, 1993) was used in all modelling procedures. The co-crystal structure of the CAP-DNA complex (Parkinson *et al.*, 1996) was used as structural template to model the primary DNA-binding site of GH5. The secondary DNA-binding site was modelled on the DNA of the crystal structure of the nucleosome solved by Luger *et al.* (1997). The core histones were reconstituted to this structure by RMS superposition (PDBViewer; Guex and Peitsch, 1997).

2.1.2. Software

Structural alignments were performed using the web-based DALI server (Holm and Sander, 1995). C_α and charge centre superpositions were performed using Hyperchem (Hypercube, Inc., Gainesville, Florida, USA) while RMS superpositions were performed using either the PDBViewer (Guex and Peitsch, 1997) or WhatIf (Vriend, 1990). Coordinate manipulations were performed using Hyperchem and WhatIf. Inter-atomic distances were calculated using Hyperchem and the PDBViewer. Hydrogen bonds were detected using Hyperchem with a minimum heteroatom distance of 1.20 Å, a maximum

distance of 2.81 Å and a minimum angle of 120°. Therefore, a hydrogen bond was only recognised as such when the distance between the hydrogen bond donor and acceptor was between 1.20 Å and 2.81 Å, and the angle formed between these two atoms and the hydrogen was between 120° and 180°. A distance cut-off of 4 Å was used to define ionic interactions. Molecular surfaces were calculated with the PDBViewer. Energy minimisations were carried out with Hyperchem using the AMBER force field (Cornell *et al.*, 1995) and default parameters, except that the termination parameter was set to 0.1 kcal/Å.mol. Residue mutations were done with Hyperchem or PDBViewer.

2.2. Organic syntheses

2.2.1. Chemical compounds and solvents

For all syntheses, the highest quality chemicals available were employed. Fluka (Sigma-Aldrich, Pf., Steinheim, Germany) *Microselect*, *Purum* or *Puriss.* were used where possible. Distilled solvents were used for chemical extractions, while analytical grade reaction solvents were used.

2.2.2. General

Chromatography:

For all TLC analyses, aluminium-back F^{254} silica plates (Merck, KgaA, Darmstadt, Germany) were employed. Substances to be examined were dissolved in eluent, applied to the TLC plate in small spots using a glass capillary, and the sample dried before the plate was put in the eluent. The separated samples were visualised by iodine vapour, 2% ninhydrin in ethanol and cerium ammonium sulphate or with UV transillumination at 250 nm, as indicated in the text. Column chromatography was carried out using Merck Kieselgel 60: 70-23- mesh for gravity columns and 230-400 mesh for flash chromatography.

^{1}H NMR Spectroscopy:

^{1}H NMR spectra were recorded on a Varian Unity Spectrometer

2.2.3. Synthesis of *S*-(*tert*-butyl)-cysteamine

The synthesis of *S*-(*tert*-butyl)-cysteamine was performed exactly as described by Flaus and Richmond (1999). In accordance with this procedure, 5.6 g (50 mmol) of cysteamine-HCl was added to 10 mℓ *tert*-butanol and 25 mℓ 2 M HCl in a 100 mℓ round-bottom flask. The mixture was gently refluxed for 14 h. Excess *tert*-butanol was removed by rotary evaporation at 50°C, and the product was allowed to crystallise from the remaining HCl at room temperature. The crystals were recovered by filtration and dried in a desiccator. The purity of the product was determined by TLC [*n*-butanol:acetic acid:water (4:1:1) stained with cerium ammonium sulphate] and the identity of the product was confirmed by spectroscopic methods (^{1}H NMR, data not shown).

2.2.4. Synthesis of triethyl-EDTA

The selective hydrolysis of one ethyl group on the tetra-ethyl ester of EDTA (kindly provided by E. Sickle, University of Cape Town) was performed as described previously (Burks *et al.*, 1998) with the following minor changes. EDTA (2.00 g, 4.95 mmol) was dissolved in 10 mℓ acetone and 190 mℓ 50 mM sodium phosphate buffer (pH 8.0). Pig Liver Esterase (PLE, Aldrich (Sigma-Aldrich) 37305-2), 170 U/mmol tetra-ethyl-EDTA in an ammonium sulphate suspension, was added and the suspension was shaken at 250 rpm, in a 500 mℓ conical flask in a shaking incubator at 27°C for 5.5 h. It was observed that this procedure could also be performed by stirring using a magnetic stirrer at 25°C or 30°C, without significant difference in the yield of the reaction (data not shown). The

reaction mixture was washed four times with 100 mℓ petroleum ether to remove any unreacted tetra ethyl ester and the product extracted by washing three times with 100 mℓ dichloromethane. The dichloromethane extracts were combined, and dried by filtration through solid anhydrous $MgSO_4$. The solvent was removed by rotary evaporation and the resulting product, a clear viscous oil, was dried by a high-vacuum pump. The purity of the compound was determined by TLC [ethyl acetate: hexane (1:1) negatively stained by cerium ammonium sulphate or detected by UV]. The identity of the compound was confirmed by ^{1}H-NMR (300 MHz, $CDCl_3/D_2O$) 1.16 (m, 9H), 2.6 - 2.7(m, 4H), 3.1 -3.3 (m, 8H) and 4.07(m, 6H).

2.2.5. Synthesis of *S*-(*tert*-butyl)-cysteaminyl-triethyl-EDTA

As water negatively affects the yields obtained in this reaction (Kerstin Bysticki, personal communication), the following procedure was carried out under anhydrous conditions. Tri-ethyl EDTA (3.047 g, 8.095 mmol; Fluka), 1-hydroxybenzotriazole (1.18 g, 8.11 mmol; Fluka) and dicyclohexylcarbodiimide (1.9 g, 9.21 mmol; Fluka) was weighed into a round-bottom flask fitted with a drying tube. 38.6 mℓ dry *N,N*-dimethylformamide (Fluka) was added after which 6.6 mℓ *N*-ethyldiisopropylamine (50.76 mmol; Fluka) was added. The mixture was cooled to -10°C in an ice-salt bath, *S*-(*tert*-butyl)-cysteamine (0.907 g, 8.098 mmol) was added to the cooled mixture, which was stirred overnight on the ice-salt bath. During this time the reaction mixture was allowed to warm to room temperature. Ethyl acetate (250 mℓ) was added, and the mixture was filtered through a 2 cm layer of Celite to remove all insoluble precipitate. The ethyl acetate was removed by

rotary evaporation at room temperature and the product was allowed to crystallise from the remaining *N,N*-dimethylformamide fraction at -20°C overnight. White, shiny, needle-like crystals were recovered from the mother liquor by filtration, and washed with ethyl acetate to remove any remaining *N,N*-dimethylformamide. The product was dried under high vacuum and the identity of the product was confirmed by ^{1}H-NMR, (300 MHz, $CDCl_3/D_2O$) 1.07 (m, 9H), 1.23 (s, 9H), 2.4 - 2.5 (m, 6H), 2.63 (br s, 2H), 3.0-3.3 (m, 8H) and 4.1- 4.3 (m, 6H).

2.2.6. Synthesis of *S*-(nitrophenylsulphenyl)-cysteaminyl-triethyl-EDTA

S-(*tert*-butyl)-cysteaminyl-triethyl-EDTA (155 mg, 245 mmol) was stirred for 150 min with 2-nitrophenylsulphenylchloride (108.5 g, 368 μM; Fluka or Aldrich) in 15.5 mℓ acetic acid at room temperature. Distilled water (15 mℓ) was added. The material was extracted three times with 20 mℓ ethyl acetate. Organic extracts were combined and washed once with 20 mℓ of distilled H_2O. Acetic acid was neutralised by repeatedly washing with saturated aqueous $NaHCO_3$ (20 mℓ per wash), until effervescence ceased upon the addition of further $NaHCO_3$. The material was washed once with 20 mℓ of a saturated aqueous solution of sodium chloride and dried over solid anhydrous $MgSO_4$. Solvent was removed by rotary evaporation to give bright yellow crystals, which were dried under high vacuum. Material was analysed by TLC [ethyl acetate: petroleum ether (1:1), detection by UV]. In order to obtain a ^{1}H-NMR spectrum, a small amount of material was purified by chromatography on silica gel. ^{1}H-NMR (300 MHz, $CDCl_3/D_2O$)

1.08 (m, 9H), 2.5 - 2.7 (m, 6H), 2.99 (m, 2H), 3.1 -3.3 (m, 8H), 4.2 - 4.4 (m, 6H), 7.40 (td, 1H), 7.58 (ddd, 1H), 7.86 (dd, 1H) and 8.34 (dd, 1H).

2.2.7. Synthesis of *S*-(nitrophenylsulphenyl)-cysteaminyl-EDTA from *S*-(nitrophenylsulphenyl)-cysteaminyl-triethyl-EDTA

S-(nitrophenylsulphenyl)-cysteaminyl-triethyl-EDTA crude extract (38 mg, 67 μmol) was dissolved in 5 mℓ methanol. Five equivalents (46 mg) solid K_2CO_3 was added and the mixture was stirred over night at room temperature. The mixture was filtered through 2 cm of Celite to remove any precipitate and the solvent was removed from the filtrate by rotary evaporation at room temperature (25°C). The product was dissolved in 250 mℓ ethylacetate and dried over solid anhydrous $MgSO_4$. The ethyl acetate was removed by rotary evaporation, dried under high vacuum, and the material analysed by TLC [ethyl acetate : petroleum ether (1:1), detection by UV]. The identity of the product was confirmed by ^{1}H-NMR (300 MHz, DMSO/D_2O) 2.5 - 2.8(m, 6H), 3.0(m, 2H), 3.1 -3.3 (m, 8H), 7.40(td, 1H), 7.58(ddd, 1H), 7.86(dd, 1H) and 8.34(dd, 1H). The product was stored dry at 4°C.

2.2.8. Synthesis of *S*-(*tert*-butyl)-cysteaminyl-EDTA

This compound was synthesised and purified according to the method proposed by Flaus and Richmond (1999). As water negatively affects the yields obtained in this reaction, the following procedure was carried out under anhydrous conditions. Acid-form EDTA

(3.047 g, 8.095 mmol), 1-hydroxybenzotriazole (1.18 g, 8.11 mmol) and dicyclohexylcarbodiimide (1.9 g, 9.21 mmol) was weighed into a round-bottom flask fitted with a drying tube. A 38.6 mℓ volume of dry *N,N*-dimethylformamide was added after which 6.6 mℓ *N*-ethyldiisopropylamine (50.76 mmol) was added. The mixture was cooled to -10°C by stirring in an ice-salt bath. *tert*-Butyl-cysteamine (0.907 g, 8.098 mmol) was added to the cooled mixture, and the mixture was stirred overnight. During this time the reaction mixture was allowed to warm to room temperature. The material was filtered through 2 cm of Celite and *N,N*-dimethylformamide was removed by repeatedly preheating the flask to 50°C in a water bath and connecting it to high vacuum in between. The dried material was re-suspended in 10 mℓ 0.1% (v/v) trifluoroacetic acid and centrifuged repeatedly at top speed (12 000 × g) in ten aliquots in a microcentrifuge until no more precipitate was present. The material (2 mℓ per run) was loaded onto a C_{18} reverse phase HPLC column (Hiasil 100, 5 µm; 250 mm × 10 mm; Higgins analytical) equilibrated in degassed 0.1% trifluoroacetic acid (HPLC buffer A). The HPLC was performed at a flow rate of 3.5 mℓ/min. Impurities were separated from the product by applying a linear gradient from 0 to 80% acetonitrile in 0.1% (v/v) TFA over 55 minutes. The liquid chromatograph time function is described in more detail by Table 2.1. Peak fractions were collected and the product was identified by mass-spectrometry (MALDI-TOF) in a 2,5-dihydrobenzoic acid matrix. The product eluted with a retention time of 20 min. Acetonitrile was removed by rotary evaporation at 25°C, after which the remaining aqueous fraction was frozen on an ethanol bath and lyophilised.

Table 2.1. Liquid chromatograph time function for the purification of *S*-(*tert*-butyl)-cysteaminyl-EDTA.

Step	*Time (minutes)*	*Buffer $B_{1)}$ (%)*
1	5	0
2	30	50
3	35	100
4	40	100
5	50	0
6	60	0
7	61	3.5

1) The composition of the liquid phase is indicated as % (v/v) of Buffer B (80% (v/v) acetonitrile; 0.1% (v/v) trifluoroacetic acid). The remainder of the liquid phase consisted of Buffer A (0.1% (v/v) trifluroacetic acid).

Table 2.2. Liquid chromatograph time function for the purification of *S*-(nitrophenylsulphenyl-cysteaminyl)-EDTA

Step	*Time (minutes)*	*Buffer $B_{1)}$ (%)*
1	0.1	20
2	10	20
3	25	50
4	65	100
5	70	100
6	75	0
7	85	0

1) The composition of the liquid phase is indicated as % (v/v) of Buffer B (80% (v/v) acetonitrile; 0.1% (v/v) trifluoroacetic acid). The remainder of the liquid phase consisted of Buffer A (0.1% (v/v) trifluroacetic acid).

2.2.9. Synthesis of *S*-(nitrophenylsulphenyl)-cysteaminyl-EDTA from *S*-(tert-butyl)-cysteaminyl-EDTA

This compound was synthesised and purified according to the method proposed by Flaus and Richmond (1999). *S*-(*tert*-butyl)-cysteaminyl-EDTA (300 mg, 230 μM) was stirred for 150 min with 2-nitrophenylsulphenylchloride (70 g, 368 μM) in 10 mℓ acetic acid at 25°C. Acetic acid was removed by rotary evaporation and the solid product was re-suspended in 8 mℓ 0.1% (v/v) trifluoroacetic acid and centrifuged repeatedly at maximum speed in a microcentrifuge until no precipitate was present. The material was loaded (2 mℓ per run) onto a C_{18} reverse phase HPLC column (Hiasil 100, 5μm; 250 mm × 10 mm; Higgins analytical) equilibrated in 20% (v/v) acetonitrile. Impurities were separated from the product by applying a linear gradient from 20% to 80% acetonitrile over 60 min. This procedure was performed essentially as described for the synthesis of *S*-(nitrophenylsulphenyl)-cysteaminyl-EDTA from *S*-(nitrophenylsulphenyl)-cysteaminyl-tri-ethyl-EDTA (see above and Table 2.2). Peak fractions were collected and the product identified by mass-spectrometry (MALDI-TOF) in a 2,5-dihydrobenzoic acid matrix. The product (detected at 277 nm) eluted with a retention time of 20 min. Acetonitrile was removed by rotary evaporation at room temperature, after which the remaining aqueous fraction was frozen on an ethanol bath and lyophilised.

2.2.10. Mass spectrometry (MALDI-TOF) of chemical products

Aliquots of HPLC-purified products (1 mℓ) were concentrated in a rotary evaporator (Savant) overnight to a final volume of 20 μℓ. A 5 μℓ aliquot of concentrated sample was mixed with 2,5-dihydroxybenzioc acid (10 mg/mℓ) and spotted onto a gold-coated MALDI sample plate. The matrix, 2,5-dihydrobenzoic acid (10 mg/mℓ) was also spotted onto the grid at a separate location. The grid was air-dried and inserted into a MALDI-TOF mass spectrometer for analysis.

2.3. Molecular biology techniques

All reagents used were Molecular Biology Grade. Unless mentioned otherwise, all techniques were performed according to standard protocols (Ausubel *et al.,* 1995).

2.3.1. Plasmid purification

Small-scale plasmid purification was performed using the Wizard Miniprep Kit (Promega Corp., Madison, USA). Plasmid DNA was eluted in TE (10 mM Tris, 1 mM EDTA, pH 8.0) for storage, or Milli-Q water in cases where enzymatic manipulations of DNA required free metal ions in solution.

2.3.2. Oligonucleotides

2.3.2.1.Oligonucleotide purification

Lyophilised synthetic primers (Integrated DNA Technologies, Inc. or Oligonucleotide Synthesis Facility, Department of Molecular and Cell Biology, University of Cape Town) were suspended in TE to a concentration of 200 μM. A 100 μℓ aliquot of each oligonucleotide was loaded with 100 μℓ formamide loading buffer onto a 15% (w/v) polyacrylamide gel containing 50% (w/v) urea. Oligonucleotides were electrophoresed for 2 h at 200 V in 1 × TBE buffer (89 mM Tris base, 89 mM borate, 2 mM EDTA).

Following electrophoresis, oligonucleotides were visualised by UV-shadowing on F^{254} silica plates (Merck). Appropriate bands were excised from the gel and incubated for 12 h at 37°C in 1 mℓ elution buffer (0.3 M sodium acetate, pH 5.2). Residual gel particles were removed by centrifugation in a desktop centrifuge at 10 000 × g for 5 min, and the supernatant filtered through sterile glass wool. The recovered samples were extracted with an equal volume of neutralised phenol and an equal volume of chloroform:isoamylalcohol (24:1; v/v), followed by precipitation twice with 95% ethanol according to standard procedures (Ausubel *et al.,* 1995). Oligonucleotides were dried in a rotary evaporator (Savant) and resuspended in 1 × TE.

2.3.2.2. Oligonucleotide sequences

In all oligonucleotide sequences listed below, sequences engineered to introduce restriction enzyme recognition sites are underlined. The restriction endonuclease *Nde* I recognises the sequence 5'-catatg-3', while *Xho* I recognises 5'-ctcgag-3'.

The following oligonucleotides were used to amplify the coding sequence for residues 20 - 109 of chicken GH5: glob(f), 5'-gggcatatgTCGCGGCGCTCGGCATCGCACCCCA-3' and glob(r), 5'- gggctcgagCTTCTTCCCGGGGGACCTCT-3'.

The following oligonucleotides were used as flanking primers for PCR mutagenesis: MTE-glob(f): 5'-gggcatatgatgacggagTCGGCATCGCACCCCAC-3' and MTE-glob(r): 5'-gggctcgagGGGGGGGACCTCTTGGCCTTG-3'. The final amplification products

corresponded to the coding sequence for residues 23 – 106 of chicken GH5. These oligonucleotides were engineered to introduce the coding sequence for an N-terminal MTE-motif (5'-ATGGACGGAG-3') into final amplification products, a requirement for over-expression of the domain corresponding to residues 23 - 106 in *E. coli* (Gerchman *et al.*, 1993).

Three sets of template mismatch oligonucleotides were employed: S41C(f), 5'-CGTGAAAAGTGCCGCGGCGGCTCC-3' and S41C(r), 5'-GGAGCCGCCGCGGCACTTTTCCGCACG-3' were designed to substitute the Ser 41 codon with a cysteine codon. The codon for alanine at position 96 was substituted with a codon for cysteine using A96C(f), 5'-TCCTTCCGCTTGTGCAAGAGCGACAAG-3' and A96C(r), 5'-CTTGTCGCTCTTGCACAAGCGGAAGGA-3', while the oligonucleotide pair R37C(f), 5'-GCGGCGGCCATCTGCGCGGAAAAGAGC-3' and R37C(r), 5'-GCTCTTTTCCGCGCAGATGGCCGCCGC-3' were used to substitute the Arg 37 codon for cysteine.

2.3.3. Polymerase Chain Reactions

The template for amplification of the domains of H5 corresponding to both residues 20 – 109 and 23 – 106, was pET9-H5 (kindly provided by V. Ramakrishnan).

2.3.3.1. Amplification of native H5 domains

In order to amplify the coding sequence for GH5 corresponding to residues 20 – 109 by the Polymerase Chain Reaction (PCR), 10 ng of pET9-H5 was added to 5 μℓ 10 × *Taq* buffer (Promega), 5 U *Taq* DNA polymerase (Promega), 120 μM of each of the deoxyribonucleotide triphosphates, 1.5 mM magnesium chloride and 3 μM of both glob(f) and glob(r) in a reaction volume of 50 μℓ. The template was denatured at 94°C for 1 min, followed by 25 cycles of a 1 min denaturing step (94°), 1 min annealing step (55°C) and one-minute extension step (72°C). Finally, partial extension products were allowed to complete by incubating the sample at 72°C for 5 min, after which the sample was kept at 4°C.

After completion of PCR amplifications, 10 μℓ of samples was loaded on 1% (w/v) analytical agarose gels in 1 × TAE electrophoresis buffer (40 mM Tris acetate, 2 mM EDTA) stained with ethidium bromide (0.5 μg/mℓ) and electrophoresed at 120 V for 40 min. Reaction products were visualised by short wavelength UV transillumination. Hereafter, 40 μℓ of the sample was loaded onto a 1% (w/v) low melting point agarose gel in 1 × TAE, containing 0.5 μg/mℓ ethidium bromide, and electrophoresed for 70 min at 60 V. Reaction products were visualised by brief exposure to long wavelength UV, and the appropriate bands were excised with a sterile blade. Gel slices were melted by incubation at 70°C and DNA purified using the Wizard PCR Preps Kit (Promega) using the manufacturer's supplied protocol.

Using identical cycling conditions to those described above, the H5 domain corresponding to residues 23 – 106 was amplified by PCR. However, this procedure differed in that it involved adding 10 ng of template DNA (pET9-H5) to 5 μℓ of 10 × *Pfu* DNA polymerase buffer (Promega), 3 U of *Pfu* DNA polymerase (Promega), 120 μM of each of the deoxyribonucleotide triphosphates and 0.5 μM of both MTE-glob(f) and MTE-glob(r) in a total reaction volume of 50 μℓ. PCR products were precipitated from 95% (v/v) ethanol and 0.3M sodium acetate according to standard procedures (Ausubel *et al.*, 1995), dried in a rotary evaporator and resuspended in 44 μℓ of milli-Q water and 5 μℓ of 10 × *Taq* in buffer (Promega). This procedure was essential to desalt the sample. *Taq* DNA polymerase was added (5 U) and the sample was denatured at 94°C for 1 min followed by 3 cycles of a 1 min denaturing step (94°), 1 min annealing step (55°C) and 1 min extension step (72°C). During these cycles, *Taq* DNA polymerase incorporates a 3' deoxyadenosine in a template-independent manner into PCR products, facilitating subsequent ligation of PCR products into the pGEM-T Easy single nucleotide T-overhang vector (Promega). Following completion of the reaction, the PCR products were analysed on a 1% (w/v) agarose gel and purified from 1% (w/v) low melting agarose as described. Reaction products were analysed on a 1% (w/v) agarose gel and purified from 1% (w/v) low melting agarose as described above.

2.3.3.2. Template mismatch PCR mutagenesis

This two-step procedure was based on the method of Kadowaki *et al.* (1989). Firstly, separate amplification of overlapping areas of the coding sequence for GH5 (residues 23 - 106) was performed by PCR. This was achieved by adding 5 ng of template DNA (pET9-H5) to 10 × *Pfu* DNA polymerase reaction buffer (Promega), 3 U of *Pfu* DNA polymerase (Promega), 120 μM of each of the deoxyribonucleotide triphosphates, 0.5 μM of mismatch oligonucleotide needed to incorporate a specific cysteine codon, and 0.5 μM of a flanking oligonucleotide (MTE-glob(f) or MTE-glob(r)) in a total reaction volume of 50 μL.

In order to generate two overlapping amplification products needed in the construction of GH5 mutant A96C, the above reaction was performed twice, using A96C(f) as mismatch primer together with MTE-glob(r) as flanking primer in one reaction, and A96C(r) as mismatch primer together with MTE-glob(f) as flanking primer in a separate reaction. Similarly, in order to create overlapping segments coding for S41C, S41C(f) in conjunction with MTE-glob(r), and S41C(r) in conjunction with MTE-glob(f) were used in separate amplification reactions. Oligonucleotides R37C(f) together with MTE-glob(r) and R37C(r) together with MTE-glob(f) were used to generate overlapping fragments of the sequence coding for A37C.

PCR products were analysed on 1% (w/v) agarose gels stained with 0.5 μg/mℓ ethidium bromide and purified from 1% (w/v) low melting point agarose gels as described.

Products were diluted to a concentration of 3 ng/μℓ with milli-Q water. The following combinations of purified PCR products were used as template in subsequent overlap extension PCR reactions: The product of the PCR employing the oligonucleotide pair MTE-glob(f) and S41C(r) as well as the PCR product of oligonucleotides MTE-glob(r) and S41C(f) were used as template for the amplification of the full-length coding sequence of mutant S41C. The product of the PCR employing the oligonucleotide pair MTE-glob(f) and R37C(r) as well as the PCR product of oligonucleotides MTE-glob(r) and R37C(f) were used as template for the amplification of the full-length coding sequence of mutant R37C, while the oligonucleotides MTE-glob(f) and A96C(r) as well as the PCR product of the oligonucleotide pair MTE-glob(r) and A96C(f) were used as template for the amplification of the full-length coding sequence of mutant A96C.

Full-length mutant coding sequences were amplified by overlap extension PCR. According to this technique, 6 ng of template DNA (3 ng of both of the appropriate overlapping PCR products as discussed) was added to 10 × *Pfu* DNA polymerase reaction buffer (Promega), 3 U of *Pfu* DNA polymerase (Promega) and 120 μM of each of the deoxyribonucleotide triphosphates, in a total reaction volume of 48 μℓ. This was followed by an extension procedure employing 2 cycles of denaturation for 1 min at 94°C, template annealing for 1 min at 55°C and *Pfu*-catalysed extension for 1 min at 72°C. This procedure allowed the formation of full-length template prior to the initiation of PCR and was performed to minimise the amplification of possible full-length contaminants. Hereafter, MTE-glob(r) and MTE-glob(f) was added to a final concentration each of 0.5 μM and DNA was denatured for 1 min at 94°C, followed by 25

cycles of a 1 min denaturing step (94°), 1 min annealing step (55°C) and 1 min extension step (72°C). Finally, partial extension steps were allowed to complete by incubating the sample at 72°C for 5 min. PCR products were precipitated with 95% (v/v) ethanol according to the standard protocol (Ausubel *et al.,* 1995), dried in a rotary evaporator (Savant) and resuspended in 44 μℓ of milli-Q water and 5 μℓ of 10 × *Taq* in buffer (Promega). This procedure was essential in desalting the sample. *Taq* DNA polymerase was added (5 U) and the sample was denatured at 94°C for 1 min followed by 3 cycles of a 1 min denaturing step (94°), 1 min annealing step (55°C) and 1 min extension step (72°C). During these cycles, *Taq* DNA polymerase incorporates a 3' deoxyadenosine in a template-independent manner into PCR products, facilitating subsequent ligation of PCR products into the pGEM-T Easy single nucleotide T-overhang vector. Following completion of the reaction, PCR products were analysed on a 1% (w/v) agarose gel and purified from 1% (w/v) low melting agarose as described.

2.3.4. Sub-cloning of PCR products into pGEM-T Easy vector

PCR products (corresponding to full-length native and mutated domains coding for GH5) were subcloned into the pGEM-T vector using the pGEM-T Easy Kit (Promega). The procedure was performed according to the manufacturer's recommendations. Briefly, 14 ng of PCR product (in three-fold molar excess to vector) was added to 50 ng pGEM-T Easy vector (Promega), 5 μℓ 2 × Rapid Ligation Buffer (Promega) and 1 μℓ of T7 DNA ligase (Promega, 3 Weiss U/μℓ) in a total volume of 10 μℓ. Two experimental controls were performed: a background control where the PCR product was omitted, and a

positive control where the PCR product was substituted with control insert DNA, supplied by the manufacturer (Promega). All reactions were incubated overnight at 4°C in 1.5 mℓ eppendorf tubes. Hereafter, 4 μℓ of each ligation reaction was added to 3 eppendorf tubes containing 100 μℓ of calcium chloride competent XL1 Blue cells, freshly thawed on ice. The suspension was mixed gently and incubated on ice for 20 min after which the cells were heat-shocked at 42°C for 40 s. The cells were immediately returned to ice for two min and 900 μℓ of LB medium (5 g/ℓ yeast extract, 5g/ℓ sodium chloride, 10 g/ℓ tryptone peptone, 1mM NaCl) equilibrated at room temperature was added. Cells were incubated in a shaking incubator (150 rpm) for 1.5 hr at 37°C. Following incubation, 100 μℓ of each cell suspension was plated on LB agar plates, supplemented with ampicillin (50 μg/mℓ; USB, Amersham Pharmacia Biotech UK Ltd., Buckinghamshire, England), isopropyl-β-D-thiogalactoside (IPTG; 0.5 mM; Roche) and 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal; 0.05 μg/mℓ; USB). Plates were grown for a maximum of 14 hrs at 37°C. Putative colonies that were positive for inserts were identified by blue/white selection. Individual white colonies were inoculated into 5 mℓ LB culture medium, supplemented with 50 μg/mℓ ampicillin and grown to an OD_{600} of approximately 0.6, after which plasmid was isolated according to the miniprep procedure described. Plasmid (0.5 μg) was digested with 10 U *Pst* I (New England Biolabs Inc., Beverly, MA, USA) for 2 hr at 37°C in order to verify the presence of the correct inserts in plasmid constructs. The pGEM-T Easy vector contains a single *Pst* I restriction enzyme recognition site, while pGEM-T Easy constructs containing the correct GH5 inserts, contain 2 *Pst* I recognition sites, resulting in fragment sizes of either 142 or

162 bp and a fragment of approximately 3 kb following *Pst* I cleavage, depending on the orientation of the insert. Digested constructs were analysed on a 1% (w/v) agarose gel, and constructs found to contain correct inserts were sequences by automated nucleotide sequencing. Glycerol stocks (15% v/v) were prepared of *E. coli* XL1 Blue cells found to contain the correct inserts and stored at -70°C.

2.3.5. Sub-cloning into pET20b(+) expression vector

Approximately 1 μg of pGEM-T Easy constructs found to contain the correct inserts as judged by automated nucleotide sequencing, was digested with 10 U *Nde* I (New England Biolabs) for 3 hr at 37°C. Hereafter, DNA was precipitated from 95% ethanol and 0.3 M sodium acetate according to standard procedures (Ausubel *et al.,* 1995) and digested with 10 U *Xho* I (Amersham) for 1 h at 37°C. Digestion products were analysed on a 1% (w/v) agarose gel and fragments (288 bp and 279 bp for MTE-GH5 constructs and GH5 constructs, respectively) were recovered from 1% (w/v) low melting agarose gels. Approximately 14 ng of recovered DNA fragment (in three-fold molar excess to vector) was added to 50 ng of the purified 3 kb fragment from *Nde* I and *Xho* I digested pET20b(+) vector (Novagen Inc., Madison, WI, USA), 1 μℓ 10 × T7 DNA ligase reaction buffer (Promega) and 1 μℓ of T7 DNA ligase (Promega, 3 Weiss U/μℓ) in a total volume of 10 μℓ. A background control was included in these experiments, where DNA fragments to be ligated into the vector were omitted. Samples were incubated overnight at 4°C. A 4 μℓ aliquot of each ligation reaction was added to 3 eppendorf tubes containing 100 μℓ of calcium chloride competent XL1 Blue cells, freshly thawed on ice.

Cells were transformed as described. Hererafter, plasmid was extracted from transformed cells as described. Plasmid (0.5 μg) was digested with 10 U *Pst* I for 2 h at 37°C in order to verify the presence of the correct inserts in plasmid constructs. The pET20b(+) vector contains a single *Pst* I restriction enzyme recognition site, while pET20b(+) constructs containing the correct GH5 inserts, contain 2 *Pst* I recognition sites, resulting in a 1419 bp and 2441 bp DNA fragment following *Pst* I cleavage. Digested constructs were analysed on a 1% (w/v) agarose gel, and constructs found to contain correct inserts were sequences by automated nucleotide sequencing. Glycerol stocks (15%, v/v) were prepared of positive clones and stored at -70°C. Expression plasmids containing the correct GH5 inserts (100 ng per transformation) were transformed into *E. coli* BL21(DE3)pLysS cells (Novagen) for subsequent expression.

2.3.6. Expression of recombinant GH5 proteins in *E. coli*

In order to establish whether proteins were stably expressed by BL21(DE3)pLysS cells harbouring GH5 expression constructs, a single colony of BL21(DE3)pLysS cells containing pET20b(+)-GH5 or pET20b(+)-MTE-GH5 and its corresponding mutants, was inoculated into 10 mℓ of LB medium supplemented with 50 μg/mℓ ampicillin and 30 μg/mℓ chloramphenicol. Cultures were grown at 37°C to an OD_{600} of 0.7 after which 1 mℓ was removed to serve as pre-induction control. IPTG was added to a final concentration of 0.4 mM and the cultures were returned to 37°C. Aliquots (1 mℓ) were taken at various time intervals (45, 90, 135 and 180 min for native domains, and 60, 120 and 180 min for mutant domains). All 1 mℓ aliquots were centrifuged for 10 s in a

desktop centrifuge, after which medium was decanted and the cell pellet frozen at -70°C. Following collection of the last aliquot, 20 μℓ of 5 × SDS-PAGE sample application buffer (0.5 M Tris·Cl (pH 6.8), 10% (w/v) sodium dodecyl sulphate (SDS), 0.1% (w/v) bromophenol blue and 10% (v/v) glycerol), to which β-mercaptoethanol (Merck) had been added to a final concentration of 2% (v/v), was added to each pellet. Samples were denatured at 95°C in a heating block, and loaded directly onto 15% (w/v) polyacrylamide gels. Gels were electrophoresed at 150 V for 1 h. Proteins were visualised by Coomassie stain as described below.

In order to express large amounts of recombinant protein in *E. coli*, a single colony of BL21(DE3)pLysS cells containing a specific pET20b(+)-MTE-GH5 expression vector was inoculated into 5 mℓ of LB growth medium supplemented with 50 μg/mℓ ampicillin and 30 μg/mℓ chloramphenicol and grown at 37°C to an OD_{600} of 0.7 after which the whole culture was added to 1ℓ fresh LB growth medium supplemented with the same complement of antibiotics. Cells were grown at 37°C to an OD_{600} of 0.6 (approximately 4.5 h) after which cultures were induced by the addition of IPTG to a final concentration of 0.4 mM and allowed to grow for an additional 2 h at 37°C. Cells were harvested by centrifugation (4 × 250 mℓ) in a JA14 rotor (Beckman) at 10 000 × g for 20 min. Cell pellets were dried by inverting centrifuge tubes to remove growth medium and frozen at -20°C overnight. Cell pellets were resuspended in 10 mℓ binding buffer (5 mM imidazole, 500 mM NaCl, 20 mM Tris·Cl, pH 7.9) each. At this point, lysis occurred, making complete resuspension extremely difficult. The 4 suspensions were transferred to 2 × 50

mℓ centrifuge tubes for sonication. This procedure was performed using the 0.4 cm tip (red) fitted on a Bronwill Biosonik III sonicator (Bronwill Scientific, Inc., New York) at an intensity setting of 30. Both samples were sonicated alternately on ice with 40 s pulses for 45 min, until samples were no longer viscous. This procedure caused complete lysis of cells, without detectable denaturation of proteins. Hereafter, the sonicated lysates were centrifuged in a JA20 rotor (Beckman) at 39 000 × g for 20 min. The supernatant was collected and pooled, after which the crude protein extracts were filtered through a 0.22 μm syringe filter. An aliquot (1 mℓ) of extract was stored at -20°C.

2.3.7. Protein purification

Recombinant proteins (MTE-GH5 and MTE-GH5 mutants) were purified by immobilised metal affinity chromatography (IMAC) using a His·Bind kit (Novagen). A 50% slurry of nickel agarose resin (2.5 mℓ) was packed into a plastic column (approximately 7 cm × 1 cm) supplied with the kit. The storage buffer was allowed to drop to the bed level after which 7.5 mℓ milli-Q H_2O was added. Hereafter, the column was washed with 12.5 mℓ charge buffer (50 mM $NiSO_4$) and 7.5 mℓ of binding buffer. Filtered, crude protein extract (approximately 40 mℓ) was applied to the column after which the column was washed with 25 mℓ of binding buffer, 15 mℓ of wash buffer (40 mM imidazole, 500 mM NaCl, 20mM Tris·Cl, pH 7.9). Aliquots (1 mℓ) of each wash fraction were stored at -20°C. Hereafter, 15 mℓ of elution buffer (1 M imidazole, 0.5 M NaCl, 20 mM Tris·Cl, pH 7.9) was added and 1 mℓ fractions were collected. Fractions were stored at -20°C and

10 μℓ aliquots of each fraction (crude extract, binding wash, wash and elution fractions) was loaded on a 15% (w/v) SDS polyacrylamide gel and electrophoresed at 150 V for 1 h. Proteins were visualised by Coomassie stain as described below.

2.3.8. In-gel trypsin digest and MALDI-TOF analysis

Following SDS-PAGE electrophoresis and Coomassie Brilliant Blue Staining, gels were destained in 7% (v/v) acetic acid, 25% (v/v) ethanol overnight. Protein bands to be analysed by in-gel trypsin digest were excised from the gel. The band corresponding to carboxypeptidase in the marker lane containing a Broad Range Molecular Marker (Bio-Rad Laboratories, Hercules, CA, USA) was excised as positive experimental control, while a blank gel piece was excised as negative control. Each slice was washed twice with 200 μℓ 50% (v/v) acetonitrile, 25 mM NH_4HCO_3. A 15 min incubation period was allowed before removing the solution. Hereafter, gel slices were washed once with 200 μℓ 100% (v/v) acetonitrile in the same way, after which slices were dried in a rotary evaporator for 20-30 min. Trypsin (20 μℓ of 20 μg/mℓ, Sigma-Aldrich) and 25 mM NH_4HCO_3 was added, and the sample was incubated overnight at 37°C to allow digestion. Acetonitrile and TFA (50% (v/v) and 5% (v/v), respectively) were added to a final volume of approximately 70 μℓ, and samples were allowed to stand for 30 min. The remaining solution was transferred to a new eppendorf tube and then dried to completion. Samples were dissolved in 60 μℓ 0.05% (v/v) TFA, 5% (v/v) acetonitrile and mixed in a 1:1 dilution with α-cyano-4-hydroxycinnamic acid (10 mg/mℓ in 60% (v/v) acetonitrile,

0.3% (v/v) TFA). A 2 μℓ aliquot of this solution was spotted on a golden grid. After air-drying, samples were analysed by MALDI-TOF. Typically, analyses were performed at an accelerating voltage of 20 000 V and a grid voltage of 74 %, while the negative ion selector was switched off.

2.3.9. SDS-PAGE gel electrophoresis of proteins

All SDS-PAGE analyses of protein samples were performed according to the discontinuous method of Laemmli (1970). Running gels consisted of a 15% (w/v) polyacrylamide matrix (acrylamide:bisacrylamide, 29:1) in 375 mM Tris·Cl and 0.1% (w/v) SDS, and were overlaid with a 6% (w/v) polyacrylamide stacking gel (acrylamide:bisacrylamide, 29:1) in 125 mM Tris·Cl and 0.1% SDS. Samples were denatured at 95°C in SDS sample application buffer (200 mM Tris·Cl (pH 6.8), 2% (w/v) SDS, 0.02% (w/v) bromophenol blue and 2% (v/v) glycerol), cooled to room temperature, and loaded onto the gel. Gels were electrophoresed at 150 V for 1h in a MiniProtean III (Bio-Rad) electrophoresis system. Following electrophoresis, gels were stained with Coomassie Blue stain solution for 1 h (0.25% (w/v) Coomassie Brilliant Blue in methanol), and destained overnight in destain solution (7% acetic acid, 25% ethanol). Ethanol was substituted with methanol in the destain solution where protein samples were to be analysed by MALDI-TOF.

CHAPTER 3: MODEL OF THE GH5 CHROMATOSOME

3.1. Introduction

In this chapter, a structural model of the binding of the globular domain of linker histone H5 to the nucleosome is proposed. The primary DNA-binding site of GH5, which is the site of attachment to linker DNA, was modelled by homology to the binding of *E. coli* CAP to DNA. The crystal structure of the nucleosome core (Luger *et al.,* 1997), in conjunction with the derived primary site model, was employed to model the secondary DNA-binding site. Specific emphasis is placed on the putative secondary DNA-binding site, as the nature of this site is still unclear (Travers, 1999; Thomas, 1999).

3.2. Model of the primary DNA-binding site of GH5

3.2.1. Template selection

In order to investigate structural models for the binding of the secondary DNA-binding site of the globular domain of chicken linker histone H5 in a nucleosome, the solved crystal structure of GH5 was used (Ramakrishnan, *et al.,* 1993). However, this crystal structure contains two slightly dissimilar GH5 monomers in the crystal unit cell, probably due to the strong dimerisation propensity of GH5 in solution at physiological pH and ionic conditions. As shown in Figure 3.1, the β-hairpin of the B-chain of the GH5 dimer is hinged to the helical body of the chain, while the β-hairpin of the A-chain is in a more

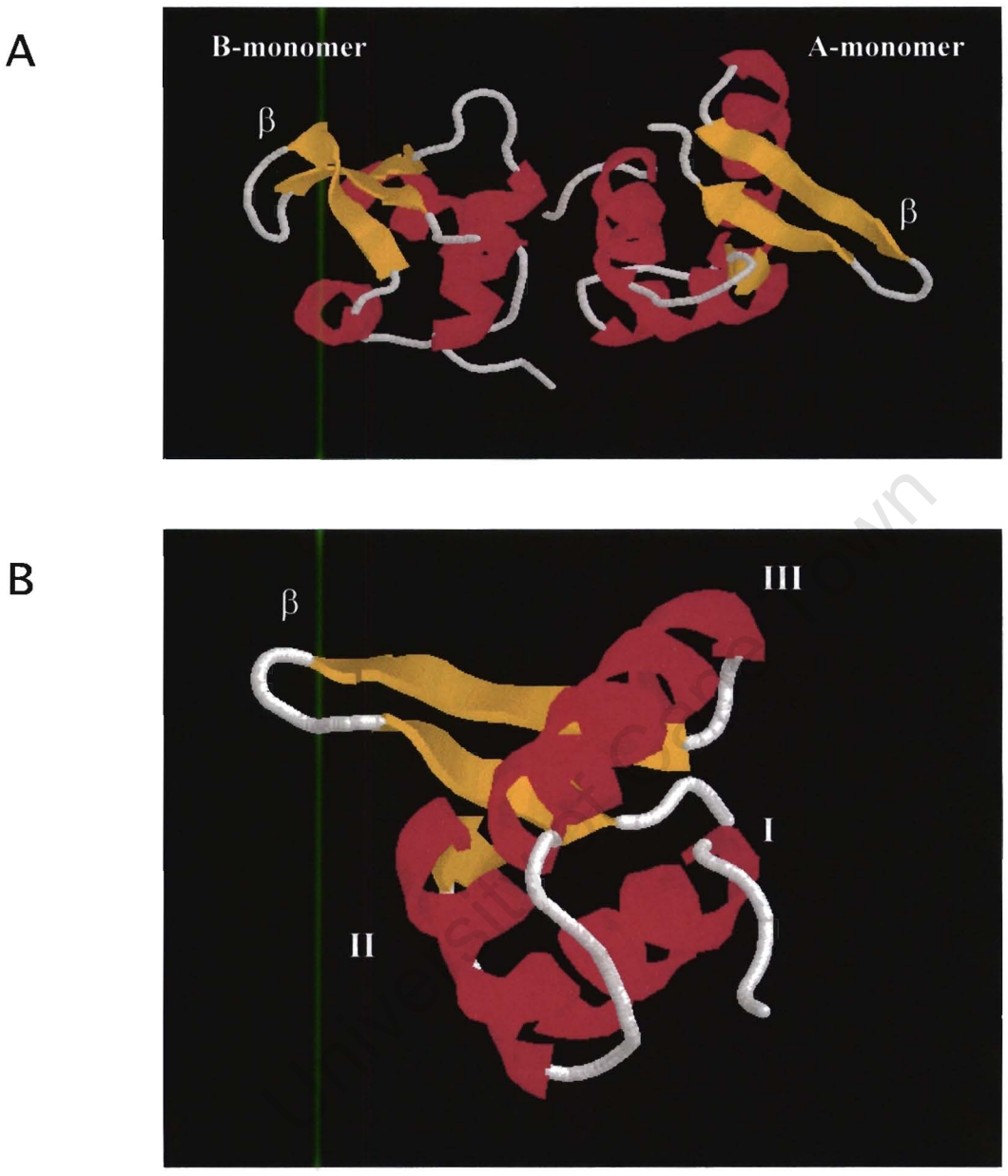


Figure 3.1. Ribbon representations of the crystal structure of the globular domain of H5. Alpha-helices are indicated in magenta and beta-sheets in yellow. The C-terminal beta-hairpin motif is annotated by β. **A.** The GH5 crystal dimer, showing the B-monomer on the left and the A-monomer on the right. The β-hairpin motif is extended in the A-monomer. **B.** Ribbon diagram of the A-monomer, indicating the positions of helices I, II and III.

extended conformation, demonstrating the flexibility of the β-hairpin. This hairpin contains two conserved residues, Lys 85 and Arg 95, which might be involved in linker DNA binding (Ramakrishnan *et al.,* 1993). Both these residues are situated at the base of the β-hairpin and therefore the effect of the conformation of the β-hairpin on their relative positions is likely to be minimal. The A-monomer has an electron density map that is better defined than the B-monomer (Ramakrishnan *et al.,* 1993), mostly due to the stabilization of its β-turn by inter-molecular crystal contacts. Similar inter-molecular contacts are expected between conserved residues located in the β-hairpin (Lys 85 and Arg 95) and linker DNA. In addition, the β-hairpin of CAP (Parkinson *et al.,* 1996) is in this conformation when associated with DNA. For these reasons, all modelling procedures were performed using the experimental structure of only the A-chain of GH5.

It is known that GH5 has the topology of a three-helix bundle and it is likely that helix III is the DNA recognition helix (Ramakrishnan *et al.,* 1993). Since it is known that helix III binds in the major groove of linker DNA (Zhou *et al.,* 1999), a model built by homology to a structural homologue of which the co-crystal structure in a complex with DNA has been solved, should reveal the nature of the primary DNA-binding site of GH5.

It can be assumed that GH5 would have the same qualitative DNA-binding features as the DNA-binding domain of CAP, a prokaryotic DNA-binding protein, since these two domains share a common fold topology (Ramakrishnan *et al.,* 1993). The Root Mean Square (RMS) difference between the structure of GH5 and the DNA-binding domain of CAP is 1.3 Å over equivalent C_α atoms and 1.7 Å over all atoms (Ramakrishnan *et al.,*

1993), suggesting that the co-crystal structure of the DNA-binding domain of CAP in complex with double-stranded DNA is a suitable modelling template. A similar value (1.3 Å over equivalent C_α atoms) has been reported for GH5 and the HNF-3/*fork head* DNA recognition motif, of which the co-crystal structure has been published (Clark *et al.*, 1993). It is known that HNF-3 and GH5 bind similarly to DNA in chromatin (Cirillo *et al.*, 1998). However, the experimental evidence of various workers (Ramakrishnan *et al.*, 1993; Thomas and Wilson, 1986; Mirzabekov *et al.*, 1990) suggests that GH5 recognises DNA in a manner consistent with the binding of helix III to the major groove. Since the same mode of DNA-binding is observed for CAP (Parkinson *et al.*, 1996), the binding of GH5 to duplex DNA was modelled by homology to the solved co-crystal structure of the DNA-binding domain of CAP in a complex with DNA.

3.2.2. Superposition of GH5 (chain A) and CAP

In order to model the primary DNA-binding site of GH5, which is the GH5-linker DNA interface, the crystal structure of the A-chain of the globular domain of GH5 (Ramakrishnan *et al.*, 1993) was superimposed onto the co-crystal structure of the DNA-binding domain of CAP (Parkinson *et al.*, 1996). The most important residues contacting the DNA-backbone in the CAP-DNA-complex were identified by analysing the ionic and hydrogen-bonded interactions involved, using the angle and distance cut-offs listed in Chapter 2. By comparing secondary structural elements of both proteins (GH5 and CAP), areas with similar secondary structures were identified, and an alignment revealed the equivalent residues in GH5 that are candidates to contact linker DNA (Figure 3.2).

A

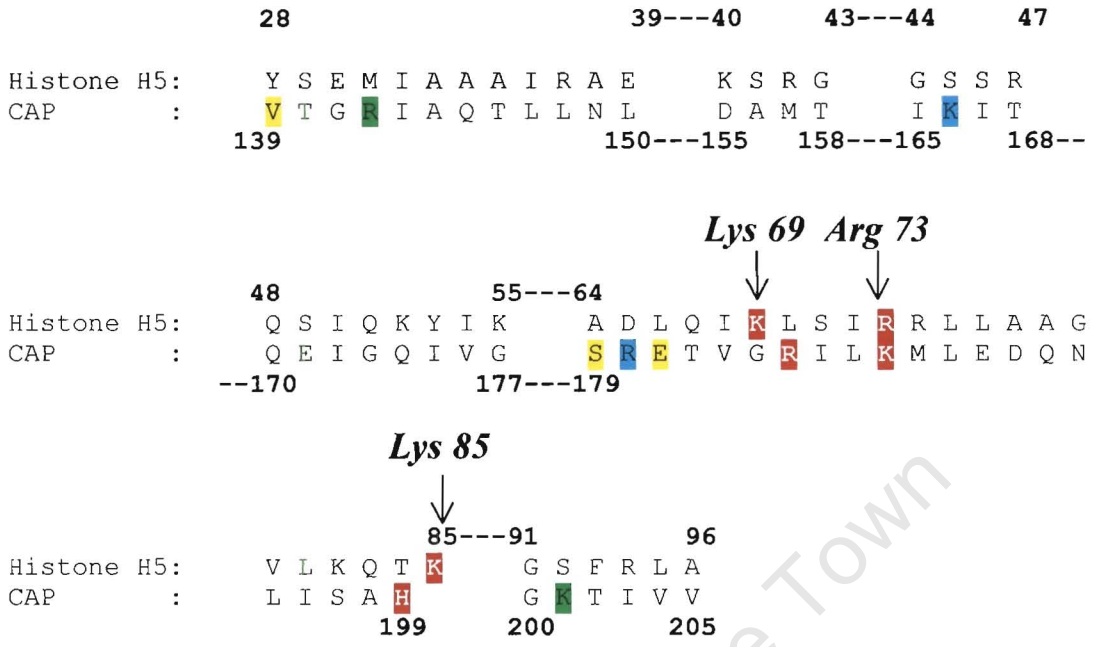


Figure 3.2. Functionally equivalent residues in GH5 and CAP A. DALI structural alignment of equivalenced areas in GH5 and CAP. The type of interaction of CAP residues with DNA is indicated by colour: **Ionic interaction/H-bond**, **H-bond**, **Ionic interaction**. Highly conserved basic residues (GH5) and their structural equivalents (CAP) are highlighted in red.

B

```
           1            13 14           28 29           43 44           58 59           73 74           88
H5 chicken -TESLVLSPAPAK-P KRVKASRRSASHPTY SEMIAAAIRAEKSRG GSSRQSIQKYIKSHY KVGHNADLQIKLSIR RLLAAGVLKQTKGVG
H5 duck    MTDSPIPAPAPAAKP KRAKAPRKPASHPSY SEMIVAAIRAEKSRG GSSRQSIQKYVKSHY KVGQHADLQIKLSIR RLLAAGVLKQTKGVG
H5 duck2   -TDSPIPAPAPAAKP KRARAPRKPASHPTY SEMIAAAIRADKSRG GSSRQSIQKYVKSHY KVGQHADLQIKLAIR RLLTTGVLKQTKGVG
H5 goose   -TDSPIPAPAPAAKP KRARAPRKPASHPTY SEMIAAAIRADKSRG GSSRQSIQKYVKSHY KVGQHADLQIKLAIR RLLTTGVLKQTKGVG
H5B Xenopus --MTENSA--PAAKP RRSKASKKSTDHPKY SDMILDAVQAKKSRS GSSRQSIQKYIKNNY TVGENADSQIKLSIK RLVTSGTLKQTKGVG
H10 mouse  --MTENSA--PAAKP RRSKASKKSTDHPKY SDMILDAVQAEKSRS GSSRQSIQKYIKNNY TVGENADSQIKLSIK RLVTSGTLKQTKGVG
H5A Xenopus --MAENSAATPAAKP KRSKALKKSTDHPKY SDMILAAVQAEKSRS GSSRQSIQKYIKNHY KVGENADSQIKLSIK RLVTSGALKQTKGVG
GH1 chicken ---------------- ---------MAGPSV TELITKAVSASKERK GLSLAALKKALAAGG YDVEKNNSRIKLGLK SLVSKGTLVQTKGTG

           89          103 104         115 116         129 130         142 143         153 154         168
H5 chicken ASGSFRLAKSDKAKR SPG---KKKKAVRRS TSPKKAARPRK-ARS PAK--KPKATARKAR KKSR----ASPKKAK KPKTVKAKSRKASKA
H5 duck    ASGSYRLAKGDKAKK SPAGRKKKKKAARRS TSPRKAARPRK-ARS PAK--KPKAAARKAR KKSR----ASPKKAK KPKTVKAKSLKTSKV
H5 duck2   ASGSFRLAKGDKAKR SPAGRKKKKKAARKS TSPKKAARPRK-ARS PAK--KPKAAARKAR KKSR----ASPKKAK KPKTVKAKSLKTSKP
H5_goose   ASGSFRLAKGDKAKR SPAGRKKKKKAARKS TSPKKAARPRK-ARS PAK--KPKAAARKAR KKSR----ASPKKAK KPKTVKAKSLKTSKP
H5B Xenopus ASGSFRLAKADEVKK -PAK--KPKKEIKKA VSPKKAAKPKKAAKS PAKAKKPKVAEKKVK KAPKKKPAPSPRKAK KTKTVRAKPVWASKA
H10 mouse  ASGSFRLAKADEVKK -PAK--KPKKEIKKA VSPKKAAKPKKAAKS PAKAKKPKVAEKKVK KAPKKKPAPSPRKAK KTKTVRAKPVWASKA
H5A Xenopus ASGSFRLAKADEGKK -PAK--KPKKEIKKA VSPKKVAKPKKAAKS PAKAKKPKVAEKKVK KVAKKKPAPSPKKAK KTKTVKAKPVRASKV
GH1 chicken ASGSFRLSK------ ---------------- ---------------- ---------------- ---------------- ----------------
```

C

```
                  1            15 16           30 31           45 46           60 61           75
H5 chicken        SASHPTYSEMIAAAI RAEKSKGGSSRQSIQ KYIKSHY----KGVH NADLQIKLSIRRLLA AGVLKQTKGV---GA
H1.1 human        P-AGPSVSELIVQAA SSSKERGGVSLAALK KALAAAG----YDVE KNNSRIKLGIKSLVS KGTLVQTKGT---GA
H1.0 human        STDHPKYSDMIVAAI QAEKNRAGSSRQSIQ KYIKSHY----KVGE NADSQIKLSIKRLVT TGVLKQTKGV---GA
H1 Drosophila     TPSHPPTQQMVDASI KNLKERGGSSLLAIK KYITATYK---CDAQ KLAPFIKKYLKSAVV NGKLIQTKGK---GA
H1.2 Drosophila   PPSHPPTQQMVDASI KNLKERGGSSLLAAI KYIGATYK---CDAQ KLAPFIKKYLKNAVA NGKLIQTKGK---GA
H1 Xenopus        PSSGPSVSEQIVTAV SASKERSGCSLAALK KTLAAGG----YDVD KNNSRLKLALKALVT KETLLQVKGS---GA
H1M Xenopus       TQSHPPTLSMVVEVL KKNTERKGTSVQAIR TRILSAHP--TVDPL RLKFLLRTALNKGLE KGILIRPLNSSATGA
H1 L. pictus      APAHPPTSQMVVAAI TALKERGGSSNQAIK KYIAANYK---VDIN KQATFIKRALKAGVA NGTLVQVKGK---GA
H1 S. purpuratus  PAAHPPAAEMVATAI TELKDRNGSSLQAIK KYIATNFD---VQMD RQLLFIKRALKSGVE KGKLVQTKGK---GA
H1 trout          --AGPSVGELIVKAV SASKERSGVSLAALK KSLAAGG----YDVE KNNSRVKIAVKSLVT KGTLVQTKGT---GA
H1 Pisum          PASHPTYEEMIKDAI VSLKERNGSSQYAIA KFIEEKQK-Q--LPA NFKKLLLQNLKKNVA SGKLIKVKGS-----
Hholp GDI Yeast   A-SSKSYRELIIEGL TALKERKGSSRPALK KFIKENYP-IVGSAS NFDLYFNNAIKKGVE AGDFEQPKGP-----
Hholp GDII Yeast  P-SSLTYKEMILKSM PQLNDGKGSSRIVLK KYVKDTFSSKLKTSS NFDYLFNSAIKKCVE NGELVQPKGP-----

                  76           90 91           109

H5 chicken        SGSFRLAKSD----- ----------KAK
H1.1 human        SGSFKLNKKAS---- ----------SV
H1.0 human        SGSFRLAKSD----- ----------EP
H1 Drosophila     SGSFKLSASAKKEKD PK---------AKSKVLS
H1.2 Drosophila   SGSFKLSRSAK--KD PKP--------KASAVEK
H1 Xenopus        SGSFKLNKKQL---- ----------QS
H1M Xenopus       TGRFKLAKPVKTTKA GKENVASENVDPNAEQET
H1 L. pictus      SGSFKLGKVKAG--- ----------KTEAQ
H1 S. purpuratus  SGSFKVNVQAAK--- ----------AQASE
H1 trout          SGSVKLNKK------ ----------AV
H1 Pisum          ---FLSAAAKK---P AV---------AKPK-AK
Hholp GDI Yeast   AGAVKLAKKKSP--- ----------EV
Hholp GDII Yeast  SGIIKLNKKK----- ----------V
```


Figure 3.2. (Continued). Sequence alignment of the globular domains of H5 linker histones and chicken linker histone H1 from different sources **(B)** and sequence alignment of linker histones from different organisms **(C)**.

This approach was followed to compensate for the poor sequence identity shared by these proteins (less than 10%), which makes pair-wise sequence alignments unreliable for detecting functional relationships (Sander and Schneider, 1991).

In this way, CAP residues Lys 188, His 199 and Arg 185, implicated in DNA-binding based on an analysis of the structure of the CAP-DNA complex, were found to be structurally equivalent to GH5 residues Lys 85, Lys 69 and Arg 73, respectively. Thomas and Wilson (1986) demonstrated the importance of Lys 85 and Lys 69 in DNA-binding, while Arg 73 is highly conserved and has been suggested to be important in DNA-binding by Ramakrishnan *et al.* (1993). These findings provide strong support for the choice of CAP as a structural template to investigate the binding of GH5 to DNA.

To investigate the association of the primary DNA-binding site of GH5 with DNA, the crystal structure of the A-monomer of GH5 was superimposed onto the co-crystal structure of the DNA-binding domain of CAP. Several different superpositioning approaches were employed.

A Root Mean Square (RMS) superposition, corresponding to the canonical "best structural fit", produced several serious steric improbabilities. This might be attributed to the fact that CAP contains an additional β-hairpin between helices D and E (corresponding to helix I and II of GH5) that is not present in the crystal structure of GH5.

To avoid the effect of this sheet on the overall superposition, Lys 73, Lys 69 (associated with helix III of GH5) and Lys 85 (associated with the C-terminal β-hairpin) were directly superimposed onto their structural counterparts in CAP (Arg 185, His 199 and Lys 188, respectively). Atoms corresponding to the location of the side-chain charge of structurally equivalent residues (as defined by the Arg *ipso* C-atom (CZ), the Lys ε-amino nitrogen and the C_γ atom of His) were superimposed to create a *charge centre* model. This manipulation resulted in severe steric clashes of putative functionally important residues of GH5 (Lys 73, Lys 69 and Lys 85) with the DNA backbone in the CAP-DNA complex. Less severe steric clashes were obtained by the superposition of C_α atoms of equivalent residues. The steric hindrance in both these models, are summarised in Table 3.1. The distances between the atoms involved in inter-atomic clashes are shown, as well as the allowable Van der Waals radii, which is the minimum distance that the atoms in question can approach each other in space (see Table 3.1).

The majority of atomic clashes in the C_α model are in the region of helix III, which is equivalent to the DNA-binding domain of CAP (Ramakrishnan *et al.,* 1993). These errors could easily be corrected by local optimisation in geometry. Contrary to this, the atomic clashes between GH5 and the DNA-backbone of the CAP-DNA complex in the *charge centre* model occur throughout the structure (Table 3.1), and would need more than local refinement to be correct. This suggests that the placement of GH5 onto DNA by superposition of C_α-atoms is more reliable than superimposing atoms corresponding to the location of charges. The C_α model was therefore used in subsequent modelling procedures.

Table 3.1. Steric hindrance in C_α and *charge centre* models. The residues in bold print are conserved in linker histones.

GH5 Residue	*Atoms involved*	*Distance (Å)*	*Allowable limit Van der Waals radii (Å)*	*Associated secondary structure*[1]
		C_α superposition		
Arg 47	N-O	2.18	2.70	Helix II
Asn 63	N-C	0.83	2.90	Helix III
Lys 69	N-O	2.09	2.70	Helix III
Gln 67	O-C	0.94	2.80	Helix III
Arg 73	N-O	1.24	2.70	Helix III
		Charge centre superposition		
Gln 48	C-C	1.74	3.20	Helix II
Lys 52	C-O	2.02	2.80	Helix II
Lys 69	C-C	2.82	3.20	Helix III
Arg 73	C-O	0.14	2.80	Helix III

1) Refer to Figure 3.1 for the location of the secondary structural elements and Figure 3.2 B for sequence alignments.

3.2.3. Removal of atomic clashes by geometry optimisation

In order to remove the inter-atomic clashes discussed above, the conformation of the proposed DNA-binding site of GH5, modelled by superposition of C_α atoms of equivalent residues (see above), was optimised by changing the conformation of the protein and evaluating the potential energy associated with that particular conformation. This was achieved by applying an iterative energy minimisation algorithm, the steepest descent algorithm (Mackay *et al.,* 1989), which is a crude tool to rapidly optimise structural conformation. The model of GH5 bound via helix III to double-stranded DNA that had been optimised accordingly, showed no significant signs of steric hindrance.

3.2.4. Structural analysis of the model of the primary DNA-binding site

All the potential interactions between the primary DNA-binding site of GH5 and DNA, as predicted by the model, are summarised in Tables 3.2 and 3.3. Some features are described below.

Six basic residues are proposed by this model to take part in DNA-binding (Arg 73, Arg 74, Lys 69, Lys 85, His 25 and His 62), have previously been implicated in linker DNA binding (see Figure 3.3 A – D). This DNA binding was proposed by the degree of conservation of some of these residues in the linker histones (Arg 73, Arg 74, Lys 69 and Lys 85; Ramakrishnan *et al.*, 1993), the ability of these residues to be covalently cross-linked to DNA in chromatin (His 25 and His 62; Mirzabekov *et al.*, 1990), or their

Table 3.2. Electrostatic interactions between GH5 residues and phosphate groups on the DNA backbone.

Residue	*Shortest distance to DNA backbone phosphate (Å)*	Comment
His 25†	2.55	Highly conserved in H5 proteins, but not well conserved in H1 proteins. Can be covalently cross-linked to DNA in chromatin.
Arg 47	2.53	Not conserved in all linker histones, but conserved in H5 family.
Lys 52	2.59	Highly conserved in linker histones. Protected from chemical modification in chromatin.
His 62†	2.56	Not conserved in linker histones. Can be covalently cross-linked to DNA in chromatin.
Lys 69*‡	2.58	Highly conserved in linker histones. Protected from chemical modification in chromatin.
Arg 73*	2.65	Highly conserved in H1 family.
Arg 74*	2.60	Charge conserved in all linker histones.
Lys 85*‡	2.61	Highly conserved in linker histones. Highly protected from chemical modification in chromatin.

* *Proposed to be part of the primary DNA-binding site by Ramakrishnan* et al. *(1993)*

‡ *Protected from chemical modification in chromatin as shown by Thomas and Wilson (1986)*

† *Can be cross-linked to DNA in chromatin (Mirzabekov* et al., *1989)*

Table 3.3. Hydrogen bonds between GH5 residues and phosphate groups on the DNA backbone.

Residue	*Heteroatom distance* (Å)	*Bond angle* (°)	*Comment*
Tyr 28	2.61	174	Hydrophobicity conserved in linker histones.
His 62†	2.59	155	Not conserved. Can be cross-linked to DNA in chromatin.
Asn 63	2.66	157	Not conserved.
Gln 67	2.68	173	Not conserved in linker histones, but conserved in H5 proteins.
Arg 73*	2.65 2.68 2.66	167 170 172	Charge highly conserved in linker histones.
Arg 74*	2.69	167	Charge not well conserved
Lys 85*‡	2.61 2.62	156 157	Highly conserved in linker histones. Highly protected from chemical modification in chromatin.

* *Proposed to be part of the primary DNA-binding site by Ramakrishnan* et al. *(1993)*

‡ *Protected from chemical modification in chromatin as shown by Thomas and Wilson (1986)*

† *Can be cross-linked to DNA in chromatin (Mirzabekov* et al., *1989)*

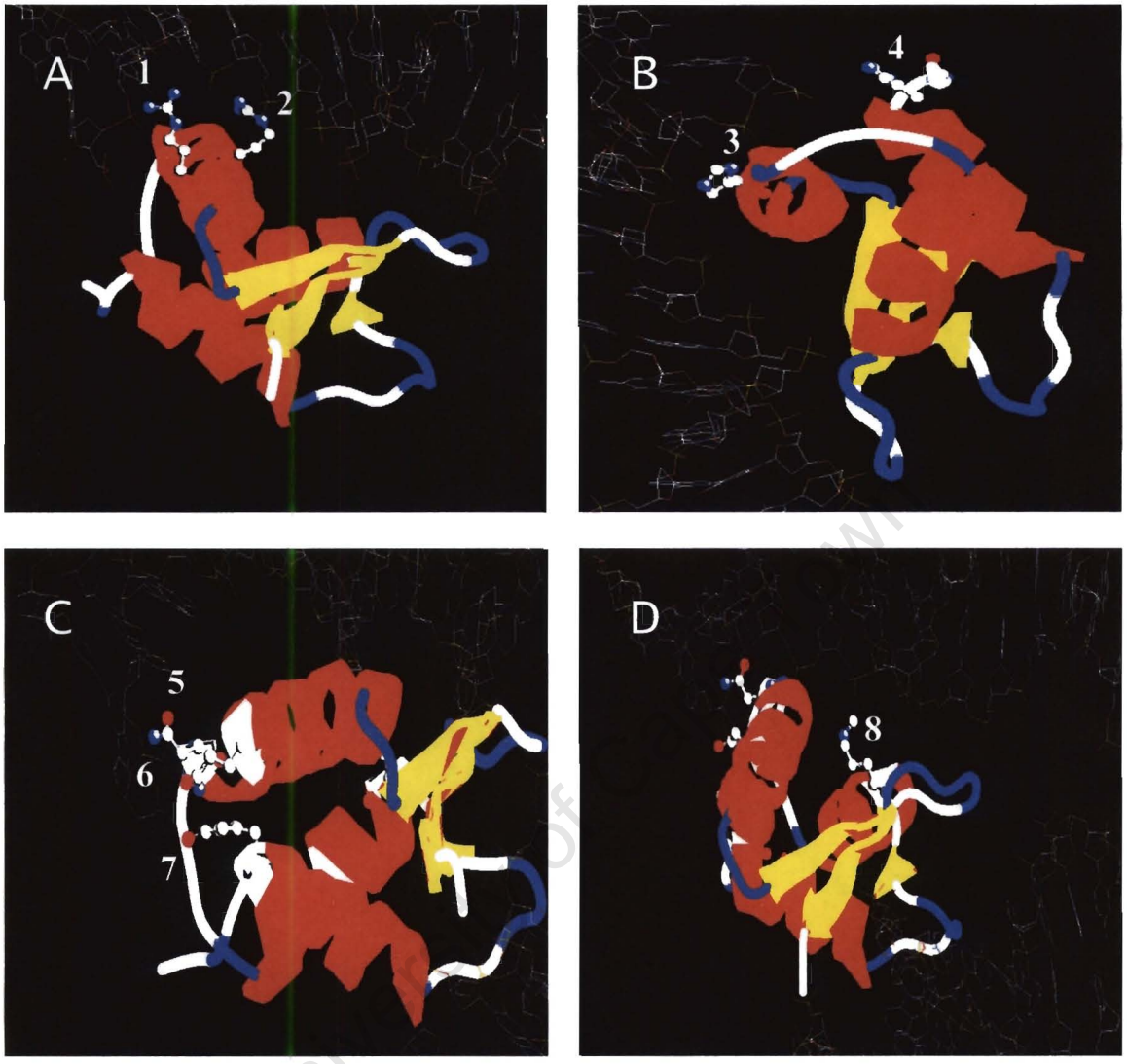


Figure 3.3. Relative position of residues of GH5 involved in linker DNA-binding. All images were generated using the program Rasmol. GH5 is represented as a ribbon, while the residues of interest are depicted in "ball-and-stick" format. Panels **A** and **B** show residues that had been previously implicated (see text), while panels **C** and **D** show newly implicated residues (this study). Arg 74 (**1**) and Arg 73 (**2**) are highly conserved among linker histones (Ramakrishnan *et al.*, 1993). His 62 (**3**) and His 25 (**4**) can be cross-linked to DNA in chromatin (Mirzabekov *et al.*, 1989). According to the model of the primary DNA-binding site proposed here, Asn 63 (**5**), Tyr 28 (**6**) and Gln 67 (**7**) form hydrogen bonds with the DNA backbone, while Arg 47 (**8**) is involved in an ionic interaction with the DNA backbone. Although these residues are not conserved among linker histones, their interactions with the DNA backbone offers an explanation for the higher affinity of GH5 for four-way junction DNA than that of GH1 (Varga-Weiss *et al.*, 1994).

protection from chemical modification when bound to DNA in chromatin (Lys 69 and Lys 85; Thomas and Wilson, 1986). The importance of all these residues can be fully explained in terms of their interaction with the linker DNA via hydrogen bonds, ionic interactions, or both (see Tables 3.2 and 3.3 and Chapter 6).

Although Lys 85 donates two strong hydrogen bonds to the DNA backbone (Table 3.3), the model predicts a moderate solvent accessibility of the side-chain surface, contrary to Lys 69, which is buried in a polar environment (Figure 3.4). However, it is conceivable that the protection of Lys 85 from chemical modification in chromatin might be due to its intimate association with DNA (see Chapter 6).

Arg 47 (Figure 3.3 D), not previously probed to investigate its role in DNA binding due to a low degree of conservation in linker histones, strongly interacts with the DNA-backbone in this model, with a contact distance of 2.53 Å (Table 3.2). This residue lies on the transition from the β-hairpin into helix II, suggesting that an arginine at this position is not important in maintaining the topology of the globular domain. The arginine clamp observed in this area, together with hydrogen bonds contributed by Gln 67 and Asn 63 (residues that are specific to H5, Table 3.3; Figure 3.3 C), might account for the higher binding affinity of GH5 to DNA compared to that of GH1. Indeed, it was proposed that the tight binding of H5 to chromatin in terminally differentiated avian erythrocytes are required for proper transcriptional silencing of select gene loci.

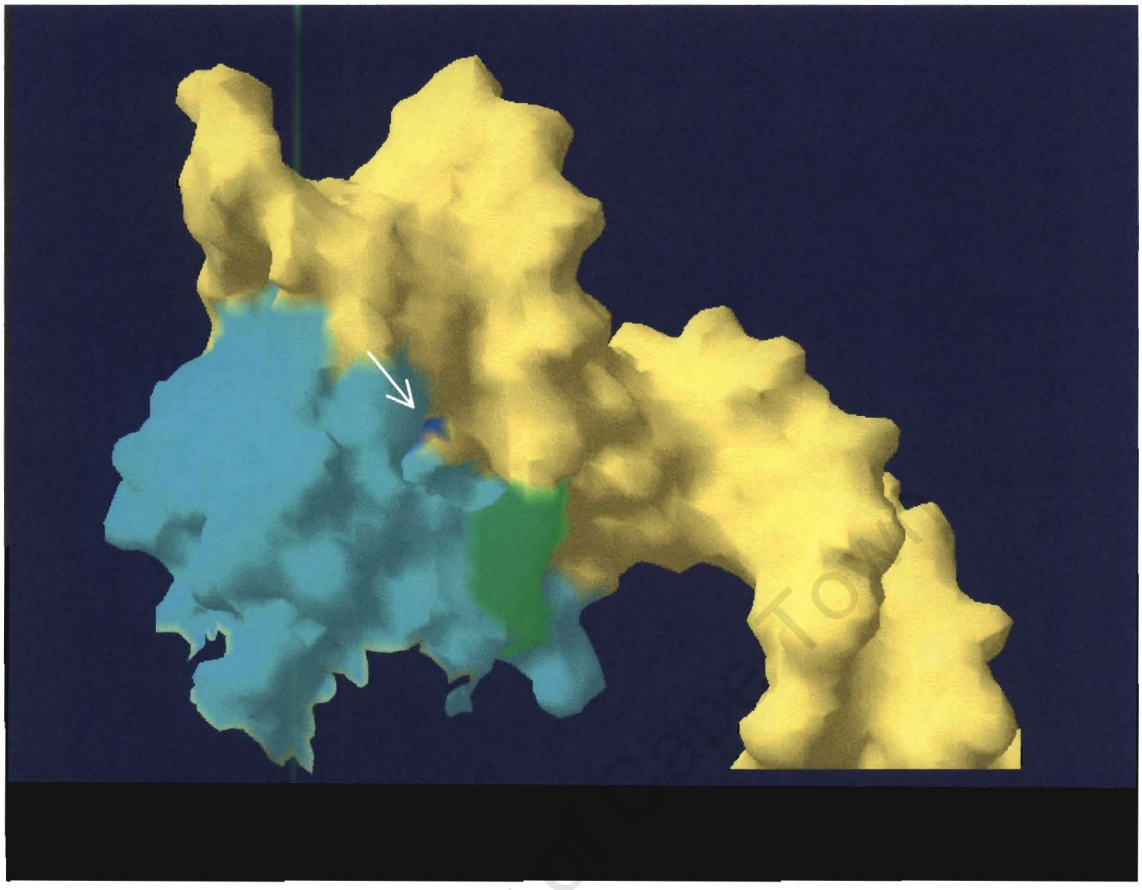


Figure 3.4. Surface model of the primary DNA-binding site of GH5. The surface of GH5 is shown in blue, while the linker DNA is indicated in yellow. The side-chain surface of Lys 85, which is protected from chemical modification in chromatin (Thomas and Wilson, 1986), is indicated in green. The white arrow indicates a dark blue spot, correlating with the surface of Lys 62 that is accessible to the solvent.

Tyr 28 (Figure 3.3 C) forms a strong, near-linear (174°) hydrogen bond to a phosphate group on the DNA-backbone (Table. 3.3). The hydrophobic character at this position is well conserved and the structurally equivalent residue in CAP, Val 139, also contacts the DNA via a strong hydrogen bond.

Ramakrishnan *et al.* (1993) and Goytisolo *et al.* (1996) have assigned Lys 52 to the secondary DNA-binding site of GH5. However, the assignment of Lys 52 to the primary DNA-binding site (Table 3.3) is entirely consistent with the experimental results of Goytisolo *et al.* (1996; see discussion, Chapter 6).

3.3. Elucidation of the secondary DNA-binding site of GH5 to nucleosomal DNA

3.3.1. Conformation of linker DNA at the entry/exit of the nucleosomes

Although the path of DNA in the core particle is well established (Luger *et al.*, 1997), the conformation of DNA at the entry/exit points of the nucleosome is unknown (see Chapter 1). Cryoelectron microscopy revealed that the arms of nucleosomes reconstituted on a 256 bp DNA fragment bend away from the nucleosome core, thereby preventing the occurrence of DNA crossing in the entry/exit region (Furrer *et al.*, 1995). Since the crystal structure of the nucleosome core particle solved by Luger *et al.* (1997) contains only 146 bp of DNA, and it is known that GH5 protects an additional 20 bp of linker DNA (Allan *et al.*, 1980), the modelling of GH5 in complex with the nucleosome requires an additional source of linker DNA. The DNA in association with GH5 in the

primary DNA-binding site model exhibited similar curvature to that observed by Furrer *et al.* (1995) after geometry optimisation. This observation, together with the fact that the DNA in association with the primary DNA-binding site of GH5 could be safely truncated without any loss of contacts, suggested that this DNA could serve as suitable source of modelled linker DNA. However, in order to test all possible conformations of linker DNA, probable nucleosomal locations for the secondary DNA-binding site of GH5 were investigated geometrically.

3.3.2. Probable nucleosomal locations of GH5 considering extreme conformations of linker DNA

Two extreme situations were considered: a *trajectory model*, where linker DNA exits the nucleosome in a straight line, and a *circular model*, where the inherent curvature of nucleosomal DNA is perpetuated to complete two negative super-helical turns (Figure 3.6).

Nucleosomal DNA was represented as a superhelical cylinder with a pitch (p) of 28 Å, and diameter (Ø) of 22 Å. The diameter (Ø) of the superhelix is 87 Å (see Figure 3.5). These dimensions correlate with the experimentally determined values (Luger *et al.*, 1997). GH5 is represented as a disk with radius equal to the distance between the primary and putative secondary DNA-binding site, centred over the 28 Å-cylinder. A definition of this distance is a complex problem, due to the dispersed nature of the

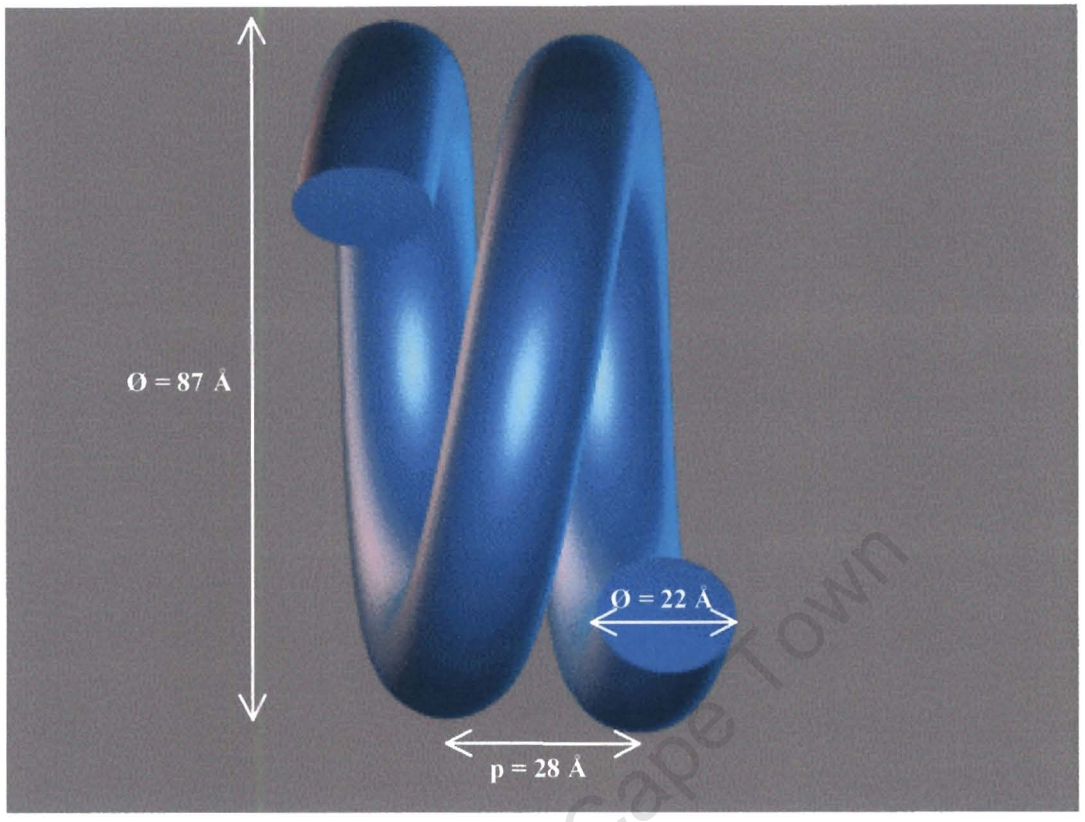


Figure 3.5. Geometry of the nucleosomal core DNA. Nucleosomal DNA is represented as a superhelical cylinder of which the dimensions correspond to the experimentally determined values (Luger *et al.*, 1997). These include a pitch (p) of 28 Å, and diameter (Ø) of 22 Å. The diameter (Ø) of the superhelix is 87 Å.

secondary DNA-binding site in particular. A maximum distance of 30 Å has been proposed in literature (Ramakrishnan *et al.,* 1993), assuming a linear angle between these two sites and the centre of the protein. However, this distance was calculated to be 19 Å by directly measuring the distance between a theoretical point located at the centre of mass of helix III (the primary DNA-binding site) and a theoretical point in the centre of the putative secondary DNA-binding site using the program Hyperchem. The surface of the disk represents a cross-section of the space that GH5 could occupy, if it could rotate freely around the DNA axis. The diameter of the disk represents the distance between the centre of the secondary DNA-binding site, and the centre of mass of helix III (19 Å). By translating the disc perpendicular to the helix axis, all possible contact surfaces between GH5 and nucleosomal DNA could be identified, where a contact surface is defined as the intersection of the disk with the DNA.

The potential locations of the secondary DNA-binding site of GH5, given the dimensions of nucleosomal DNA and the distance between the GH5 binding sites, are shown in Figure 3.6. As can be seen from this Figure, assuming a distance of 30 Å (Ramakrishnan *et al.,* 1993) between the primary and secondary DNA-binding sites, the binding of GH5 to a nucleosome resembling the circular model would require an angle smaller than 180° between the primary and secondary DNA-binding sites, since the constant distance between adjacent DNA gyres (28 Å) would preclude an angle of 180° on steric grounds. Furthermore, if the distance was 19 Å, GH5 would be able to bind anywhere on a circular model, assuming that GH5 binds without sequence-specificity. This is highly unlikely,

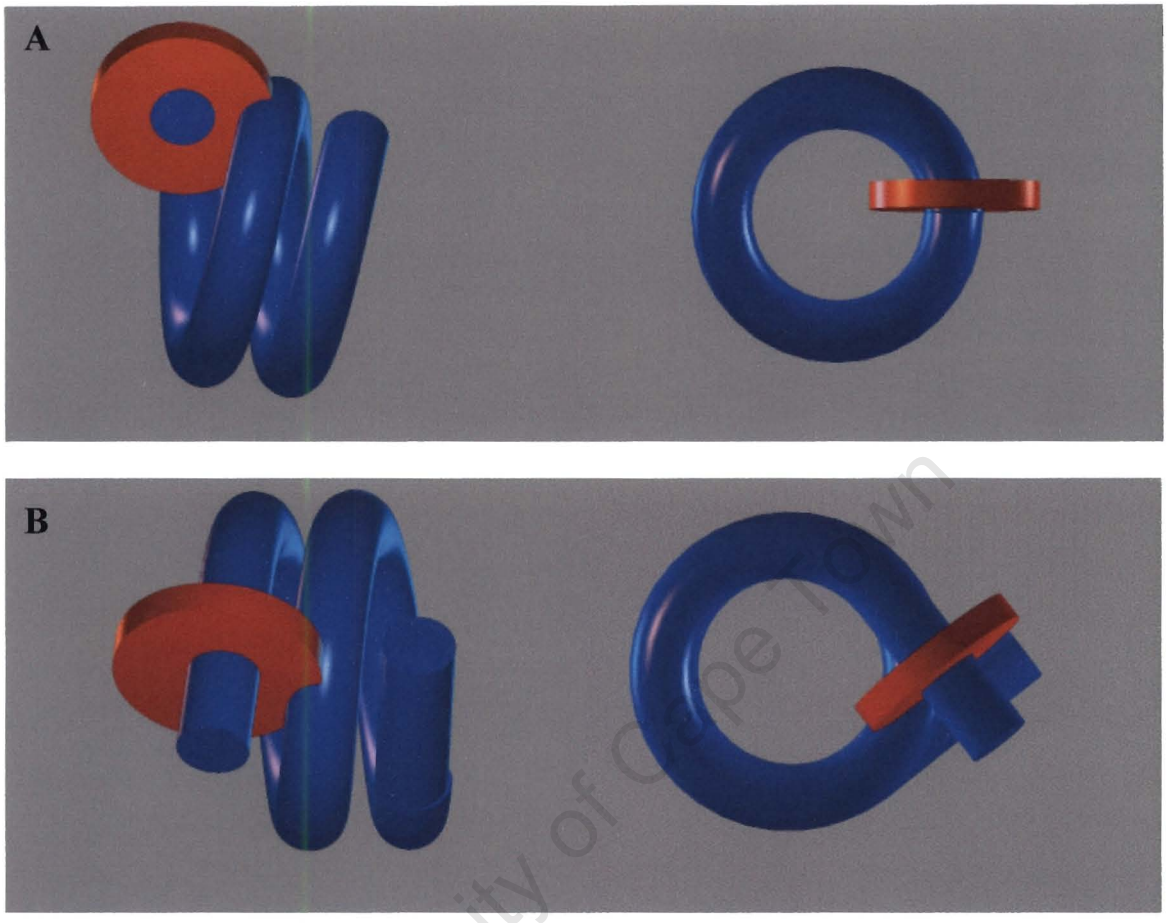


Figure 3.6. Geometrical analysis of the possible locations of the secondary DNA-binding site of GH5 in the nucleosome. Nucleosomal DNA is represented as a blue superhelical cylinder with dimensions as indicated in Figure 3.5. The surface of the red disk represents a cross-section of the space that GH5 could occupy, if it could rotate freely around the DNA axis. The diameter of the disk is 19 Å and represents the calculated distance between the centre of the putative secondary DNA-binding site, and the centre of mass of helix III, the primary DNA-binding site. **A.** A *circular model* where linker DNA at the entry/exit points of the nucleosome core follows the same curvature as nucleosomal DNA to complete 2 full superhelical turns. **B.** A *trajectory* model, where linker DNA at the entry/exit points of the nucleosome cross orthagonally.

since GH5 assumes a single preferred binding site in bulk chromatin (Zhou *et al.,* 1998). In contrast, the trajectory model clearly shows (see Figure 3.6) that at positions close to the entry/exit points of the linker DNA, distances between successive DNA gyres are compatible with simultaneous binding of the primary and secondary DNA-binding sites to the nucleosomal DNA. Therefore, the conformation of linker DNA as seen in the trajectory model was used to model the location of GH5 to the nucleosome.

As noted above, the core 13 bp of DNA associated with GH5 in the model of the primary DNA-binding site exhibits a curvature similar to that described by Furrer *et al.* (1995) for linker DNA. In addition, when this 13 bp core DNA was positioned at the entry/exit region of the structure of the core particle, it resembled the linker DNA in the trajectory model. The linker DNA-GH5 construct was attached to a nucleosomal core particle terminus, by defining phospho-diester bonds between the terminal base-pair of core particle DNA and the 13 bp DNA construct, in a manner that placed the secondary DNA-binding site in the vicinity of the octamer dyad as suggested by the data of Zhou *et al.* (1998). In this orientation, the β-hairpin of GH5 extends into the nucleosome. It should be noted that it is possible in principle to construct a model where the β-hairpin of GH5 extends away from the nucleosome. These two possibilities (see Figure 3.7) are discussed separately below.

3.3.3. Model where the β-hairpin of GH5 extends into the nucleosome

3.3.3.1. Model refinement

In order to optimise the bond lengths and angles between the nucleosomal DNA and the 13 bp fragment that was defined as "linker DNA", the geometry of the *in silico* ligation site was optimised *in vacuo* by applying a crude iterative steepest descent algorithm (Mackay *et al.,* 1989) with the program Hyperchem. This method is useful where large regions of conformational space are expected to be traversed to optimise the geometry of a structure. The manipulation had the effect of smoothening the DNA-backbone at the transition from nucleosomal DNA to linker DNA, without disturbing the contacts between GH5 and linker DNA.

A second round of energy minimisation was then performed on the linker DNA, five base-pairs of nucleosomal DNA in the vicinity of the nucleosomal dyad and GH5, by applying the conjugate gradient method of Polak and Ribiére (1969). The method of Polak and Ribiére is an iterative energy minimisation algorithm, which is more sensitive than the steepest descent algorithm (Mackay *et al.,* 1989). These regions were selected to optimise both the structure of the 13 bp linker DNA, and also the molecular contacts between the putative secondary DNA-binding site of GH5 and the central gyre of nucleosomal DNA. The protein-DNA contacts observed after these energy minimisation procedures are summarised in Tables 3.4 and 3.5.

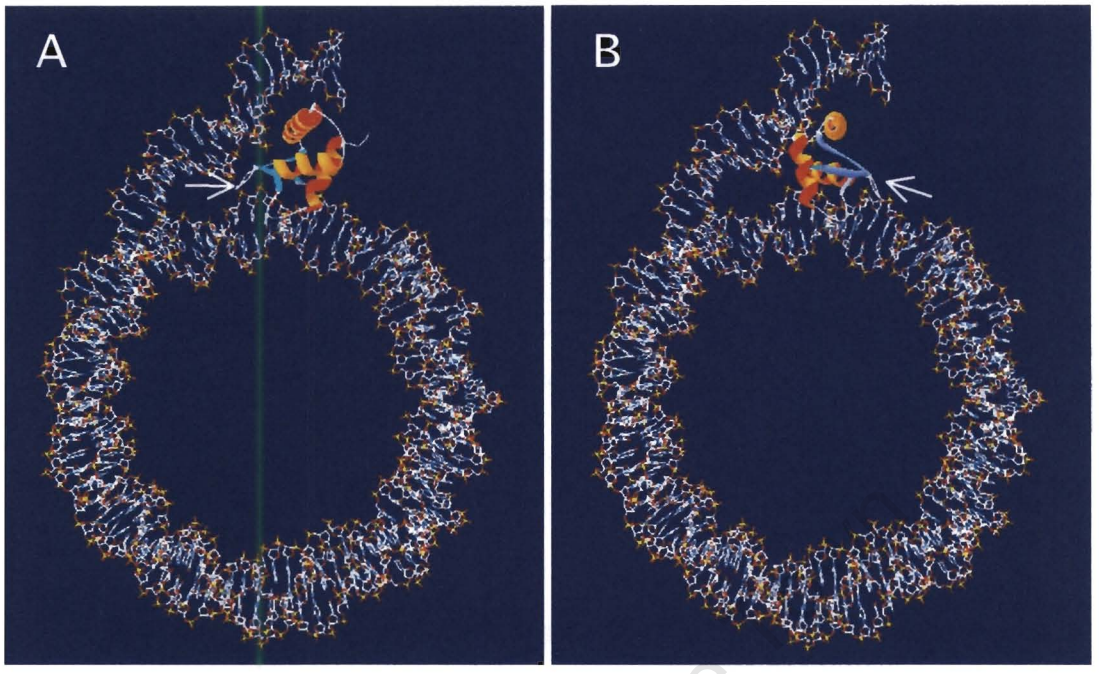


Figure 3.7. Models for the orientation of GH5 on the nucleosome. Nucleosomal DNA is represented in wire-frame format. GH5 is shown as a ribbon. A white arrow indicates the position of the β-hairpin of GH5. Panel **A** shows the orientation where the β-hairpin of GH5 extends into the nucleosome, while panel **B** shows the orientation where the β-hairpin of GH5 extends away from the nucleosome. Images were generated with the PDBViewer.

3.3.3.2. Analysis of the putative secondary DNA-binding site

The secondary DNA-binding site proposed here cannot be assigned to a specific secondary structural region. Instead, it consists of a contact surface generated by the C-terminal end of helix I, the inter-helical segment connecting helix I and II, and distinct contact points in both strands of the C-terminal β-sheet (Figure 3.8). These regions co-localize over the central gyre of nucleosomal DNA and straddles DNA in the vicinity of the dyad axis. The highly conserved basic residues, Arg 42 (inter-helical segment), Lys 40 (helix I) and Arg 94 (C-terminal β-strand), are in close contact with nucleosomal DNA near the nucleosomal dyad (see Figure 3.9). In addition, Gly 44 (inter-helical segment), which is also highly conserved among linker histones, donates a strong (2.68 Å) hydrogen bond to the DNA-backbone. Two less conserved residues belonging to the aforementioned GH5 segments (Arg 37 and Gly 43), also contribute to GH5 binding, mainly via hydrogen-bonded interactions (Table 3.6).

3.4. Models where the orientation of GH5 with respect to the nucleosome is reversed

In order to avoid an assumption regarding the two possible orientations of GH5, two models of the reverse orientation, i.e. where the β-hairpin of GH5 extends away from the nucleosome, were created using the optimised model described above as starting point.

Table 3.4. Electrostatic interactions between GH5 secondary DNA-binding site residues and phosphate groups on the DNA backbone.

Residue	*Shortest distance to DNA backbone phosphate* (Å)	*Comment*
Arg 37	2.26	Not conserved.
Arg 42	2.61	Highly conserved in linker histones.
Lys 40	2.54	Highly conserved in linker histones.
Arg 94	2.60	Highly conserved in linker histones.

Table 3.5. Hydrogen bonds between secondary DNA-binding site GH5 residues and nucleosomal DNA. All H-bonds are to the DNA backbone, unless specified otherwise.

Residue	*Heteroatom distance* (Å)	Bond angle (°)	Comment
Arg 37	2.61	172	Not conserved.
Gly 43	2.65	161	Not conserved. H-bond to the same backbone phosphate as Gly 44.
Gly 44	2.68	166	Highly conserved in linker histones.
Thr 84	2.56	159	Fairly well conserved in linker histones.
Lys 97	2.73	163	Fairly well conserved in linker histones. H-bond to cytosine carbonyl oxygen.

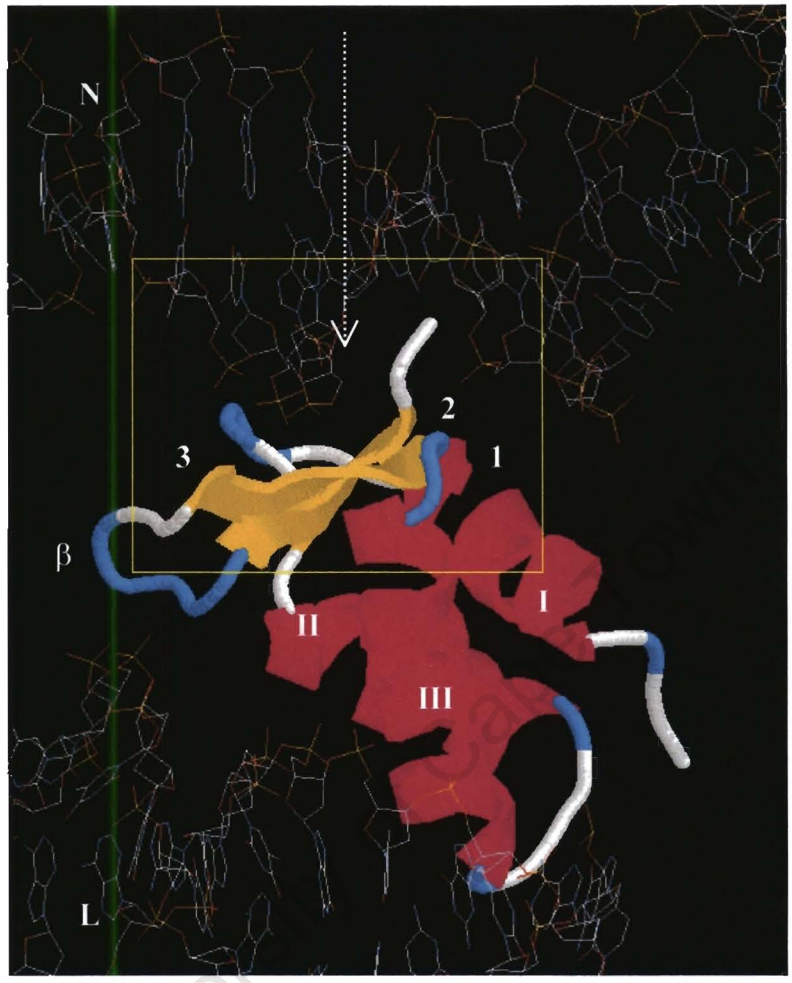


Figure 3.8. Location of the secondary DNA-binding site of GH5 on nucleosomal DNA. GH5 is presented as a ribbon, with helices numbered **I** to **III** from the N- to the C-terminus. The C-terminal β-hairpin is labelled β. DNA associated with the central gyre of nucleosomal DNA is labelled **N**, while linker DNA is labelled **L**. The area of GH5 corresponding to the secondary DNA-binding site is boxed. Specifically, this area consists of, **1.** The C-terminus of helix I; **2.** The inter-helical loop connecting helices I and II; and **3.** The C-terminal β-sheet. The white arrow indicates the nucleosomal dyad axis.

The *rotated* model was constructed by rotating GH5 in the major groove of the modelled linker DNA around the centre of mass of helix III (using the program Hyperchem). The *cut-and-paste* model was constructed by removing the linker DNA major groove associated with helix III and inserting GH5 bound to this DNA back in the reverse orientation at the same position.

The orientational setting in these two models is consistent with the cross-linking data of Mirzabekov *et al.* (1990). According to this study, His 25 interacts with the terminal regions of nucleosomal DNA, whereas His 62 is associated with more distal segments of DNA. When His 25 and His 62 are aligned in this order, the β-hairpin of GH5 extends away from the nucleosome. Both models were optimised by applying the iterative energy minimisation algorithm of Polak and Ribiére (1969) *in vacuo.* An analysis of these models is summarised in Tables 3.6 to 3.7. Note that in the case of the *cut-and-paste* model, the primary site contacts are identical to those reported for the model discussed in section 3.2.3., above.

The secondary DNA-binding site in the *cut-and-paste* model is characterised by two weak ionic interactions between Lys 52 and Lys 97 with the nucleosomal DNA-backbone and a host of hydrogen-bonded interactions (see Table 3.7). However, highly conserved residues like Arg 94 and Gly 44 are not involved in DNA-binding. The fairly well conserved Thr 84 is also not involved in DNA binding, and it is unlikely that this residue is required for the tertiary structure of the protein, since no ionic or hydrogen bond

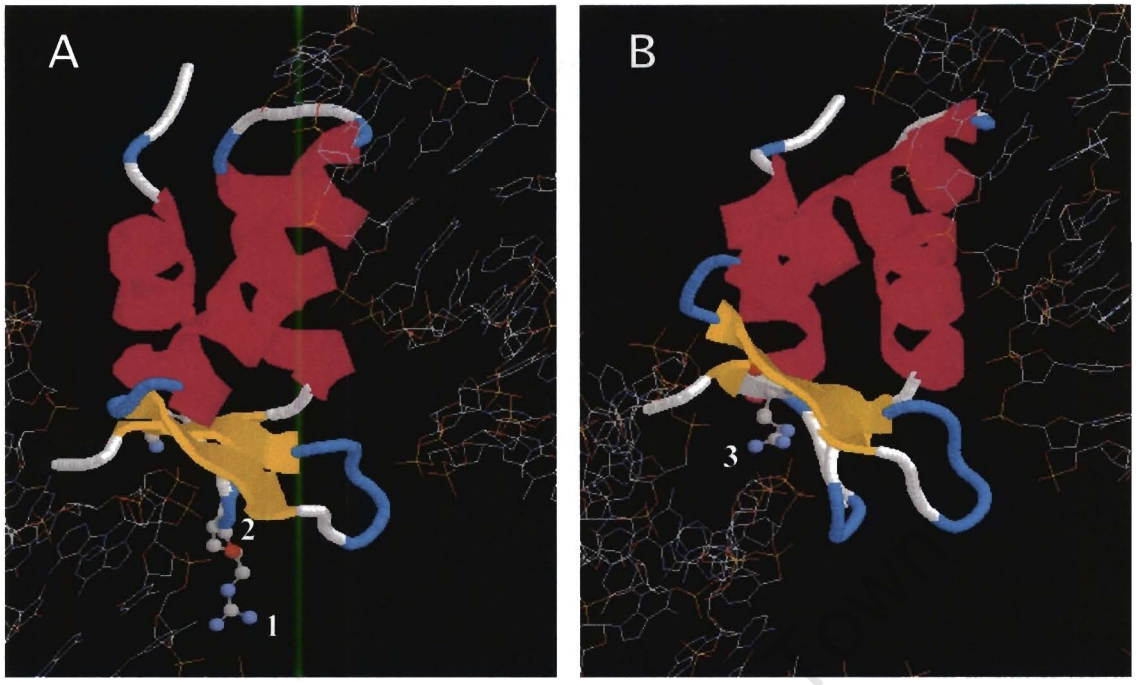


Figure 3.9. Conserved basic residues of GH5 predicted to be involved in nucleosomal DNA-binding via the secondary DNA-binding site. GH5 is represented as a ribbon, while nucleosomal DNA is represented in wire-frame format. Conserved basic residues are represented in "ball-and-stick" format. Residues shown in panel **A** belong to the inter-helical segment connecting helix I and II, whereas the residue in panel B is situated in the C-terminal β-hairpin motif. **1.** Arg 42; **2.** Lys 40 and **3.** Arg 94.

interactions to distal amino acid residues are evident in the crystal structure (Ramakrishnan *et al.,* 1993). The exposure of the hydrophobic Tyr 58 to the solvent is also problematic.

The secondary DNA-binding site of the *rotated* model has a more ionic character, although most of these interactions are quite weak. In this model, GH5 is stabilized by additional inter-residue H-bonds between Lys 59 and Lys 69.

Although these models are consistent with the orientation of GH5 as determined by Mirzabekov *et al.* (1990), they do not offer sufficient explanation for the majority of the body of evidence already in place regarding the nature of the binding of GH5 to the nucleosome. The model that is most consistent with both the chemical protection data (Thomas and Wilson, 1986) and where the highly conserved residues are involved in protein-DNA interactions (see Tables 3.4 and 3.5), contains helix III in a configuration where contacts are made in the major groove. These contacts are similar to that seen in the CAP co-crystal structure (Parkinson *et al.,* 1996). Also, the C-terminal β-hairpin is oriented towards the linker DNA in this model (see Figure 3.7).

Table 3.6. Analysis of interactions between GH5 residues and DNA as predicted by the *rotated model*.

1. Primary DNA-binding site		
Residues interacting via electrostatic interactions	***Contact distance* (Å)**	
Lys 82	2.58	
Residues interacting via hydrogen-bonded interactions	***Hetero-atom distance* (Å)**	***Bond angle* (°)**
Arg 73*	2.61 3.78 2.64	158.3 159.8 169.6
Arg 74*	2.70	167.5
Gln 83	2.81 2.71 2.73	151.5 169.1 166.8
Thr 84	2.59	158.2
Lys 85*‡	2.72	172.4
Gly 86	2.68	153.2
Arg 94	2.65 2.65	160.9 165.0
2. Secondary DNA-binding site		
Residues interacting via electrostatic interactions	***Contact distance* (Å)**	
His 25†	2.64	
Residues interacting via hydrogen-bonded interactions	***Hetero-atom distance* (Å)**	***Bond angle* (°)**
Ser 24	2.73	171.7
Tyr 58	2.57	159.4

* *Proposed to be part of the primary DNA-binding site by Ramakrishnan* et al. *(1993)*

‡ *Protected from chemical modification in chromatin as shown by Thomas and Wilson (1986)*

† *Can be cross-linked to DNA in chromatin (Mirzabekov* et al.., *1989)*

Table 3.7. Analysis of interactions between GH5 residues and nucleosomal DNA as predicted by the *cut-and-paste* model.

Secondary DNA-binding site[a]		
Residues interacting via electrostatic interactions	***Contact distance* (Å)**	
Lys 52	3.1	
Residues interacting via hydrogen-bonded interactions	***Hetero-atom distance* (Å)**	***Bond angle* (°)**
Ser 18	2.6	170
Arg 37	2.6 2.6 2.7	157 159 155
Lys 40	2.7	169
Arg 42	2.6	159
Tyr 53	2.98	155
Arg 73*	2.6 2.7 2.6	166 168 160

a Refer to Tables 3.2 and 3.3 for the primary DNA-binding site contacts of this model

** Proposed to be part of the primary DNA-binding site by Ramakrishnan* et al. *(1993)*

3.5. Conclusion

The model of GH5 bound to the nucleosome where the β-hairpin of GH5 extends into the core particle, offers the best explanation for our current understanding of the structure of the chromatosome. This model is consistent with most of the experimental evidence offered by Mirzabekov *et al.* (1990), Zhou *et al.* (1998) and Thomas and Wilson (1986). For this reason, this proposed model for the binding of GH5 to the nucleosome was used in the design of experiments to probe the nature of the secondary DNA-binding site.

CHAPTER 4: EXPRESSION AND PURIFICATION OF GH5 PEPTIDES

4.1. Introduction

In the previous chapter, a model for the binding of the secondary DNA-binding site of GH5 to nucleosomal DNA was proposed. In this chapter, specific residues that are suitable for substitution with cysteine to enable testing of the proposed model and mapping of the position of the secondary DNA-binding site on the nucleosome by site-directed metal-affinity cleavage are identified. In addition, the design, over-expression and purification of GH5 peptides with suitably substituted cysteine residues are described.

4.2. Overview of site-directed metal-affinity cleavage

Site-directed metal-affinity cleavage offers a high-resolution method for mapping protein-DNA interactions. According to this technique, an EDTA-metal derivative is bound via a disulphide linkage to strategically placed engineered cysteine residues on the protein surface. The metal moiety constitutes a reactive centre from where a Fenton hydroxyl radical-generating reaction cycle can be catalysed through the addition of hydrogen peroxide (see Figure 4.1; Tullius, 1988). In turn, released hydroxyl radicals lead to DNA strand scission by attacking C-1' and C-4' of the deoxyribose ring, creating chain breakage (Hertzberg and Dervan, 1984).

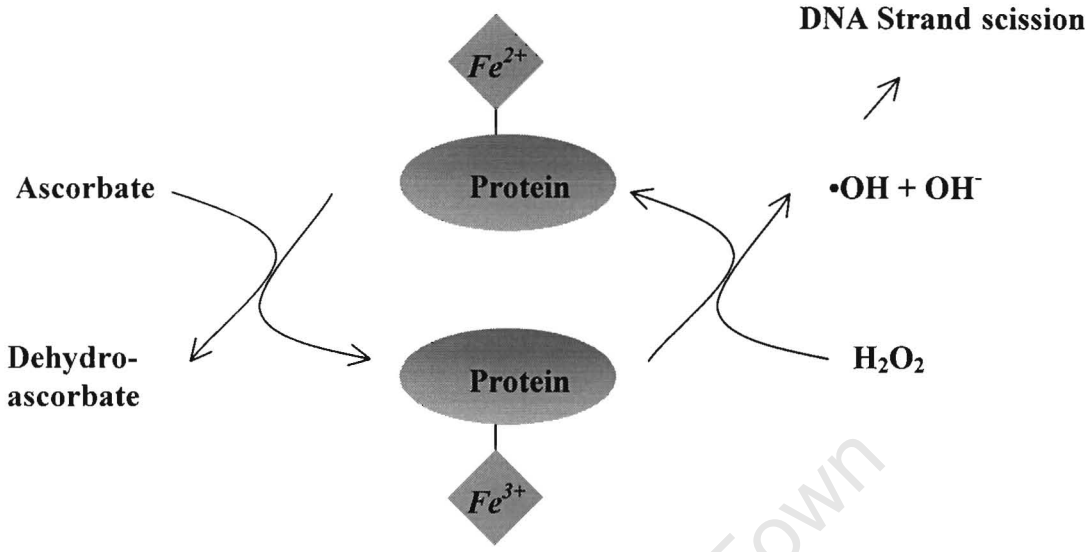


Figure 4.1. Schematic representation of the Fenton reaction cycle used in site-directed metal-affinity cleavage. Hydrogen peroxide reduces Fe^{3+} chelated by the EDTA moiety which is covalently attached to the protein via a cysteine residue. Hydroxyl radicals are released and result in DNA strand scission. The resulting Fe^{2+} is oxidised by ascorbate, allowing the cyclic generation of hydroxyl radicals (Tullius, 1988).

Strand scission at C-1' and C-4' lead to 3'-phospate and 3'-phosphoglyconate ends, respectively, with concomitant loss of the nucleotide base (Hertzberg and Dervan, 1984). By selectively end-labelling one DNA-strand, it is possible to map the protein-EDTA-metal induced cleavage sites at base-pair resolution and to infer the spatial placement of modified cysteine residues relative to the DNA helix (Flaus and Richmond, 1999).

One of the earliest reports of experiments utilising site-directed metal-affinity cleavage came from the Ebright laboratory where *S*-(2-pyridylthio)cysteaminyl-EDTA-metal (EPD-metal), an aromatic disulphide derivative of EDTA-metal, was used to probe the interaction sites of catabolite gene activating protein (CAP) and Cro with DNA (Ebright *et al.,* 1992). Ermacora *et al.* (1992) made use of EPD-metal to obtain information of the non-native state of staphylococcal nuclease. The following year, the Ebright group reported the interaction of Cro with DNA by employing a haloacetal EDTA-metal derivative, *N*-(iodoacetyl)-*p*-phenylenediamine-EDTA:metal (Ebright *et al.,* 1993). This group also demonstrated the stable incorporation of europium and Fe^{55} in complex with *N*-(iodoacetyl)-*p*-phenylenediamine-EDTA, allowing site-specific fluorescent and radioactive tagging of Cro derivatives.

In the chromatin field, the most notable studies employing affinity cleavage came from Jeff Hayes's study of the location of the linker histone H1° on nucleosomes positioned on the 5S rDNA from *Xenopus borealis* (Hayes, 1996) and Flaus and Richmond's base-pair resolution mapping of nucleosome positions (Flaus *et al.,* 1996). Hayes made use of EPD-metal, while Flaus and Richmond proposed an alternative aromatic EDTA-

derivative, *S*-(nitrophenylsulfenyl)-cysteaminyl-EDTA. The latter compound was synthesised in the study discussed here (see Chapter 5).

4.3. Identification of candidate mutation sites

This section deals with the design of GH5 mutants to characterise the secondary DNA-binding site by site-directed metal-affinity cleavage. Since the aim of this study is to map the location of the secondary DNA-binding site in the nucleosome by affinity cleavage, candidate residues suitable for appropriate chemical modification were identified.

Residues at these identified positions must be substituted with cysteine, to allow covalent linkage of the hydroxyl-radical donor, *S*-(nitrophenylsulfenyl)-cysteaminyl-EDTA, to the cysteine via a disulphide bond. A model of cysteaminyl-EDTA, coupled to the appropriate cysteine residues was used to help identify the potential cutting sites of hydroxyl radicals liberated from the donor. This was built from the chemical structure formula and energy minimised with the sensitive Polak-Ribière algorithm (Polak and Ribière, 1969; see Figure 4.2). Three residues that are in the vicinity of the proposed secondary DNA-binding site and are also proximal to the DNA molecule in the model structure (see above), were mutated *in silico* (Ala 96 → Cys 96, Ile 36 → Cys 36, Arg 37 → Cys 37). The resulting structures were optimised by applying the energy minimisation algorithm of Polak-Ribière.

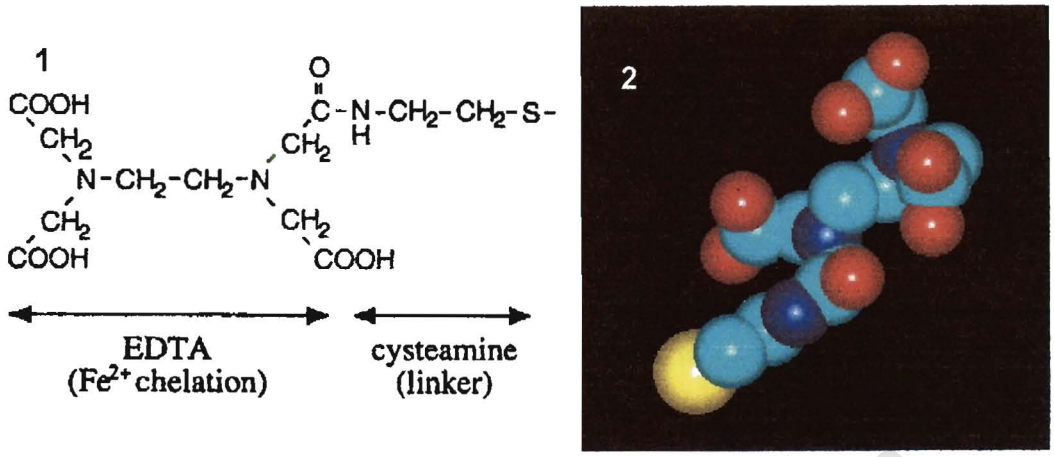


Figure 4.2. Structure of the hydroxyl radical generating adduct, cysteaminyl-EDTA. 1. Structure formula. **2.** Space-filled representation of the energy-minimised model.

It is known that hydroxyl radicals diffuse approximately 10 – 15 Å in solution (Ebright *et al.*, 1992), and that the distance between the major cleavage site in DNA and the modified cysteine thiol and α-carbon is approximately 7 Å and 10 Å, respectively (Flaus and Richmond, 1999). In addition, the adduct is rigid, containing only four rotatable bonds, and may form H-bonds to stabilize it (Flaus and Richmond, 1999). It is therefore expected that the major cutting sites should be within a 13 Å sphere of the cysteine C_α atom. In addition, the maximum distance from the cysteine thiol where hydroxyl radicals are expected to be active is 22 Å, since the cysteine thiol is 7 Å away from the reactive centre, and the hydroxyl radical migration from this centre persists for a maximum of 15 Å (Figure 4.3). The observed cutting sites from each of the three mutants suggested above are listed in Table 4.1.

Ala 96 is not associated with a secondary structural region and the incorporation of cysteine at this position would most likely not affect the proper folding of the protein. However, six ribose rings from the linker DNA are expected to be exposed to hydroxyl radicals. This wide range of DNA cleavage sites is likely to preclude a confident determination of the position of Cys 96 in the nucleosome.

Ile 36 and Arg 37 lie on the edge of helix I. Although cysteine is not known to be a helix-breaker, the change from an apolar to polar character (Ile 36 → Cys 36) might influence proper folding of the expressed mutants. However, the substitution of arginine at position 37 with cysteine retains the polar nature of the residue at this position, and is expected not to influence the folding of the Arg 37 → Cys 37 mutant GH5.

Table 4.1. Evaluation of the expected cutting sites from the modelled GH5 mutants. The major cutting site is defined by a 13 Å sphere from the cysteine C_α and the maximum diffusion distance of the hydroxyl radicals from the point of release is represented by the 15 Å sphere from the reactive centre (refer to Figure 4.3).

Mutant	*13 Å – sphere (major cutting site)*	*15 Å –sphere (after diffusion)*
Ala 96 → Cys 96	• DNA in region of dyad axis	• More extended piece of DNA at dyad axis • 6 ribose rings of linker DNA
Ile 36 → Cys 36	• Only dyad DNA	• Both DNA gyres
Arg 37 → Cys 37	• Only DNA of central gyre	• Only a few ribose rings of linker DNA

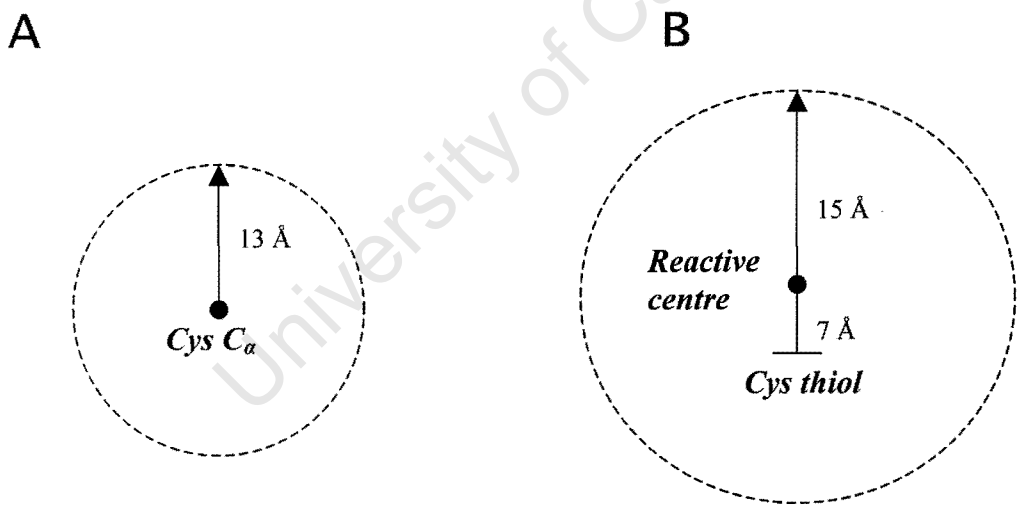


Figure 4.3. Schematic representation of the 13 Å (A) and 15 Å (B) spheres used to determine the major cutting sites of hydroxyl radicals liberated from putative substituted cysteine residues. The dotted circle delineates the extent of a hypothetical sphere within which hydroxyl radicals released from the EDTA moiety are expected to cut. In **A**, the cysteine C_α is at the centre of the circle, while the reactive centre is at the centre of the circle in **B**. The latter point is separated from the cysteine thiol by 7 Å, and from the outermost expected cutting site by 15 Å. Therefore the total distance between the cysteine thiol and expected cutting site is 22 Å.

The minimal possible exposure of the ribose rings of the linker DNA to hydroxyl radicals in the 15 Å - sphere, suggests that the Arg 37 → Cys 37 mutant is the best candidate to identify the position of the secondary DNA-binding site in a nucleosome. Zhou *et al.*, (1998) have shown that Ser 41 contacts nucleosomal DNA near the dyad axis of symmetry of the core particle. In the light of the observations made above, it was decided to design and over-express the following GH5 mutants: Arg 37 → Cys 37, Ala 96 → Cys 96 and Ser 41 → Cys 41. These proteins will be referred to as R37C, A96C and S41C, respectively. The location of these residues in the model of the chromatosome proposed in Chapter 3, is shown in Figure 4.4.

4.4. Overview of expression and purification of GH5 peptides

In order to sub-clone and express the cysteine-containing mutant globular domains of H5 as well as the native globular domains (spanning residues 20 – 109 and 23 – 106, respectively) in *E. coli*, the corresponding coding regions were mutated and amplified by PCR. Oligonucleotide pairs that introduce terminal *Nde* I and *Xho* I restriction enzyme sites as well as the coding sequence for an N-terminal MTE-motif were employed. The N-terminal motif was found to facilitate the expression of GH5 spanning residues 23 – 106 in *E. coli* (Gerchman *et al.*, 1993). The cleavage efficiency of *Nde* I close to fragment termini is very low (New England Biolabs, unpublished observations). Therefore, the 3' A single nucleotide overhang present in a large fraction of DNA molecules amplified with *Taq* DNA polymerase (Clark, 1988), was used to directly ligate

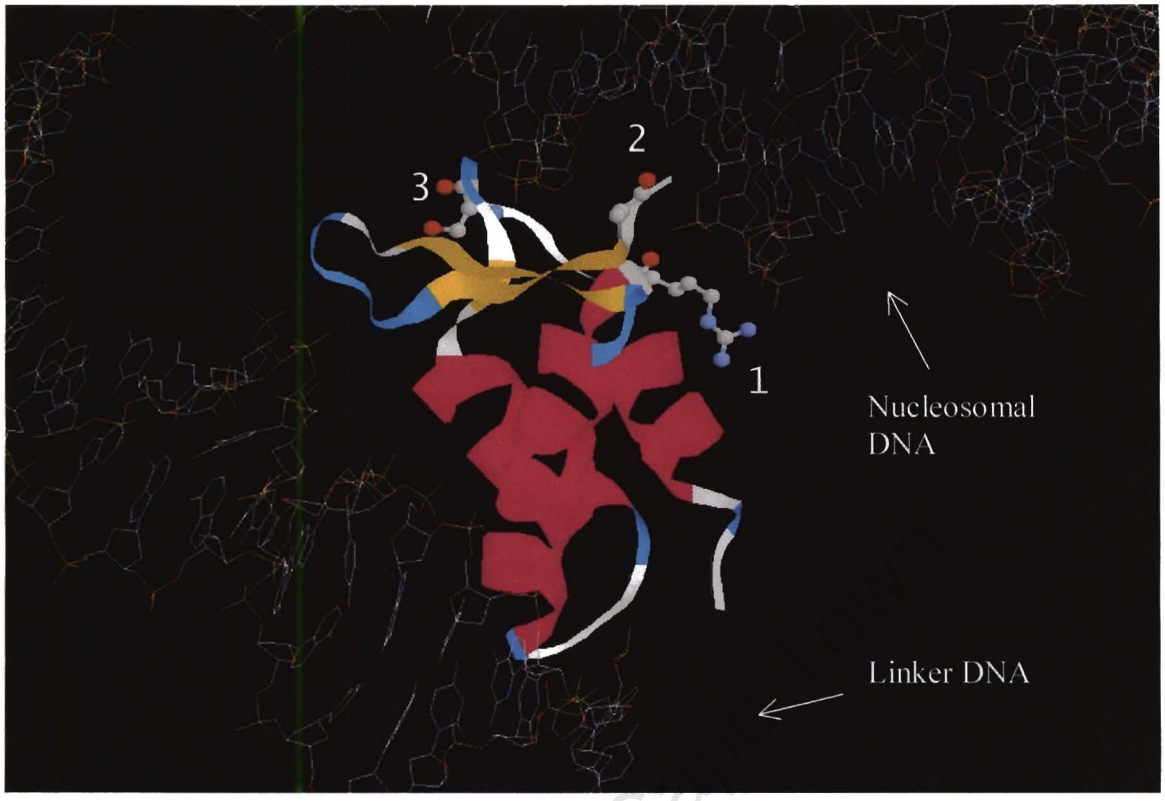


Figure 4.4. Ribbon representation of the backbone of GH5 bound to nucleosomal DNA as described by the model proposed in Chapter 3. The side-chains of the three residues chosen to be mutated to cysteine are shown in a ball-and-stick representation. Nucleosomal and linker DNA are indicated by arrows. **1.** Ser 41; **2.** Ala 96; **3.** Arg 37.

purified PCR products into the pGEM-T Easy vector (Promega), containing compatible 3' T overhangs. The fragments were subsequently sub-cloned into the *Nde* I and *Xho* I sites of the pET-20b(+) bacterial expression vector, in frame with six C-terminal histidine codons, a requirement for subsequent purification of the proteins by immobilized metal-affinity chromatography (IMAC; Sulkowski, 1989). All proteins were expressed in *E. coli* strain BL21(DE3) co-transformed with the pLysS plasmid in which expression conditions can be stringently regulated.

4.5. Template mismatch PCR mutagenesis

GH5 mutant coding sequences were created by employing a two-step template mismatch PCR technique (Kadowaki *et al.,* 1989). The first step of this technique involved the separate amplification of overlapping areas of the coding sequence for GH5 (residues 23 to 106 of full-length H5) from a plasmid template (pET9-H5, kindly provided by V. Ramakrishnan) containing the full-length H5 gene from *Gallus gallus* in the multiple cloning site. An oligonucleotide containing the required single (cysteine) codon mismatch together with an oligonucleotide complementary to either of the GH5 termini (flanking oligonucleotides) were employed in each of a pair of PCR reactions to produce overlapping segments (see Figures 4.5 and 4.6). This manipulation ensured the incorporation of the mutant cysteine codon (TGC). Flanking oligonucleotides were designed to introduce terminal *Nde* I (5' end) and *Xho* I (3' end) restriction enzyme cutting sites. In addition, since it is known that an N-terminal MTE-motif is required for the expression of GH5 (spanning residues 23 - 106) in *E. coli* (Gerchman *et al.,* 1993),

```
glob(f)(5'→3')          1     gggcatatgTCGCGGCGCTCGGCATCGCACCCCA          34
MTE-glob(f)(5'→3')      1     gggcatatgatgacggagTCGGCATCGCACCCCAC         35
pET9-H5(5'→3')          1   GGGTGAAGGCATCGCGGCGCTCGGCATCGCACCCCACCTACTCGGA 46
pET9-H5(3'→5')          1   CCCACTTCCGTAGCGCCGCGAGCCGTAGCGTGGGGTGGATGAGCCT 46

R37C(f)(5'→3')          1         GCGGCGGCCATCTGCGCGGAAAAGAGC             27
S41C(f)(5'→3')          1                    CGTGAAAAGTGCCGCGGCGGCT        22
pET9-H5(5'→3')         47 GATGATCGCGGCGGCCATCCGTGCGGAAAAGAGCCGCGGCGGCT 90
pET9-H5(3'→5')         47 CTACTAGCGCCGCCGGTAGGCACGCCTTTTCTCGGCGCCGCCGA 90
R37C(r)(3'→5')          1         CGCCGCCGGTAGACGCGCCTTTTCTCG             27
S41C(r)(5'→3')          1                    GCACGCCTTTTCACGGCGCCGCCGA      25

S41C(f)(5'→3')         23 CC                                               24
pET9-H5(5'→3')         91 CCTCGCGGCAGTCCATCCAGAAGTACATCAAGAGCCACTACAAG 134
pET9-H5(3'→5')         91 GGAGCGCCGTCAGGTAGGTCTTCATGTAGTTCTCGGTGATGTTC 134
S41C(r)(3'→5')         26 GG                                               27


pET9-H5(5'→3')        135 GTGGGCCACAACGCCGATCTGCAGATCAAGCTCTCCATCCGACG 178
pET9-H5(3'→5')        135 CACCCGGTGTTGCGGCTAGACGTCTAGTTCGAGAGGTAGGCTGC 178

pET9-H5(5'→3')        179 TCTCCTGGCTGCCGGCGTCCTCAAGCAGACCAAAGGGGTCGGGG 222
pET9-H5(3'→5')        179 AGAGGACCGACGGCCGCAGGAGTTCGTCTGGTTTCCCCAGCCCC 222

A96C(f)(5'→3')          1          TCCTTCCGCTTGTGCAAGAGCGACAAG            27
pET9-H5(5'→3')        223 CCTCCGGCTCCTTCCGCTTGGCCAAGAGCGACAAGGCCAAGAGG 266
pET9-H5(3'→5')        223 GGAGGCCGAGGAAGGCGAACCGGTTCTCGCTGTTCCGGTTCTCC 266
A96C(r)(3'→5')          1          AGGAAGGCGAACACGTTCTCGCTGTTC             27
MTE-glob(r)(3'→5')      1                                GTTCCGGTTCTCC  13
glob(r)(3'→5')          1                                     TCTCC     5

pET9-H5(5'→3')        267 TCCCCCGGGAAGAAGAAGAAGGGCCGTCAGGAGGTCCACGTCTC 310
pET9-H5(3'→5')        267 AGGGGGCCCTTCTTCTTCTTCCCGGCAGTCCTCCAGGTGCAGAG 310
MTE-glob(r)(3'→5')     14 AGGGGGGGgagctcggg                               28
glob(r)(3'→5')          6 AGGGGGCCCTTCTTCgagctcggg                        29
```


Figure 4.5. Template mismatch PCR mutagenesis to create cysteine-containing GH5 mutants. The area of the PCR template sequence (pET9-H5) is shown in double-stranded format in italics. The location and sequence of all oligonucleotides used to generate the coding sequences for native GH5 domains, as well as mutants R37C, S41C and A96C are shown. Oligonucleotides are shown in bold print. Mismatch sequences required for the incorporation of a cysteine codon are highlighted in black. Engineered overhangs are shown in lower case. Restriction enzyme cutting sites are underlined (catatg = *Nde* I; gagctc = *Xho* I). The oligonucleotide MTE-glob(f) contains a 9 nucleotide motif (atgacggag) 3' to the *Nde* I site (catatg), coding for the amino acids MTE.

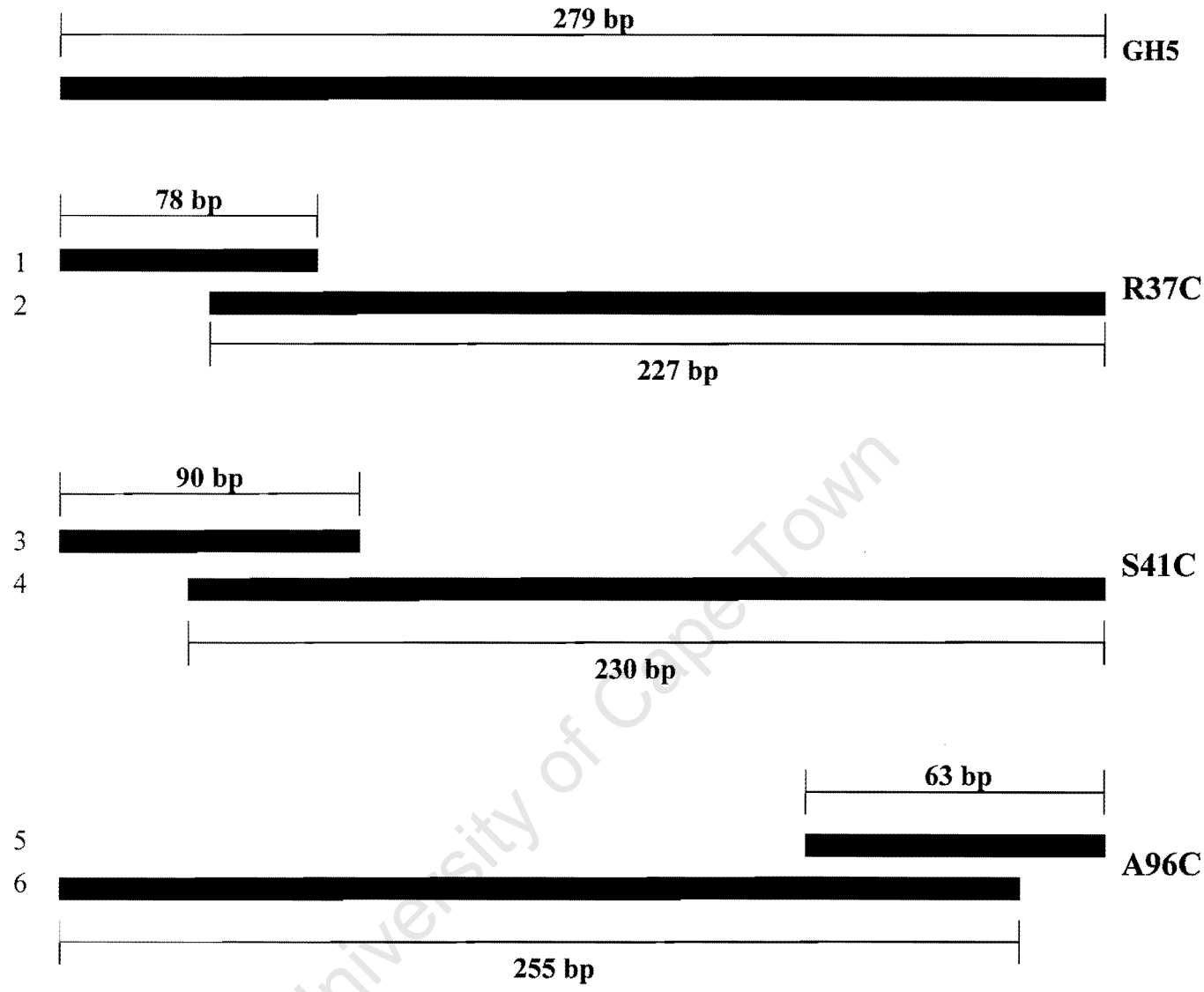


Figure 4.6. Schematic representation of the PCR products generated in the first step of template mismatch PCR mutagenesis. Two overlapping PCR products were amplified per mutant in separate PCR reactions. Each pair was then used as template in a third PCR reaction to give the full-length coding sequence (279 bp, shown at the top). **1.** Product of oligonucleotide pair MTE-glob(f) and R37C(r). Refer to Fig. 4.7 A below, lane 5; **2.** Product of oligonucleotide pair MTE-glob(r) and R37C(f). Refer to Fig. 4.7 A below, lane 4; **3.** Product of oligonucleotide pair MTE-glob(f) and S41C(r). Refer to Fig. 4.7 B below, lane 3; **4.** Product of oligonucleotide pair MTE-glob(r) and S41C(f). Refer to Fig. 4.7 B below, lane 2; **5.** Product of oligonucleotide pair MTE-glob(r) and A96C(f). Refer to Fig. 4.7 A below, lane 2; **6.** Product of oligonucleotide pair MTE-glob(f) and A96C(r). Refer to Fig 4.7 A below, lane 3.

the oligonucleotide complementary to the coding sequence for the N-terminus of GH5 was engineered to introduce codons coding for this motif. In a second round of PCR, the two products from the previous reactions were combined to serve as template for the amplification of the entire globular domain (residues 23 to 106) by overlap extension PCR, employing the same flanking oligonucleotides that had been used separately in the first round of PCR. Coding sequences for three GH5 mutants, Ala 96 → Cys 96 (A96C), Ser 41 → Cys 41 (S41C) and Arg 37 → Cys 37 (R37C) were produced in this way (Figure 4.7). As control, the coding sequence for the native globular domain (residues 23 to 106) was also amplified, utilizing the same flanking oligonucleotides employed to construct the mutant sequences.

To address the domain-specific requirement for an N-terminal MTE-motif for the expression of GH5 in *E. coli*, a broader area of the H5 gene (residues 20 to 109) was also amplified with oligonucleotides lacking the coding sequence for an N-terminal MTE-motif. It should be noted that no GH5 mutants discussed in this study were constructed in this way, and that all GH5 mutants were engineered to contain the N-terminal MTE-motif.

4.6. Fidelity of template mismatch PCR mutagenesis

Various error rates have been reported for *Taq* DNA polymerase, which lacks proofreading activity, varying between one error per 9 kb (Barnes, 1992) and one error per 13.8 kb (Ling *et al.*, 1991). The sizes of the amplified globular domains of GH5 and

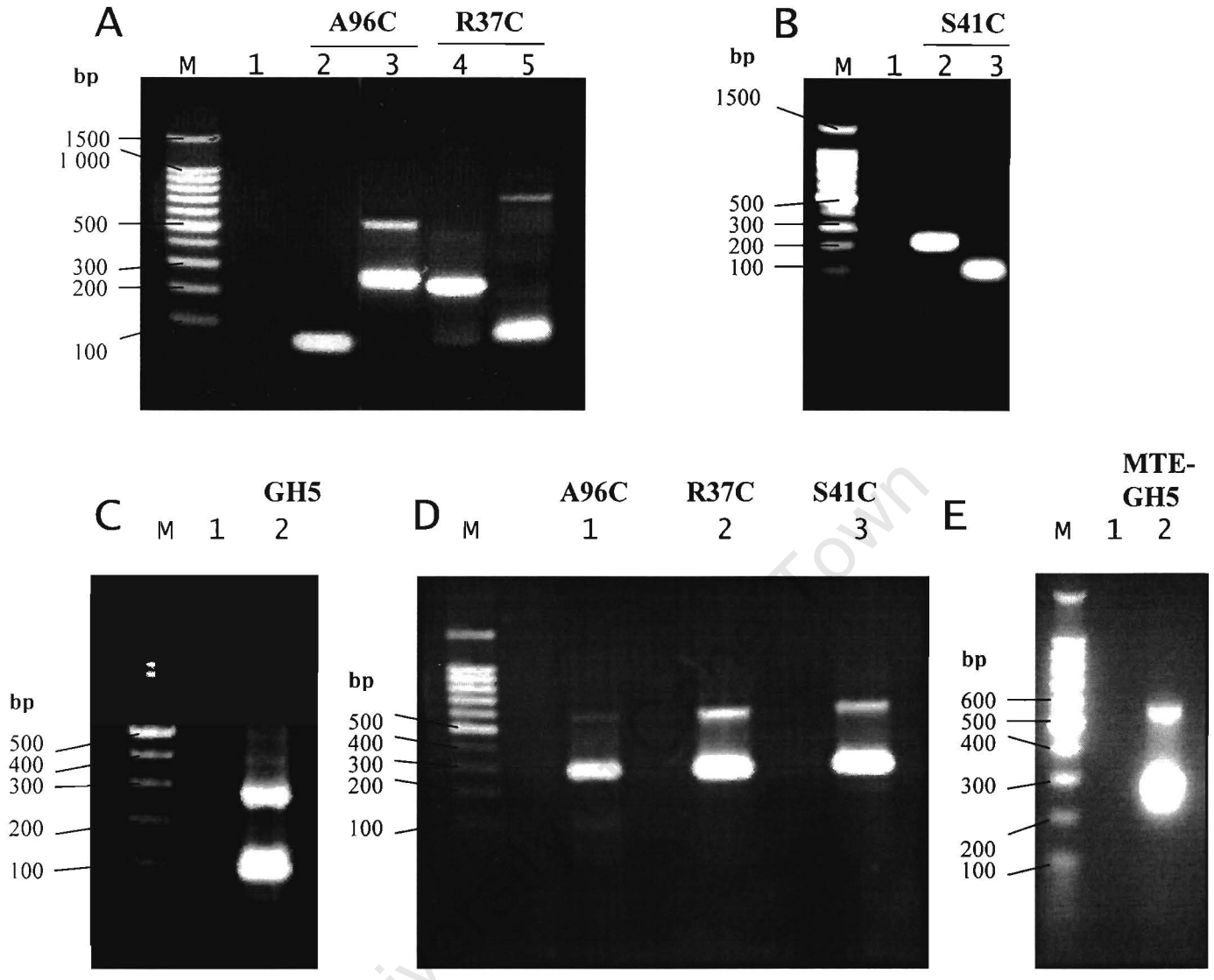


Figure 4.7. Products of template mismatch PCR mutagenesis. 1% (w/v) agarose gels in 1x TAE containing 0.5 μg/mℓ ethidium bromide and visualised under short wavelength UV transillumination. All lanes marked **M** contain the same 100 bp DNA molecular weight marker (Promega). **A.** Initial PCR products for A96C and R37C. Lane 1: Negative control (with oligonucleotides MTE-glob(f) and MTE-glob(r)); Lane 2: PCR product amplified with oligonucleotides MTE-glob(r) and A96C(f), 63 bp; Lane 3: PCR product amplified with oligonucleotides MTE-glob(f) and A96C(r), 255 bp and a 500bp non-specific product; Lane 4: PCR product amplified with oligonucleotides R37C(f) and MTE-glob(r), 227 bp; Lane 5: PCR product amplified with oligonucleotides R37C(r) and MTE-glob(f), 78 bp and a 700bp non-specific product. **B.** Initial PCR products for S41C. Lane 1: Negative control (with oligonucleotides MTE-glob(f) and MTE-glob(r)); Lane 2: PCR product amplified with oligonucleotides S41C(f) and MTE-glob(r), 230 bp; Lane 3: PCR product amplified with oligonucleotides S41C(r) and MTE-glob(f), 90 bp. **C.** PCR products for GH5. Lane 1, negative control (with oligonucleotides glob(f) and glob(r)); Lane 2: 288 bp PCR product corresponding to the coding sequence for GH5 (residues 20 to 109 lacking N-terminal MTE-motif) amplified with oligonucleotides glob(f) and glob(r)), as well as a 100bp non-specific product. **D.** Final PCR products for A96C, R37C and S41C. Lanes 1, 2 and 3: 279bp PCR products corresponding to the coding sequences for A96C, R37C and S41C (all amplified with oligonucleotides MTE-glob(f) and MTE-glob(r)), respectively, as well as a ~600bp non-specific product observed in each case. **E.** PCR products for MTE-GH5. Lane 1, negative control (with oligonucleotides MTE-glob(f) and MTE-glob(r)); Lane 2: 279 bp PCR product corresponding to the coding sequence for native GH5 (residues 23 to 106 preceded by an N-terminal MTE-motif) amplified with oligonucleotides MTE-glob(f) and MTE-glob(r)), as well as a ~600 bp non-specific product.

its mutants are 279 bp for the domain spanning residues 23 to 106, where an N-terminal MTE-motif had been incorporated, and 288 bp in the case of the broader domain (residues 20 to 109) lacking an N-terminal MTE-motif. This suggests that *Taq* DNA polymerase could faithfully catalyse the PCR reactions required in the construction of GH5 mutants and the native domains. This enzyme has also been used successfully by Kadowaki *et al.* (1989) in the pioneering site-directed PCR mutagenesis of an insulin receptor.

Because the lack of proofreading activity is a concern when the products of a PCR reaction catalyzed by such a polymerase are to be translated *in vivo*, *Taq*-catalyzed PCR products were sub-cloned into the pGEM-T Easy vector and sequenced to detect random amplification errors. These errors are not evident when sequencing PCR products. Automated sequencing of sub-cloned GH5 PCR products revealed an error rate of approximately one error per sequenced clone (Figure 4.8). The unusually high error rate observed here is most likely because magnesium chloride and deoxyribonucleotide triphosphates were not present in equimolar amounts in the PCR reaction that had been optimized in terms of yield (Eckert and Kunkel, 1991).

Therefore, with the exception of the amplification of the domain spanning residues 20 – 106 of H5, *Pfu* DNA polymerase, a proofreading enzyme, was employed in all PCR reactions. The error rate of this enzyme has been reported to be 1.3×10^{-6} errors per base pair synthesized (Cline *et al.*, 1996). However, *Pfu*-amplified DNA fragments do not

```
pGEM-S41C (clone 4)    1   CATATGATGACGGAGTCGGCATCGCACCCCACCTACTCGGAGATGATCGCGG 52
pGEM-S41C (clone 5)    1   CATATGATGACGGAGTCGGCATCGCACCCCACCTACTCGGAGATGATCGCGG 52
MTE-GH5 (correct)      1   CATATGATGACGGAGTCGGCATCGCACCCCACCTACTCGGAGATGATCGCGG 52
pGEM-GH5 (incorrect)   1   CATATGATGACGGAGTCGGCATCGCACCCCACCTACTCTGAGATGATCGCGG 52
pGEM-A96C (clone 1)    1   CATATGATGACGGAGTCGGCATCGCACCCCACCTACTCGGAGATGATCGCGG 52

pGEM-S41C (clone 4)    53  CGGCCATCCGTGCGGAAAAGTGCCGCGGCGGCTCCTCGCGGCAGTCCATCCA 104
pGEM-S41C (clone 5)    53  CGGCCATCTGTGCGGAAAAGTGCCGCGGCGGCTCCTCGCGGCAGTCCATCCA 104
MTE-GH5 (correct)      53  CGGCCATCCGTGCGGAAAAGAGCCGCGGCGGCTCCTCGCGGCAGTCCATCCA 104
pGEM-GH5 (incorrect)   53  CGGCCATCCGTGCGGAAAAGAGCCGCGGCGGCTCCTCGCGGCAGTCCATCCA 104
pGEM-A96C (clone 1)    53  CGGCCATCCGTGCGGAAAAGAGCCGCGGCGGCTCCTCGCGGCAGTCCATCCA 104

pGEM-S41C (clone 4)    105 GAAGTACATCAAGAGCCACTACAAGGTGGGCCACAACGCCGATCTGCAGATC 156
pGEM-S41C (clone 5)    105 GAAGTACATCAAGAGCCACTACAAGGTGGGCCACAACGCCGATCTGCAGATC 156
MTE-GH5 (correct)      105 GAAGTACATCAAGAGCCACTACAAGGTGGGCCACAACGCCGATCTGCAGATC 156
pGEM-GH5 (incorrect)   105 GAAGTACATCAAGAGCCACTACAAGGTGGGCCACAACGCCGATCTGCAGATC 156
pGEM-A96C (clone 1)    105 GCAGTACATCAAGAGCCACTACAAGGTGGGCCACAACGCCGATCTGCAGATC 156

pGEM-S41C (clone 4)    157 AAGCTCTCCATCCGACGTCTCCTGGCTGCCGGCGTCCTCAAGCAGACCAAAG 208
pGEM-S41C (clone 5)    157 AAGCTCTCCATCCGACGTCTCCTGGCTGCCGGCGTCCTCAAGCAGACCAAAG 208
MTE-GH5 (correct)      157 AAGCTCTCCATCCGACGTCTCCTGGCTGCCGGCGTCCTCAAGCAGACCAAAG 208
pGEM-GH5 (incorrect)   157 AAGCTCTCCATCCGACGTCTCCTGGCTGCCGGCGTCCTCAAGCAGACCAAAG 208
pGEM-A96C (clone 1)    157 AAGCTCTCCATCCGACGTCTCCTGGCTGCCGGCGTCCTCAAGCAGACCAAAG 208

pGEM-S41C (clone 4)    209 GGGTCGGGGCCTCTGGCTCCTTCCGCTTGGCCAAGAGCGACAAGGCCAAGAG 260
pGEM-S41C (clone 5)    209 GGGTCGGGGCCTCCGGCTCCTTCCGCTTGGCCAAGAGCGACAAGGCCAAGAG 260
MTE-GH5 (correct)      209 GGGTCGGGGCCTCCGGCTCCTTCCGCTTGGCCAAGAGCGACAAGGCCAAGAG 260
pGEM-GH5 (incorrect)   209 GGGTCGGGGCCTCCGGCTCCTTCCGCTTGGCCAAGAGCGACAAGGCCAAGAG 260
pGEM-A96C (clone 1)    209 GGGTCGGGGCCTCCGGCTCCTTCCGCTTGTGCAAGAGCGACAAGGCCAAGAG 260

pGEM-S41C (clone 4)    261 GTCCCCCCTCGAG 273
pGEM-S41C (clone 5)    261 GTCCCCCCTCGAG 273
MTE-GH5 (correct)      261 GTCCCCCCTCGAG 273
pGEM-GH5 (incorrect)   261 GTCCCCCCTCGAG 273
pGEM-A96C (clone 1)    261 GTCCCCCCTCGAG 273
```


Figure 4.8. ***Taq*-catalysed PCR errors.** The expected sequence of GH5 with an N-terminal MTE-motif (MTE-GH5) is aligned with sequences obtained from automated nucleotide sequencing of pGEM-clones. Each of the above clones contain a single nucleotide mismatch. Mismatched nucleotides which are a result of the template mismatch PCR are indicated in bold print, while true PCR errors are highlighted in black. Note that the clone labelled pGEM-GH5 contains the N-terminal MTE-motif.

exhibit a 3'-A single nucleotide overhang. The 3' terminal transferase activity of *Taq* DNA polymerase during a single cycle of *Taq*-catalyzed DNA extension was exploited in order to sub-clone gel-purified PCR products into pGEM-T Easy. Automated nucleotide sequencing of pGEM-T Easy constructs confirmed the fidelity of this method.

4.7. Expression of GH5 mutants in *E. coli*

The pET-20b(+) bacterial expression system (Novagen) was utilized for the expression of the mutant globular domains as well as the native domains, spanning residues 23 to 106 (with the N-terminal MTE-coding sequence) and 20 to 109, (lacking the N-terminal MTE-coding sequence) respectively, in *E. coli*. Expression is driven by an IPTG-inducible T7 promoter, which can be tightly regulated in the presence of T7 lysozyme, expressed from the pLysS plasmid. T7 lysozyme inhibits basal levels of T7 polymerase that is present prior to induction, and which can lead to counter-selection of plasmids from which toxic peptides are expressed (Baneyx, 1999).

pGEM-T Easy constructs containing native and mutant GH5 coding fragments, were digested with *Nde* I and *Xho* I to release inserts, which were subsequently ligated into appropriately digested pET-20b(+). The expression constructs containing GH5 inserts were transformed into *E. coli* strain XL1Blue for plasmid amplification and isolation, following a *Pst* I diagnostic screen (data not shown) and automated nucleotide sequencing (Figure 4.9). Hereafter, *E. coli* BL21(DE3) cells containing the pLysS

```
GH5     :       M  S  R  R    S  A  S    H  P  T    Y  S  E  M    I  A  A    A  I  R
MTE-GH5:        M  M  T  E    S  A  S    H  P  T    Y  S  E  M    I  A  A    A  I  R
R37C    :       M  M  T  E    S  A  S    H  P  T    Y  S  E  M    I  A  A    A  I  C
S41C    :       M  M  T  E    S  A  S    H  P  T    Y  S  E  M    I  A  A    A  I  R
A96C    :       M  M  T  E    S  A  S    H  P  T    Y  S  E  M    I  A  A    A  I  R

                      10           20           30           40           50           60
GH5     (ORF):  ATGTCGCGGC   GCTCGGCATC   GCACCCCACC   TACTCGGAGA   TGATCGCGGC   GGCCATCCGT
MTE-GH5(ORF):   ATGATGACGG   AGTCGGCATC   GCACCCCACC   TACTCGGAGA   TGATCGCGGC   GGCCATCCGT
R37C    (ORF):  ATGATGACGG   AGTCGGCATC   GCACCCCACC   TACTCGGAGA   TGATCGCGGC   GGCCATCTGC
S41C    (ORF):  ATGATGACGG   AGTCGGCATC   GCACCCCACC   TACTCGGAGA   TGATCGCGGC   GGCCATCCGT
A96C    (ORF):  ATGATGACGG   AGTCGGCATC   GCACCCCACC   TACTCGGAGA   TGATCGCGGC   GGCCATCCGT


GH5     :       A  E  K  S    R  G  G    S  S  R    Q  S  I  Q    K  Y  I    K  S  H
MTE-GH5:        A  E  K  S    R  G  G    S  S  R    Q  S  I  Q    K  Y  I    K  S  H
R37C    :       A  E  K  S    R  G  G    S  S  R    Q  S  I  Q    K  Y  I    K  S  H
S41C    :       A  E  K  G    R  G  G    S  S  R    Q  S  I  Q    K  Y  I    K  S  H
A96C    :       A  E  K  S    R  G  G    S  S  R    Q  S  I  Q    K  Y  I    K  S  H

                      70           80           90          100          110          120
GH5     (ORF):  GCGGAAAAGA   GCCGCGGCGG   CTCCTCGCGG   CAGTCCATCC   AGAAGTACAT   CAAGAGCCAC
MTE-GH5(ORF):   GCGGAAAAGA   GCCGCGGCGG   CTCCTCGCGG   CAGTCCATCC   AGAAGTACAT   CAAGAGCCAC
R37C    (ORF):  GCGGAAAAGT   GCCGCGGCGG   CTCCTCGCGG   CAGTCCATCC   AGAAGTACAT   CAAGAGCCAC
S41C    (ORF):  GCGGAAAAGA   GCCGCGGCGG   CTCCTCGCGG   CAGTCCATCC   AGAAGTACAT   CAAGAGCCAC
A96C    (ORF):  GCGGAAAAGA   GCCGCGGCGG   CTCCTCGCGG   CAGTCCATCC   AGAAGTACAT   CAAGAGCCAC


GH5     :       Y  K  V  G    H  N  A    D  L  Q    I  K  L  S    I  R  R    L  L  A
MTE-GH5:        Y  K  V  G    H  N  A    D  L  Q    I  K  L  S    I  R  R    L  L  A
R37C    :       Y  K  V  G    H  N  A    D  L  Q    I  K  L  S    I  R  R    L  L  A
S41C    :       Y  K  V  G    H  N  A    D  L  Q    I  K  L  S    I  R  R    L  L  A
A96C    :       Y  K  V  G    H  N  A    D  L  Q    I  K  L  S    I  R  R    L  L  A

                     130          140          150          160          170          180
GH5     (ORF):  TACAAGGTGG   GCCACAACGC   CGATCTGCAG   ATCAAGCTCT   CCATCCGACG   TCTCCTGGCT
MTE-GH5(ORF):   TACAAGGTGG   GCCACAACGC   CGATCTGCAG   ATCAAGCTCT   CCATCCGACG   TCTCCTGGCT
R37C    (ORF):  TACAAGGTGG   GCCACAACGC   CGATCTGCAG   ATCAAGCTCT   CCATCCGACG   TCTCCTGGCT
S41C    (ORF):  TACAAGGTGG   GCCACAACGC   CGATCTGCAG   ATCAAGCTCT   CCATCCGACG   TCTCCTGGCT
A96C    (ORF):  TACAAGGTGG   GCCACAACGC   CGATCTGCAG   ATCAAGCTCT   CCATCCGACG   TCTCCTGGCT


GH5     :       A  G  V  L    K  Q  T    K  G  V    G  A  S  G    S  F  R    L  A  K
MTE-GH5:        A  G  V  L    K  Q  T    K  G  V    G  A  S  G    S  F  R    L  A  K
R37C    :       A  G  V  L    K  Q  T    K  G  V    G  A  S  G    S  F  R    L  A  K
S41C    :       A  G  V  L    K  Q  T    K  G  V    G  A  S  G    S  F  R    L  A  K
A96C    :       A  G  V  L    K  Q  T    K  G  V    G  A  S  G    S  F  R    L  C  K

                     190          200          210          220          230          240
GH5     (ORF):  GCCGGCGTCC   TCAAGCAGAC   CAAAGGGGTC   GGGGCCTCCG   GCTCCTTCCG   CTTGGCCAAG
MTE-GH5(ORF):   GCCGGCGTCC   TCAAGCAGAC   CAAAGGGGTC   GGGGCCTCCG   GCTCCTTCCG   CTTGGCCAAG
R37C    (ORF):  GCCGGCGTCC   TCAAGCAGAC   CAAAGGGGTC   GGGGCCTCCG   GCTCCTTCCG   CTTGGCCAAG
S41C    (ORF):  GCCGGCGTCC   TCAAGCAGAC   CAAAGGGGTC   GGGGCCTCCG   GCTCCTTCCG   CTTGGCCAAG
A96C    (ORF):  GCCGGCGTCC   TCAAGCAGAC   CAAAGGGGTC   GGGGCCTCCG   GCTCCTTCCG   CTTGTGCAAG


GH5     :       S  D  K  A    K  R  S    P  G  K    K  L  E  H    H  H  H    H  H
MTE-GH5:        S  D  K  A    K  R  S    P  L  E    H  H  H  H    H  H  -    -  -
R37C    :       S  D  K  A    K  R  S    P  L  E    H  H  H  H    H  H  -    -  -
S41C    :       S  D  K  A    K  R  S    P  L  E    H  H  H  H    H  H  -    -  -
A96C    :       S  D  K  A    K  R  S    P  L  E    H  H  H  H    H  H  -    -  -

                     250          260          270          280          290     296
GH5     (ORF):  AGCGACAAGG   CCAAGAGGTC   CCCCGGGAAG   AAGCTCGAGC   ACCACCACCA  CCACCAC
MTE-GH5(ORF):   AGCGACAAGG   CCAAGAGGTC   CCCCCTCGAG   CACCACCACC   ACCACCAC--  -------
R37C    (ORF):  AGCGACAAGG   CCAAGAGGTC   CCCCCTCGAG   CACCACCACC   ACCACCAC--  -------
S41C    (ORF):  AGCGACAAGG   CCAAGAGGTC   CCCCCTCGAG   CACCACCACC   ACCACCAC--  -------
A96C    (ORF):  AGCGACAAGG   CCAAGAGGTC   CCCCCTCGAG   CACCACCACC   ACCACCAC--  -------
```


Figure 4.9. Sequences of open reading frames (ORFs) of pET20b(+) expression constructs as determined by automated nucleotide sequencing. Corresponding translated protein sequences are shown above. All mutations are highlighted in black.

plasmid, were transformed with pET20b(+)-GH5 expression constructs for subsequent expression. It was observed that constructs containing the coding sequence for an N-terminal MTE-motif preceding the canonical globular domain of H5 (residues 23 to 106), rapidly accumulated in BL21(DE3)pLysS cells in response to induction with IPTG (Figure 4.10). This result is consistent with the observation of Gerchman *et al.* (1993). However, the Ala 96 → Cys 96 mutant of GH5 could not be detected following induction in this way.

The broader GH5 domain also did not accumulate upon induction of a host *E. coli* BL21DE3(pLysS) culture with IPTG (Figure 4.10).

4.8. Purification of His-tagged proteins by IMAC

His-tagged recombinant GH5 mutants were purified by immobilized metal-affinity chromatography (IMAC; Sulkowski, 1989), which can be rationalized in terms of the coordination of histidine residues by immobilized heavy metals. Nickel-agarose resin (Novagen) charged with nickel sulphate was used for the purification of GH5 and GH5 mutants from *E. coli* BL21DE3(pLysS). Despite a concern that the disulphide-induced dimerisation of GH5 mutants under non-reducing conditions might occlude the His-tag from binding to nickel-agarose, all mutants expressed (S41C and R37C) could be effectively purified in this way (Figure 4.11). The only alteration to standard elution conditions was to decrease the imidazole concentration in the wash buffer from 60 mM to 40 mM in order to avoid significant premature elution of the His-tagged protein.

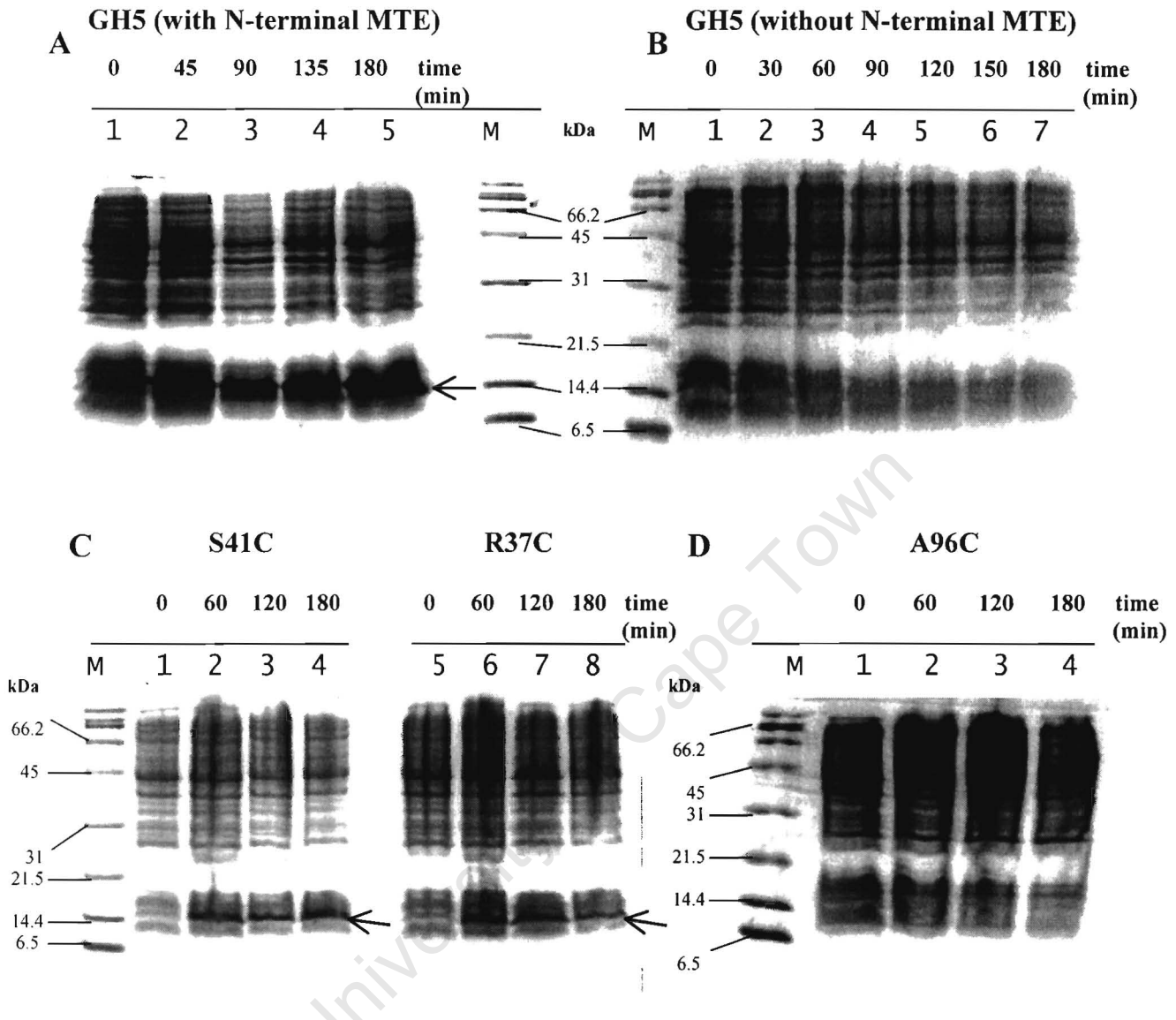


Figure 4.10. Induction of recombinant protein expression in *E. coli* strain BL21(DE3)pLysS. SDS-PAGE gels (15%) were visualised by Coomassie staining. In all of the above figures (**A-D**), lane M contains a broad range protein molecular weight marker (Bio-Rad). All the other lanes contain *E. coli* cells that had been lysed by adding sample application buffer prior to loading. **A.** Accumulation of native GH5 (with N-terminal MTE-motif) over various time intervals. Cells were induced for 0 (Lane 1), 45 min (Lane 2), 90 min (Lane 3), 135 min (Lane 4) and 180 min (Lane 5). The arrow indicates the position of GH5. GH5 accumulates and migrates at a size corresponding to ~14 400 Da. **B.** Native GH5 (residues 20 to 109) lacking an N-terminal MTE-motif is not expressed. Cells were induced for 0 (Lane 1), 30 min (Lane 2), 60 min (Lane 3), 90 min (Lane 4), 120 min (Lane 5), 150 min (Lane 6) and 180 min (Lane 7). **C.** GH5 mutants S41C and R37C accumulate in induced BL21(DE3)pLysS cultures. Lanes 1 to 4: S41C; Lanes 5 to 8: R37C. Cells were induced for 0 (Lane 1 and 5), 60 min (Lane 2 and 6), 120 min (Lane 3 and 7) and 180 min (Lane 4 and 8). GH5 mutants S41C and R37C accumulate and migrate at a size corresponding to ~14.4 kDa. Arrows indicate the positions of S41C and R37C respectively. **D.** GH5 mutant A96C is not expressed by BL21DE3pLysS cultures. Cells were induced for 0 (Lane 1), 60 min (Lane 2), 120 min (Lane 3), and 180 min (Lane 4).

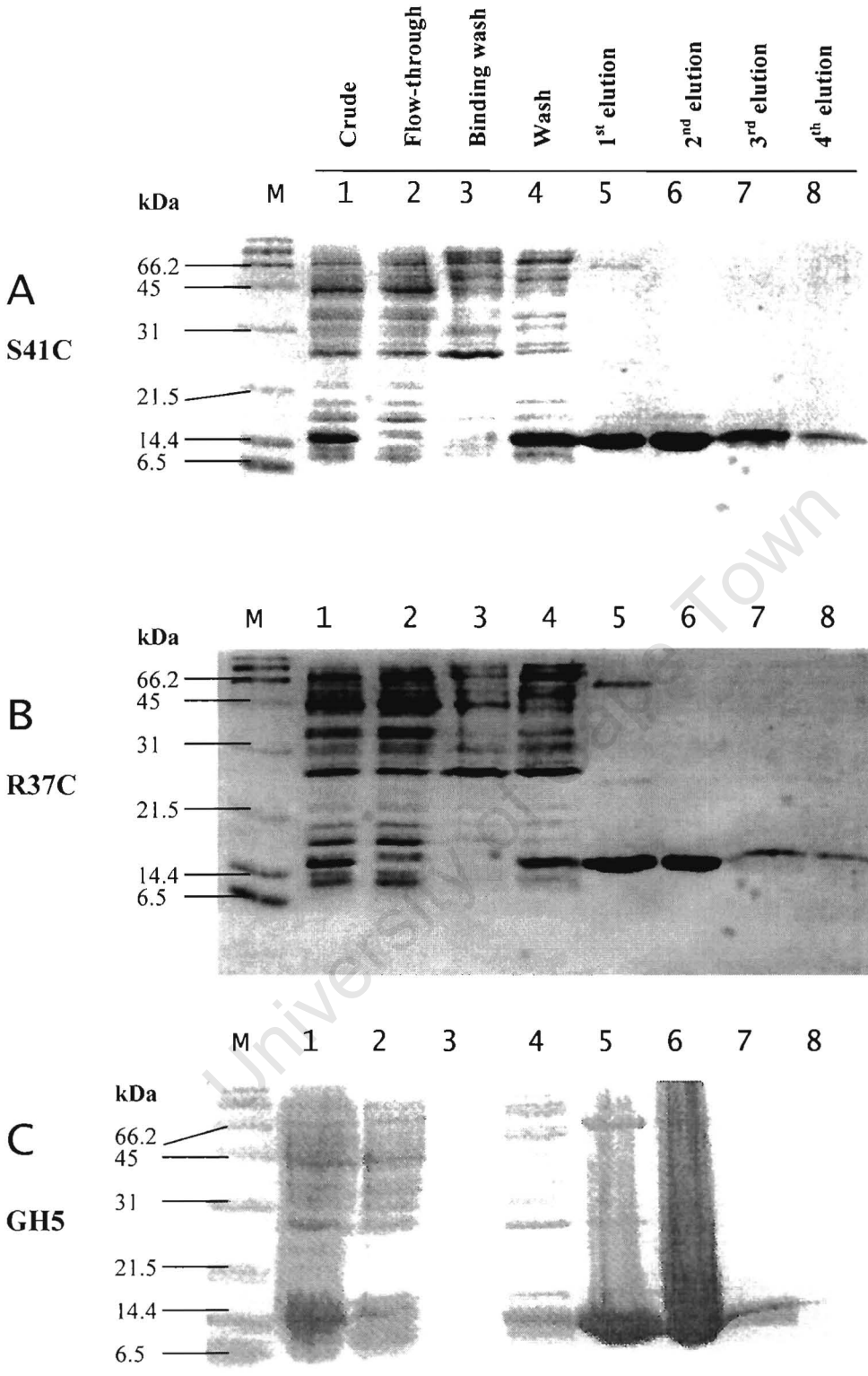


Figure 4.11. Elution profiles of His-tagged proteins from Nickel-agarose. Each figure shows a 15% SDS-PAGE gel stained with Coomassie brilliant blue dye. In all figures: Lanes M contains a broad-range protein molecular weight marker (Bio-Rad). Aliquots (20$\mu\ell$) were loaded from the following sources: the crude fraction that had been loaded onto the nickel-agarose column, flow-through, binding wash (5 mM imidazole), wash fraction (40 mM imidazole) and 4 consecutive elution fractions (1000 mM imidazole). The annotation of Lanes 1 to 8 in **A** is the same for **B** and **C**. **A.** S41C; **B.** R37C; **C.** native GH5. In panels **A** and **B**, lanes 1, 5 and 6 contain fractions that had been diluted 20×.

4.9. In-gel trypsin digest and MALDI-TOF analyses on GH5 and GH5 mutants

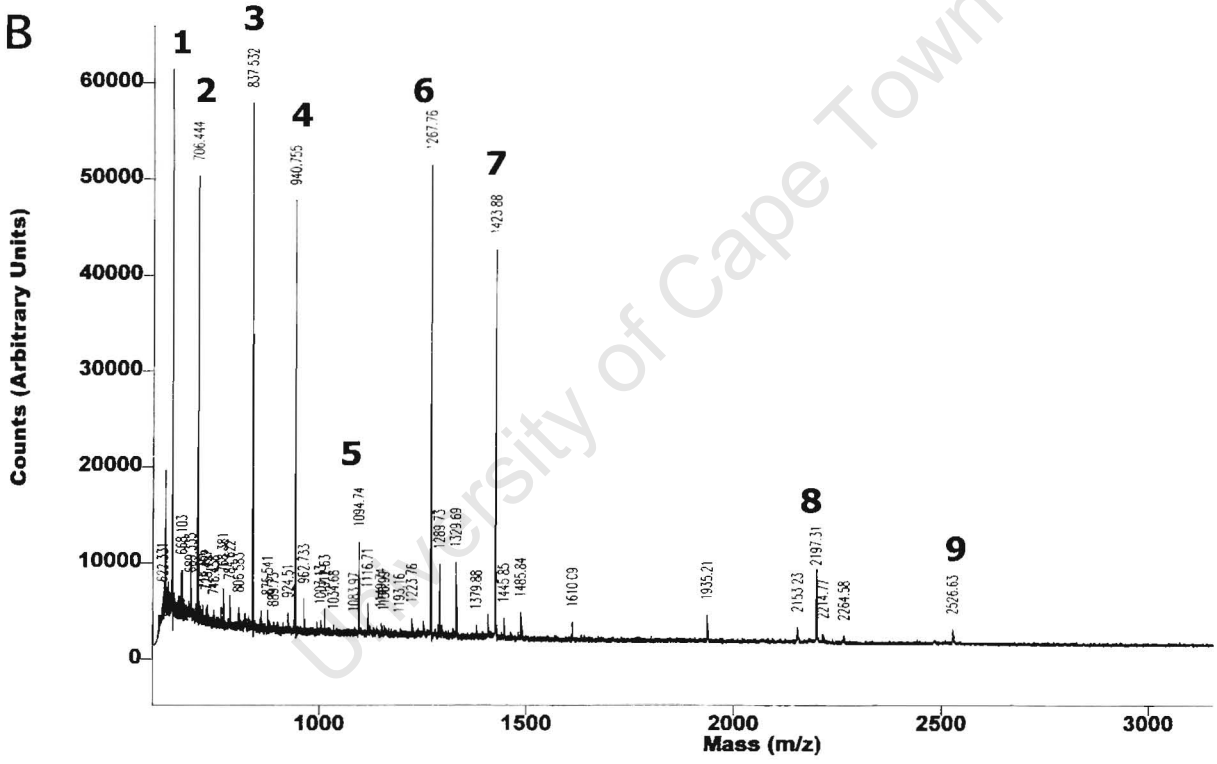
Although the size of GH5 (residues 23 to 106, preceded by an N-terminal MTE-motif) was calculated to be roughly 10 747 Da, the isolated His-tagged proteins exhibited anomalous migration on 15% SDS-PAGE gels, migrating at a size corresponding to approximately 14 400 Da. Automated nucleotide sequencing had confirmed that the coding region of the various expressed constructs were in frame with the His-tag (refer to Figure 4.9 above).

In-gel trypsin digests and subsequent MALDI-TOF analyses were employed to confirm the identity of purified proteins (Figure 4.12). Briefly, proteins were run on a 15% SDS-PAGE gel, stained with Coomassie brilliant blue and destained in a modified solution (see Chapter 2). Protein bands were excised from the gel and treated with Trypsin. The resulting peptides were analysed by MALDI-TOF. The matrix selected for these experiments, was α-cyano-4-hydroxycinnamic acid (α-CHC) and produces very intense signals from peptides and proteins. The intensities of the protonated peptide signals obtained with α-CHC are substantially higher than those obtained from other cinnamic acid derivatives, indicating that peptides are very efficiently ionized. In addition, α-CHC produces intense multiply charged ions in the positive ion spectra of proteins (Beavis *et al.,* 1992). Proteins also undergo significantly larger fragmentation in the spectrometer. As a result, α-CHC is recommended for peptides and proteins under 10 kDa of mass. Best results are obtained in the 500-5000 Da range where peptides liberated

A

MALDI-TOF Analysis of major peaks (>20000 counts)			
Segment	***Theoretical Mass (Da)***	***Experimental value (Da)***	***Assigned to:***
1	644.503	644.80	LSIR▼R
2	706.444	706.74	SR▼GGSSR
3	837.532	837.91	GVGASGSFR
4	940.755	941.21	R▼LLAAGVLK
5	1095.24	1094.74	VGHNADLQIK
6	1267.76	1268.34	SPLEHHHHHH
7	1423.88	1424.53	R▼SPLEHHHHHH
8	2197.69	2197.31	MMTEQSHPTYSEMIAAAIR
9	2526.53	2526.63	MMTEQSHPTYSEMIAAAIR▼AEK

B



C

MMTEQSHPTYSEMIAAAIR▼AEK▼SR▼GGSSR▼QSIQK▼YIK▼SHYK▼VGHNADLQIK▼

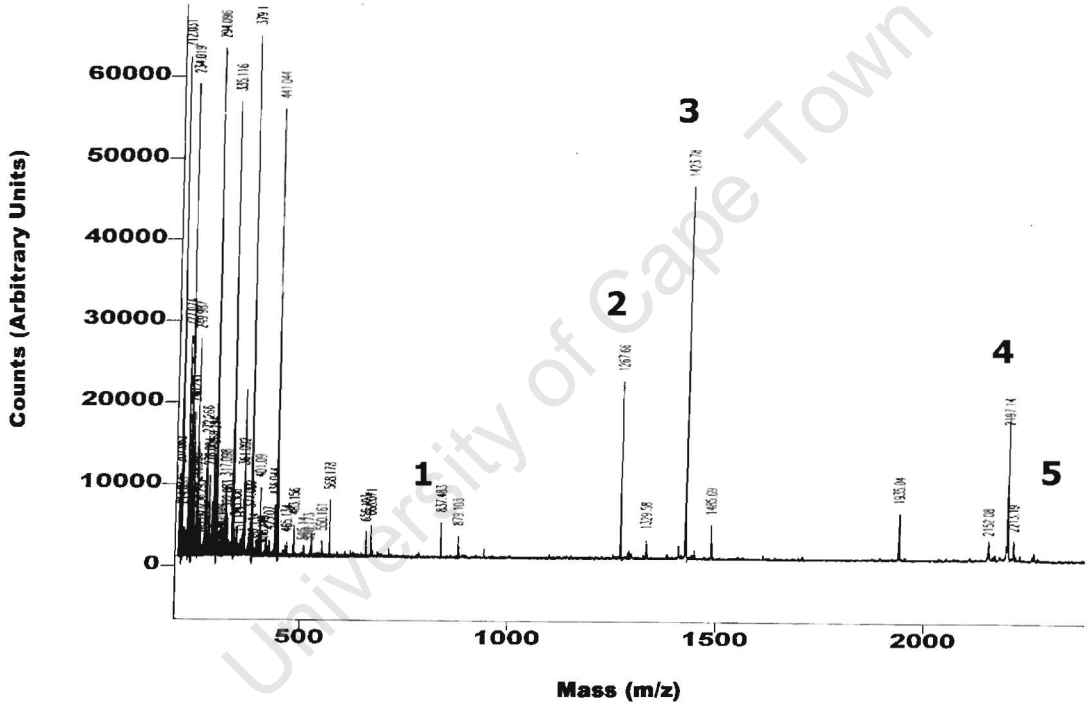
LSIR▼R▼LLAAGVLK▼QTK▼GVGASGSFR▼LAK▼SDK▼AK▼R▼SPLEHHHHHH

Figure 4.12. MALDI-TOF analyses of GH5 and GH5 mutants, S41C and R37C. A. Table showing the expected and experimental masses of peptides obtained by trypsin digest of purified GH5. Peptide sequences are also shown. **B.** Mass spectrum of GH5 trypsin digest. Mass peaks corresponding to the peptides in **A** are labelled accordingly. **C.** Expected sequence of GH5 expressed from pET20b(+)-GH5. Trypsin cleavage sites are indicated by (▼). Peptide fragments that could be assigned to experimental mass peaks are shown in colour.

D

MALDI-TOF Analysis of major peaks (>20000 counts)			
Segment	***Theoretical Mass (Da)***	***Experimental value (Da)***	***Assigned to:***
1	837.532	837.48	GVGASGSFR
2	1267.76	1267.66	SPLEHHHHHH
3	1423.88	1423.77	R▼SPLEHHHHHH
4	2197.69	2197.13	MMTEQSHPTYSEMIAAAIR
5	2526.53	2525.40	MMTEQSHPTYSEMIAAAIR▼AEK

E



F

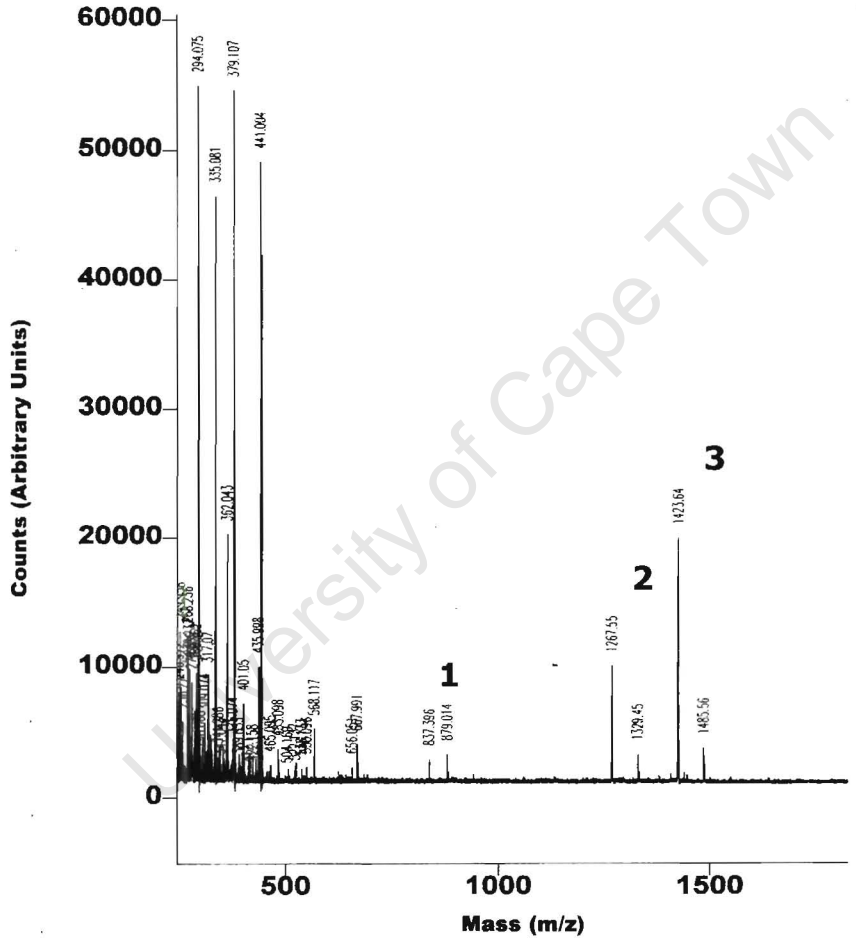
MMTEQSHPTYSEMIAAAIR▼AEK▼CR▼GGSSR▼QSIQK▼YIK▼SHYK▼VGHNADLQIK▼
LSIR▼R▼LLAAGVLK▼QTK▼GVGASGSFR▼LAK▼SDK▼AK▼R▼SPLEHHHHHH

Figure 4.12. (Continued). MALDI-TOF analysis of purified S41C. D. Table showing the expected and experimental masses of peptides obtained by trypsin digest of purified S41C. Peptide sequences are also shown. **E.** Mass spectrum of S41C trypsin digest. Mass peaks corresponding to the peptides in **D** are labelled accordingly. **F.** Expected sequence of S41C expressed from pET20b(+)-S41C. Trypsin cleavage sites are indicated by (▼). Peptide fragments that could be assigned to experimental mass peaks are shown in colour.

G

MALDI-TOF Analysis of major peaks (>20000 counts)			
Segment	***Theoretical Mass (Da)***	***Experimental value (Da)***	***Assigned to:***
1	837.532	837.48	**GVGASGSFR**
2	1267.76	1267.66	**SPLEHHHHHH**
3	1423.88	1423.77	**R▼SPLEHHHHHH**

H



I

MMTEQSHPTYSEMIAAAI**C**AEK▼SR▼GGSSR▼QSIQK▼YIK▼SHYK▼VGHNADLQIK▼L

SIR▼R▼LLAAGVLK▼QTK▼**GVGASGSFR**▼LAK▼SDK▼AK▼**R▼SPLEHHHHHH**

Figure 4.12. (Continued). MALDI-TOF analysis of purified R37C. G. Table showing the expected and experimental masses of peptides obtained by trypsin digest of purified R37C. Peptide sequences are also shown. **H.** Mass spectrum of R37C trypsin digest. Mass peaks corresponding to the peptides in **G** are labelled accordingly. **I.** Expected sequence of R37C expressed from pET20b(+)-R37C. Trypsin cleavage sites are indicated by (▼). Peptide fragments that could be assigned to experimental mass peaks are shown in colour.

from tryptic digests appear and therefore α-CHC is the matrix of choice for peptide mass fingerprinting (Beavis *et al.,* 1992).

Most mass-peaks could be assigned to masses of peptides generated by a theoretical trypsin digest on the sequences of GH5 and GH5 mutants. All experimental masses larger than 600 Da could be assigned to theoretical trypsin-generated peptides of native GH5. Because noise is generated by the matrix below this threshold, the low molecular mass area of the spectrum is not suitable for analytical purposes. A database search with the mass spectrum of GH5 identified the trypsin-digested peptide as part of H5 from *Gallus gallus* (chicken). Carboxypeptidase, isolated from a 15% SDS-PAGE was used as positive experimental control in the database search (data not shown). Peptides liberated by trypsin digest of S41C and R37C mutants gave much weaker signals than the native protein. In the case of S41C, the N- and C-termini, as well as a central nonapeptide (GVGASGSFR) could be detected by MALDI. A peptide containing the mutant cysteine 41 could not be detected. However, it should be noted that no native peptides corresponding to peptides where the mutant cysteine would be expected could be detected either.

The mass spectrum of trypsin-treated R37C could only identify C-terminal peptides and the nonapeptide noted in the case of S41C. In order to confirm the identity of the mutant proteins, the mass spectra of each were plotted against the mass spectrum of native GH5 (Figure 4.13). The absence of mass peaks larger than 2000 Da in the mass spectrum of

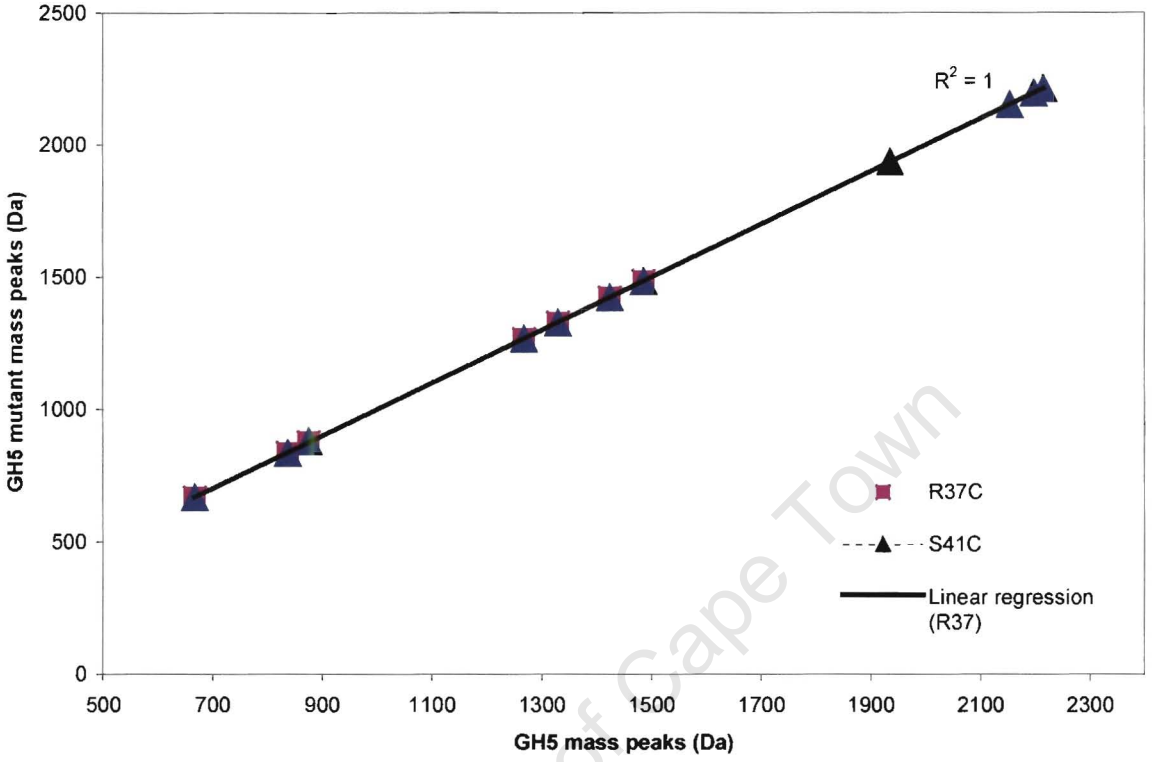


Figure 4.13. Plot of mass spectrum of GH5 and the mass spectra obtained by both R37C and S41C trypsin digests. The identity of GH5 was confirmed by both a database search and an accurate assignment of trypsin generated mass peaks to theoretical peptide masses. The R^2-value obtained from linear regression analysis is 1, showing the identity of these proteins. The reason for the observed identity, despite differences expected due to site-directed mutagenesis, is that no cysteine-containing peptides belonging to either S41C or R37C were detectable by MALDI. However, the presence of cysteine residues have been confirmed biochemically (see Figure 4.14). No R37C peptides in the range > 1500 Da was detected, as the Arg 37 → Cys 37 mutation causes a loss of the trypsin cleavage site responsible for liberating peptides in this mass range.

R37C is consistent with the loss of an N-terminal trypsin site caused by the substitution of Arg 37 for Cys.

4.10. Dimerisation of S41C and R37C under non-denaturing conditions

Although expression constructs coding for S41C and R37C had been confirmed by automated nucleotide sequencing, the fact that the mutated cysteine residues could not be detected by MALDI was a concern. However, when fractions containing pure S41C or R37C were loaded on a 15% SDS-PAGE gel without prior addition of β-mercaptoethanol to samples, an additional band migrating at double the size of native GH5 was observed (Figure 4.14 A). This suggests that these proteins dimerise under non-reducing conditions via disulphide bond formation and offers biochemical evidence for the presence of cysteine residues in R37C and S41C. The fact that S41C dimerises more strongly than R37C could be attributed to the observation that Cys 41 is more solvent accessible considering that GH5 is a dimer under physiological conditions (Figure 4.14 B). This phenomenon is indicative of the solvent-accessibility of the cysteine residues in both mutants, which will be essential for future reaction with *S*-(*tert*-butyl)-cysteaminyl-EDTA.

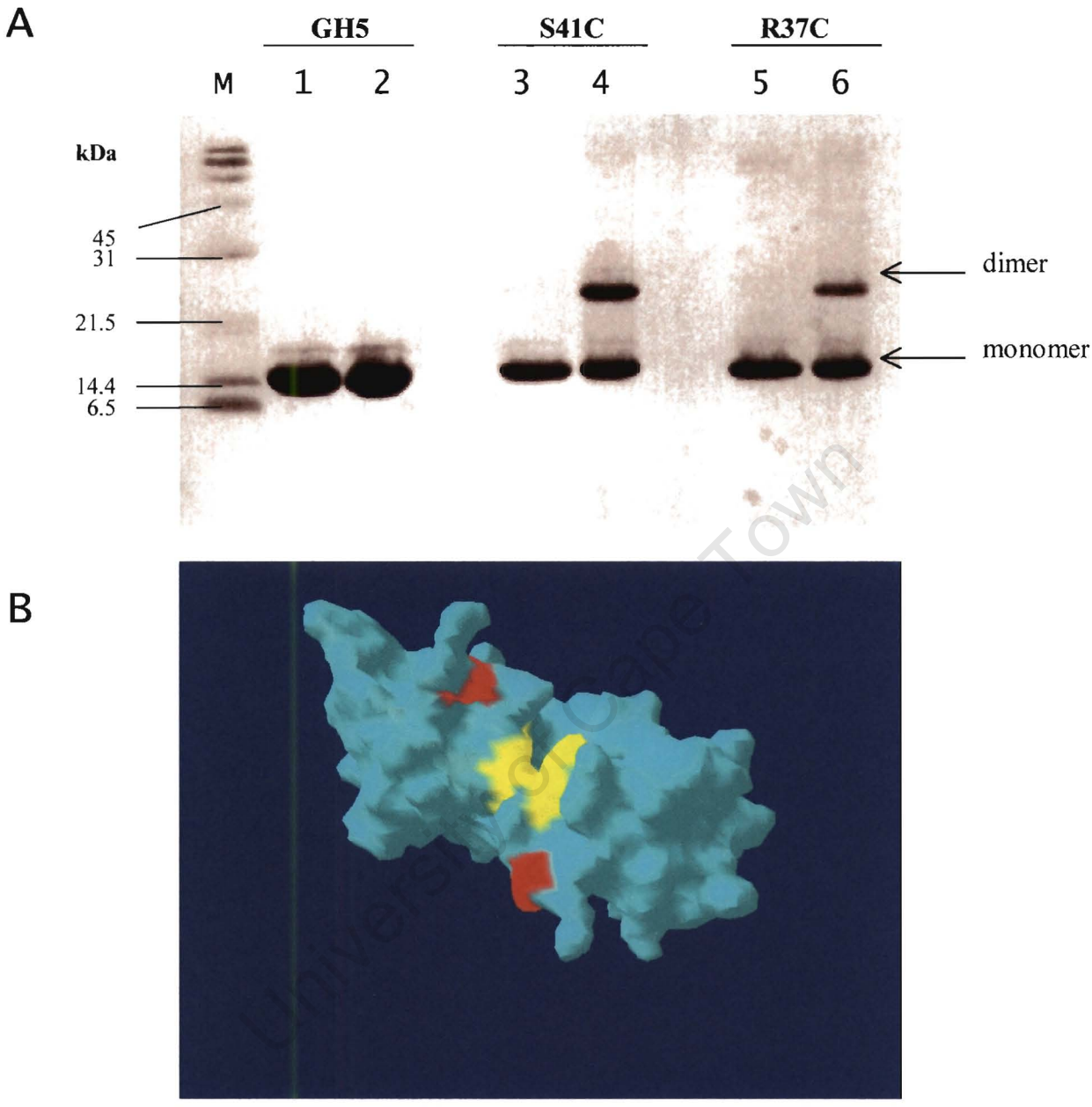


Figure 4.14. GH5 mutants S41C and R37C dimerises via disulphide linkages. **A.** 15% SDS-PAGE gel visualized by Coomassie Brilliant Blue staining. Lane M contains a broad range protein molecular weight marker (Bio-Rad). Lanes 1 and 2 contain identical GH5 fractions loaded with and without β-mercaptoethanol respectively. Lanes 3 and 4 contain identical S41C fractions loaded with and without β-mercaptoethanol respectively. The presence of a ~26 kDa band under non-reducing conditions suggests the formation of a S41C dimer. Lanes 5 and 6 contain identical R37C fractions loaded with and without β-mercaptoethanol respectively. The presence of a ~26 kDa band under non-reducing conditions suggests the formation of a R37C dimer. **B.** Molecular surface of the GH5 dimer as calculated using the program PDBViewer. The red areas indicate the areas occupied by Ser 41 while the yellow areas correspond to Arg 37.

CHAPTER 5: SYNTHESIS OF *S*-(nitrophenylsulphenyl)-cysteaminyl-EDTA

5.1. Introduction

The chemical synthesis of the site-directed metal affinity cleavage reagent, *S*-(nitrophenylsulfenyl)-cysteaminyl-EDTA is described in this chapter. The compound was synthesised according to the procedure as developed by Flaus and Richmond (1999). However, the inefficiencies of this method are pointed out, and a novel method for the synthesis of the compound is proposed.

5.2. Overview of synthesis routes

The scheme proposed by Flaus and Richmond (1999) served as basis for the synthesis of *S*-(nitrophenylsulfenyl)-cysteaminyl-EDTA (see Figure 5.1). According to this method, the thiol group of cysteamine is protected by a tertiary butyl group. A peptide bond is formed between the amine group of protected *tert*-butylcysteamine and the carboxylic acid function EDTA in a subsequent dicyclohexylcarbodiimide (DCC) mediated reaction. The resulting compound is purified by reverse phase HPLC and reacted with 2-nitrophenylsulfenyl chloride (NPS-Cl). This manipulation facilitates exchange between the tertiary butyl group with NPS to generate *S*-(nitrophenylsulfenyl)-cysteaminyl-EDTA, which is purified by reverse phase HPLC. The NPS-group is an effective leaving group

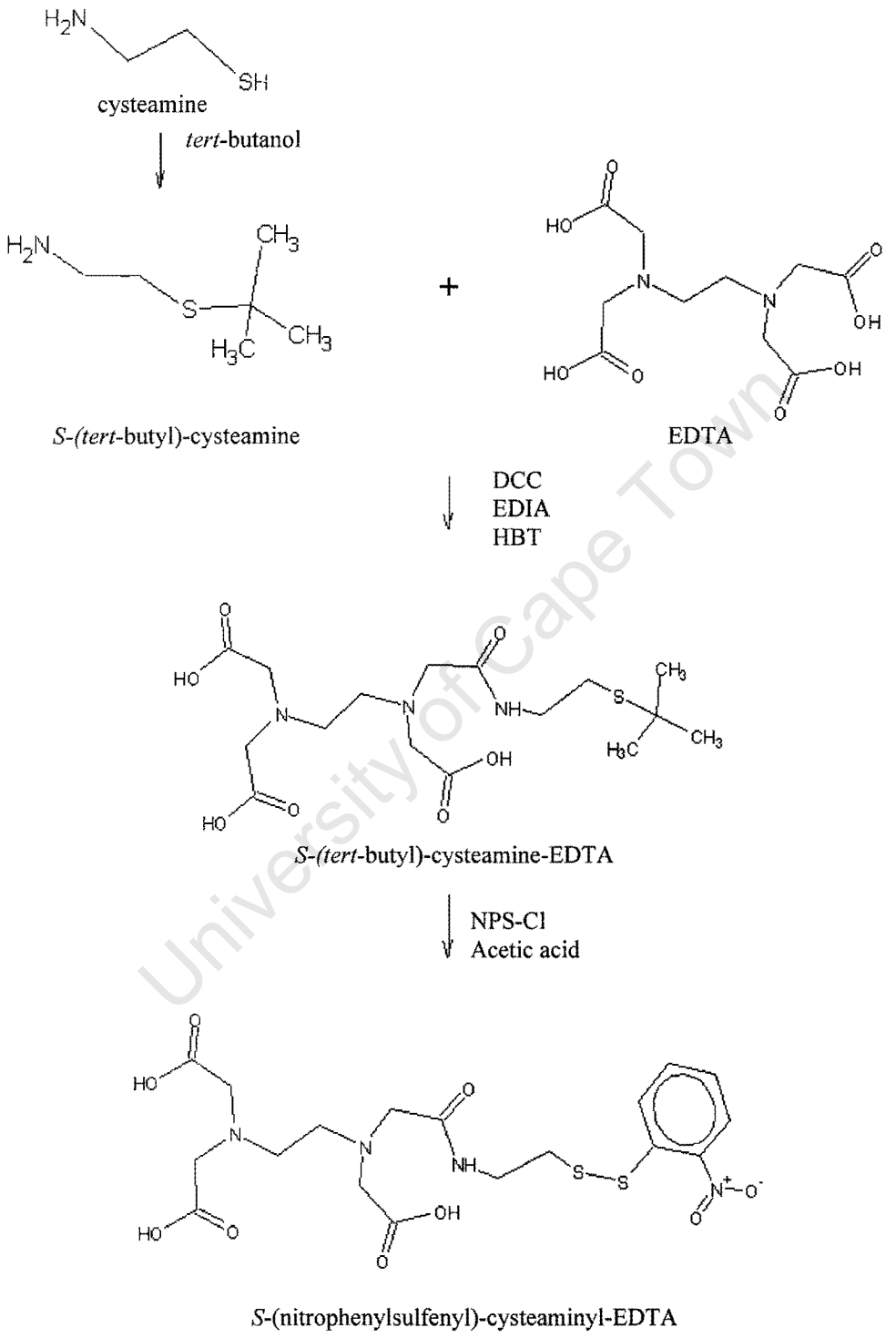


Figure 5.1. Structural representation of the synthesis scheme of *S*-(nitrophenylsulfenyl)-cysteaminyl-EDTA as published by Flaus and Richmond (1999).

and facilitates the formation of a disulphide bond between the cysteine thiol of a protein and the EDTA moiety of *S*-(nitrophenylsulfenyl)-cysteaminyl-EDTA.

Although the method of Flaus and Richmond (1999) has been successfully used in other laboratories, it has the following drawbacks:

1) The scale of the synthesis is limited by the two purification steps by HPLC. Typically, only milligram amounts of cyst-EDTA-NPS can be synthesized per run (G.K., unpublished data).
2) The yield of the DCC mediated reaction is low (typically <20%; G.K. unpublished data; see also Section 5.7.)
3) The symmetry of the EDTA molecule promotes side-reactions, which complicates effective purification procedures (Burks *et al.*, 1998).

To address these problems, a novel method was developed for the synthesis of *S*-(nitrophenylsulfenyl)-cysteaminyl-EDTA using conventional chemical techniques (Figure 5.2), which simplifies the purification procedure. As a result, the synthesis can be performed on a significantly larger scale to overcome the poor yields of the coupling reaction without the limitation of extensive preparative HPLC procedures. The tetra-ethyl ester of EDTA is used as starting material for this procedure, as opposed to acid form EDTA used in the method of Flaus and Richmond. Only one carboxyl group is selectively de-esterified by Pig Liver Esterase (PLE). This procedure yields the tri-

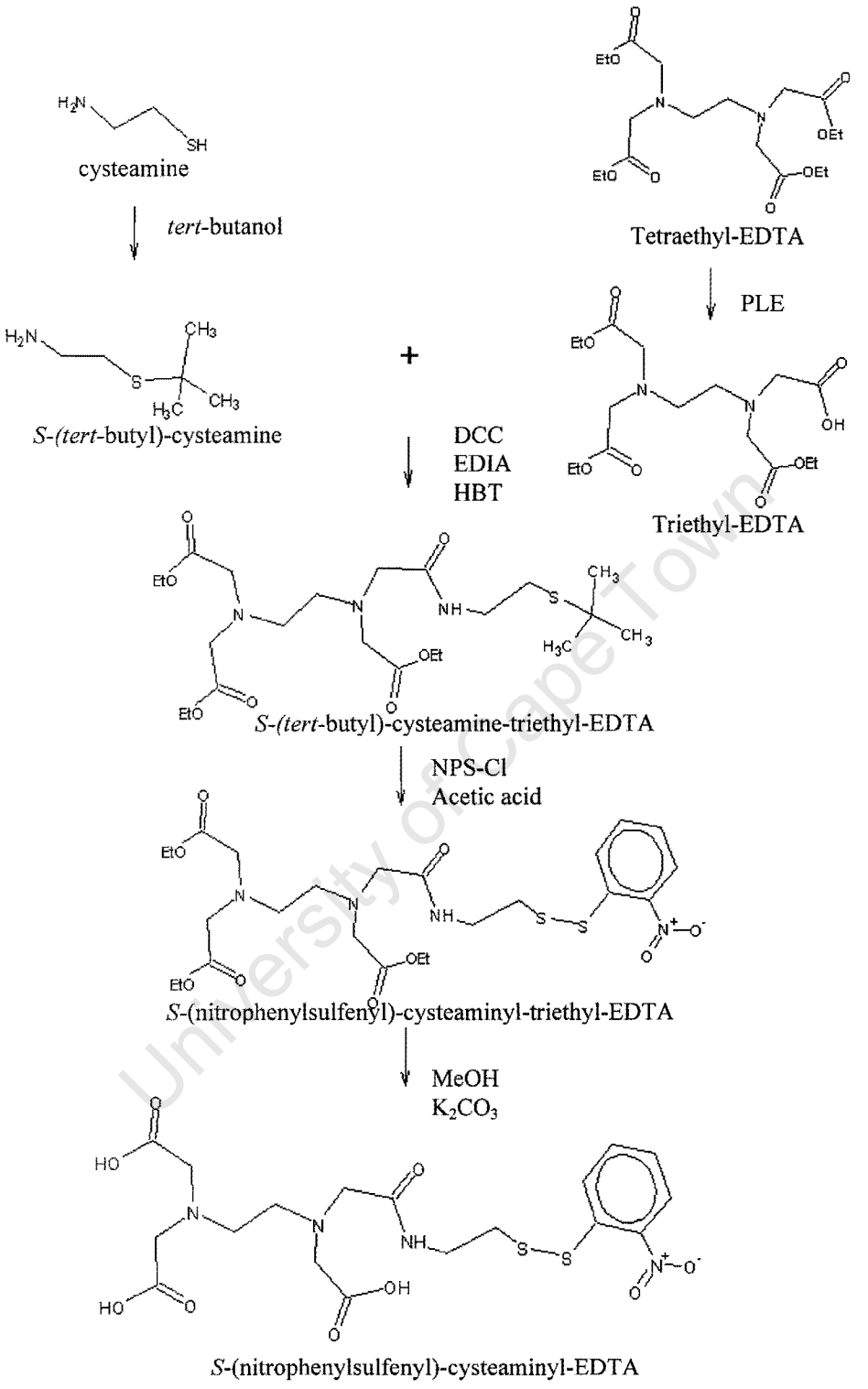


Figure 5.2. Overview of an improved synthesis scheme for *S*-(nitrophenylsulfenyl)-cysteaminyl-EDTA. See text for details.

ethyl ester of EDTA that is easily extracted from the reaction mixture with dichloromethane. Coupling between the resulting tri-ethyl ester of EDTA and *tert*-butyl protected cysteamine is carried out exactly as described by Flaus and Richmond (1999) without the need for HPLC purification. The exchange of the *tert*-butyl group for NPS is also carried out as described by Flaus and Richmond (1999). The product is subsequently extracted according to standard techniques, without the need for chromatography. The three remaining ethyl groups are readily removed from the EDTA-moiety by mild base hydrolysis, yielding *S*-(nitrophenylsulfenyl)-cysteaminyl-EDTA.

Both procedures are summarised in Figure 5.3. The two different routes of synthesis have three steps in common: protection of the cysteine thiol functionality with *tert*-butanol, linkage of an EDTA-moiety to *tert*-butylcysteamine and exchange of *tert*-butyl protection for NPS. The novel route to the synthesis of *S*-(nitrophenylsulphenyl)-cysteaminyl-EDTA involves two additional reactions. Firstly, in order to free only one carboxyl group for linkage to *S*-(*tert*-butyl)-cysteamine, a single ethyl group on tetraethyl-EDTA is selectively hydrolysed. Secondly, the remaining ethyl groups are removed following exchange of the *tert*-butyl group on the resulting *S*-(*tert*-butyl)-cysteaminyl-triethyl-EDTA with 2-nitrophenylsulfenyl (NPS). It should be noted that both these additional steps are easily executed, producing clean products in high yield. These procedures are presented separately below.

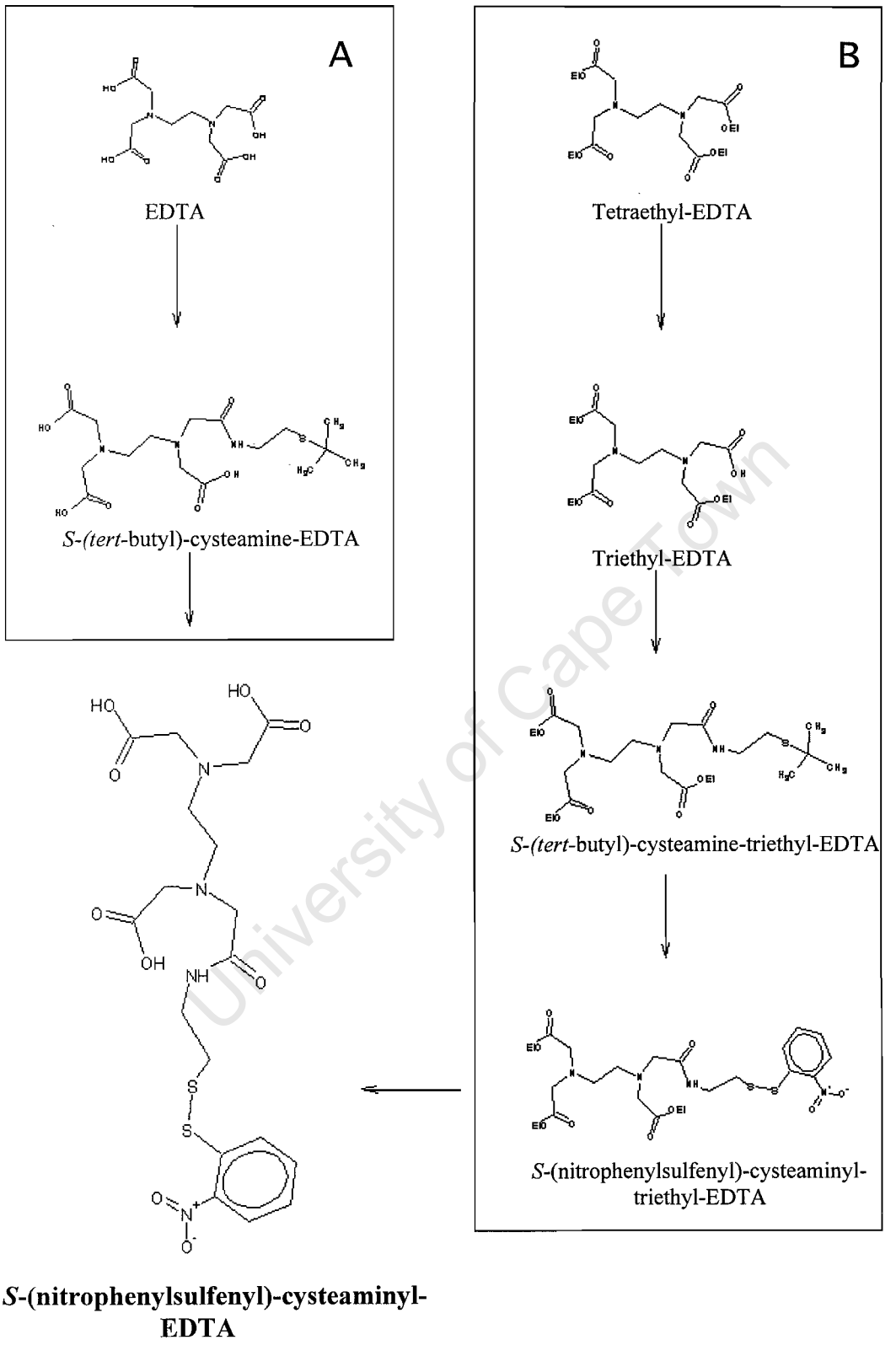


Figure 5.3. Overview of the synthesis routes to *S*-(nitrophenylsulfenyl)-cysteaminyl-EDTA. A. The method of Flaus and Richmond (1999). **B.** Novel method proposed in this thesis.

5.3. Protection of the cysteine thiol functionality with *tert*-butanol

The protection of the thiol group of cysteamine is essential to prevent reaction between the thiol group of cysteamine and one or more carboxyl groups of EDTA in subsequent reactions. This reaction between cysteamine-HCl and *tert*-butanol was performed essentially as described (Flaus and Richmond, 1999). Reagents were reacted by gentle reflux in acid medium (see Chapter 2), and recrystallised from the solvent. The purity of the recovered *S*-(*tert*-butyl)-cysteamine was determined by thin layer chromatography (TLC, Figure 5.4 A) using the solvent system described in Chapter 2. The plate was treated with cerium ammonium sulphate to stain the reaction product, which was found to be superior to iodine vapour. The structure of the product was confirmed by ^{1}H-NMR (data not shown).

5.4. The selective monohydrolysis of tetraethyl-EDTA

Due to the symmetry of EDTA, *S*-(*tert*-butyl)-cysteamine could attach to each of four carboxyl groups when these two molecules are coupled to synthesise *S*-(*tert*-butyl)-cysteamine-EDTA. We have made use of the unique ability of a crude ammonium sulphate precipitate of Pig Liver Esterase (PLE) to selectively hydrolyse a single ethyl group on the tetra-ethyl ester of EDTA (Burks *et al.,* 1998). This manipulation ensured that only one carboxyl group was available for subsequent reaction of the resulting tri-ethyl EDTA ester with *S*-(*tert*-butyl)-cysteamine, simplifying the purification of *S*-(*tert*-butyl)-cysteamine-triethyl-EDTA. The temperature reported for the biotransformation

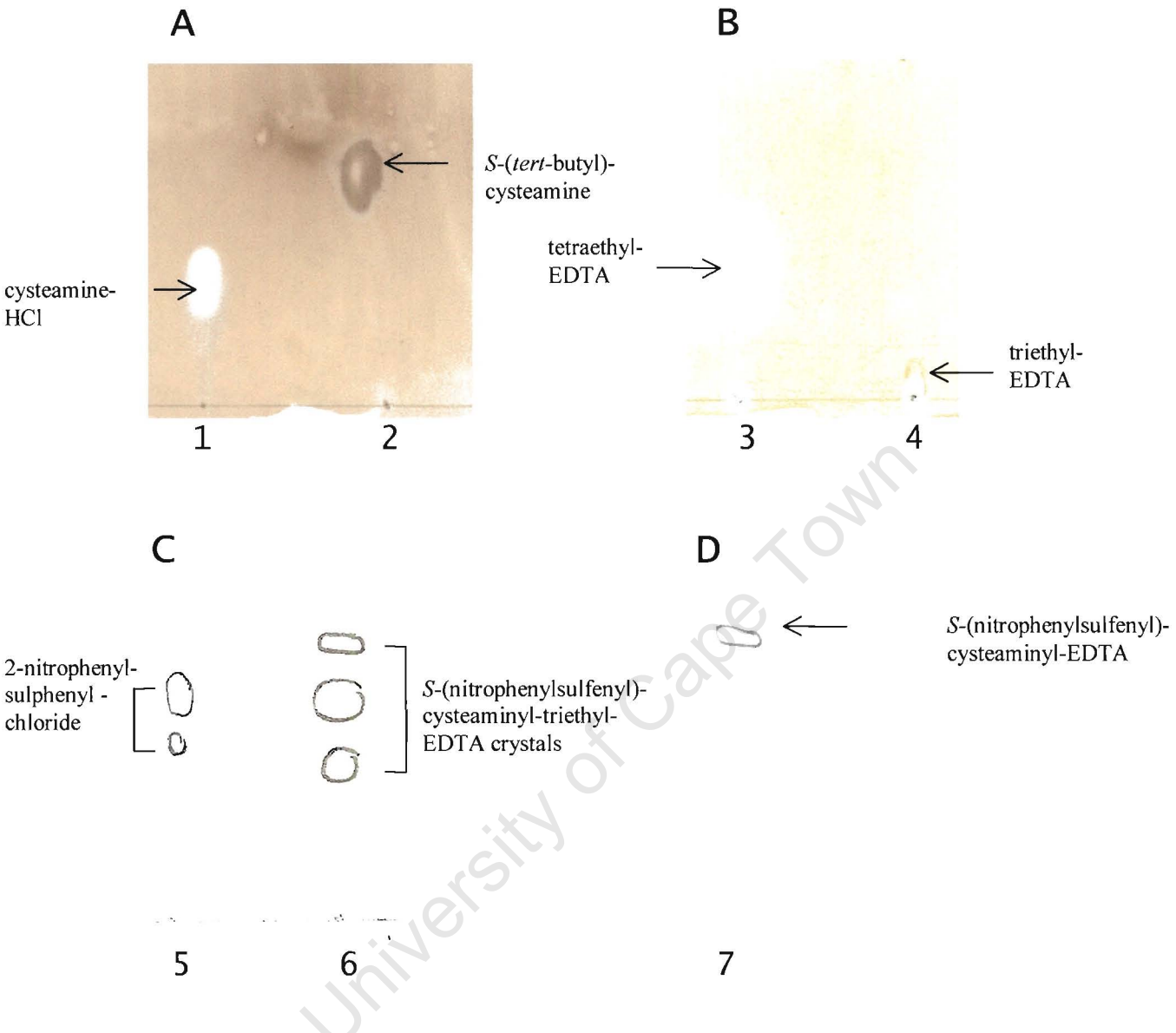


Figure 5.4. TLC analyses of reaction products. Silica TLC plates were stained with cerium ammonium sulphate (**A** and **B**) or analysed by UV (**C** and **D**). **A.** Cysteamine-HCl dissolved in eluent (*n*-butanol: acetic acid: water (4:1:1) was spotted at lane (**1**), while purified *S*-(*tert*-butyl)-cysteamine dissolved in the same eluent was spotted at lane (**2**). **B.** Tetraethyl-EDTA (**3**) and tri-ethyl EDTA (**4**) was spotted in eluent (50% ethylacetate, 50% petroleum ether). **C.** 2-nitrophenylsulphenyl chloride (**5**) and *S*-(nitrophenylsulfenyl)cysteaminyl-triethyl-EDTA (**6**) was spotted in eluent (50% ethylacetate, 50% petroleum ether). **D.** *S*-(nitrophenylsulfenyl)-cysteaminyl-EDTA (**7**) was spotted in eluent.

of tetra- to tri-ethyl ester by PLE is 27°C (Burks *et al.*, 1998). However, it was observed that the procedure could be performed successfully at 25°, which is the optimum temperature of the enzyme, or at 30°C, without significant difference in yield (data not shown). This observation points to the robust nature of this procedure. Practically, this eliminates the need for a shaking incubator with temperature regulator, since this procedure could be performed by stirring at room temperature or in a 30°C growth room.

Unreacted tetraethyl-EDTA was removed by washing with hexane, and triethyl-EDTA was extracted with dichloromethane. The purity of the product was confirmed by TLC (Figure 5.4 B), and the structure was confirmed by H^1-NMR, which showed a distinct multiplet at δ = 1.16 ppm assigned to the three methyl groups of the tri-ethyl ester function (see Chapter 2).

5.5. Linkage of *tert*-butylcysteamine to EDTA or triethyl-EDTA

The reaction of *tert*-butylcysteamine with EDTA or triethyl-EDTA was carried out under as dry conditions as possible, as water is a potent inhibitor of this reaction.

Following synthesis of *tert*-butylcysteamine-EDTA, the reaction products were separated by reverse phase HPLC. The resolution obtainable by HPLC chromatography is essential to separate the mono-, di- and tri-substituted reaction products. The purity of the desired mono-substituted product is of prime importance, since the chelating ability of the EDTA moiety depends on this. In an attempt to promote the formation of desired mono-

substituted EDTA products, EDTA was added in 5-fold molar excess to *S*-(*tert*-butyl)-cysteamine. Detection of *S*-(*tert*-butyl)-cysteamine is typically performed by measurement of the absorbance at 220 nm, the absorption maximum of the compound (Kerstin Bysticki, personal communication). However, it was observed that collection of chromatographically resolved, pure *S*-(*tert*-butyl)-cysteamine-EDTA was assisted by detection at 237 nm, since the elution profile immediately preceding *S*-(*tert*-butyl)-cysteamine-EDTA interfered less with the elution peak of the compound of interest (Figure 5.5). Purified fractions were analysed by mass spectrometry (Figure 5.6).

In Matrix-Assisted Laser Desorption Ionisation mass spectrometry (MALDI), a sample is embedded in a low molecular weight, UV absorbing matrix that enhances sample ionisation. The matrix chosen for these experiments was 2,5-dihydroxybenzoic acid (DHB). DHB is fairly insensitive to contamination by inorganic salts, buffers and detergents and is effective for the analysis of small molecules and peptides (200 to 1000 Da), because it produces only minimal amount of interference in the low molecular weight range (Strupat *et al.*, 1991). As shown in Figure 5.10, a mass corresponding to *S*-(*tert*-butyl)-cysteaminyl-EDTA (408.357 Da) was detected in the fraction that eluted from the HPLC column with a retention time of approximately 14 min. This retention time corresponds to a mobile phase composition of 13.02% (v/v) acetonitrile. Interestingly, a mass peak corresponding to *S*-(*tert*-butyl)-cysteamine-EDTA containing one chelated Fe atom (464.317 Da) was also detected. No extraneous Fe was added.

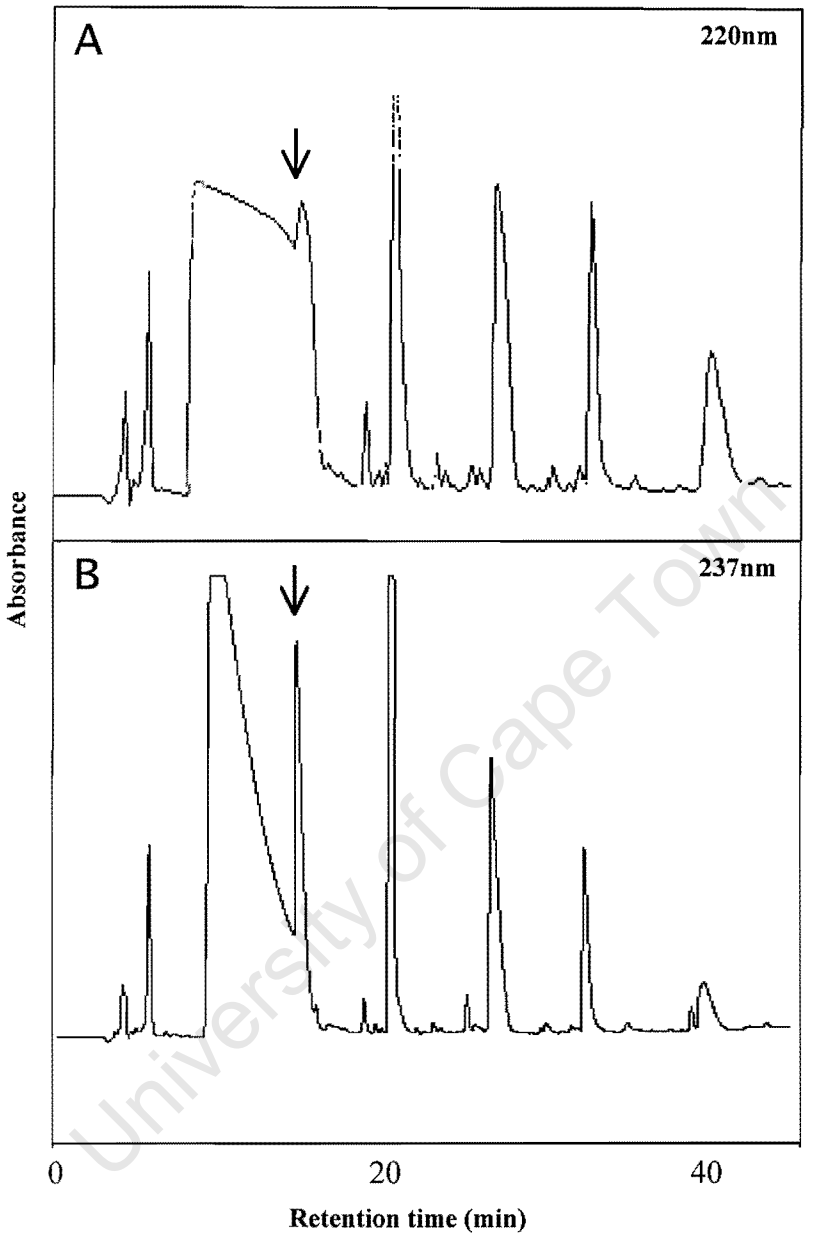


Figure 5.5. Chromatograms of the purification of *S*-(*tert*-butyl)-cysteamine-EDTA by reverse phase HPLC. The peak corresponding to *S*-(*tert*-butyl)-cysteamine-EDTA (retention time = 14 min) is indicated by an arrow. Detection in **A** was at 220 nm, while detection in **B** was at 237 nm. Although 220 nm is the absorption maximum of *S*-(*tert*-butyl)-cysteamine-EDTA, detection at 237 nm allows better visualisation of the elution of *S*-*tert*-butylcysteamine-EDTA from the column.

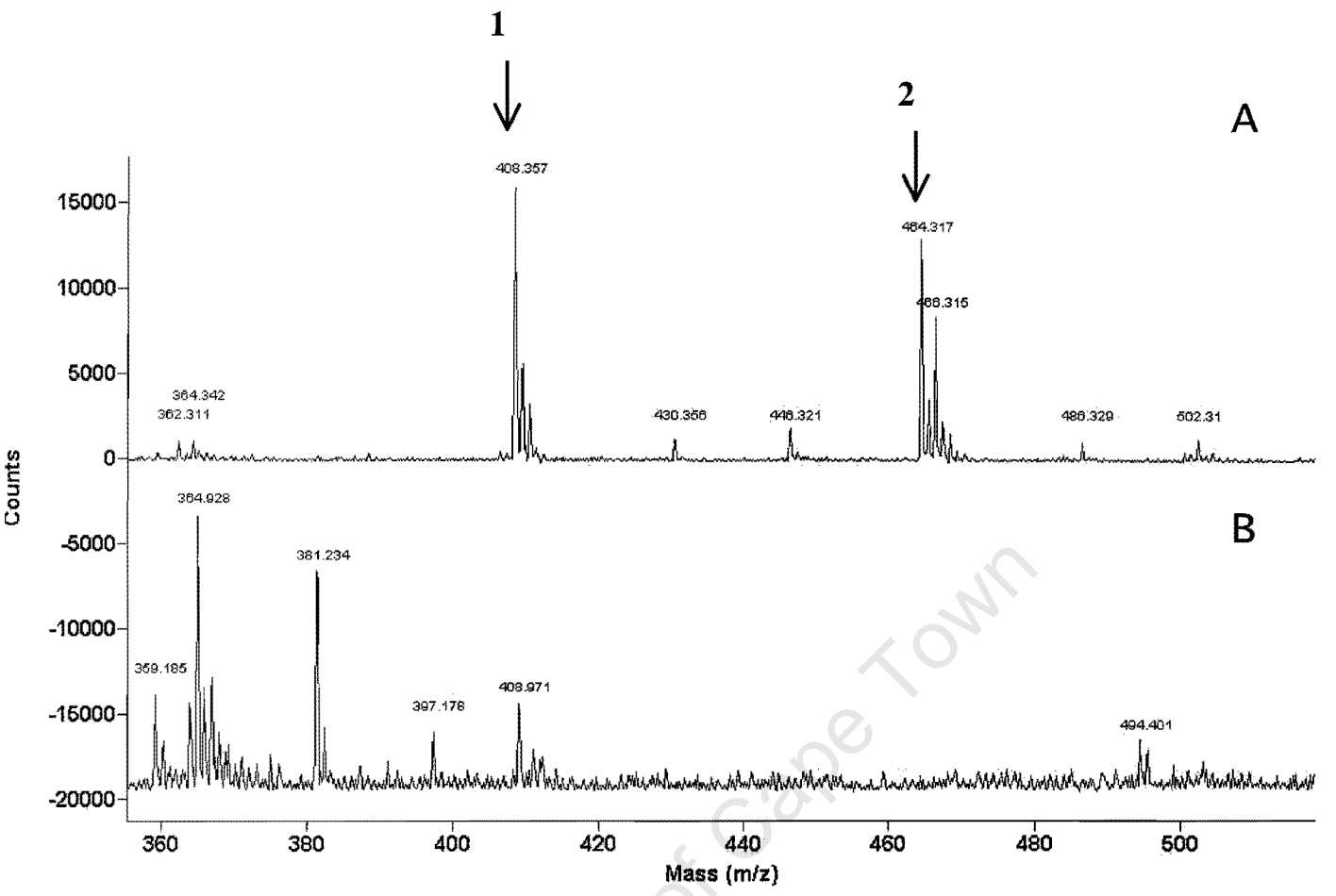


Figure 5.6. MALDI-TOF analysis of *S*-(*tert*-butyl)-cysteaminyl-EDTA. The compound was mixed with 2,5-dihydroxybenzoic acid and air dried to completion on a golden grid. The mass spectrum of purified *S*-(*tert*-butyl)-cysteaminyl-EDTA is shown in panel **A**, overlaid with the mass spectrum generated by 2,5-dihydroxybenzoic acid, shown in panel **B**. Arrow **1** indicates a 408.357 Da peak, not observed in the matrix, corresponding to the calculated mass of *S*-(*tert*-butyl)-cysteaminyl-EDTA (407.5 Da). Arrow **2** indicates 464.317 Da peak, corresponding to the calculated mass of *S*-(*tert*-butyl)-cysteaminyl-EDTA:Fe (463.3 Da). Counts are in arbitrary units.

The procedure detailed in this section was not required for the synthesis of *tert*-butylcysteaminyl-triethyl-EDTA, which forms part of the new synthesis scheme proposed in this chapter. The reaction mixture was filtered after completion of the reaction to remove any insoluble compounds, and *S*-(*tert*-butyl)-cysteaminyl-triethyl-EDTA was crystallised from (*N,N*-dimethylformamide), the reaction solvent. As can be seen from the TLC analysis in Figure 5.4 B, crystals were pure. The identity of the compound was confirmed by H^1-NMR. Thus a broad signal at δ = 2.6 ppm was assigned to the methylene protons alpha to the sulphur atom. Furthermore, the singlet at δ = 1.23 ppm is characteristic of the methyl-protons of a *tert*-butyl group.

5.6. Exchange of *tert*-butyl protection for NPS

2-nitrophenylsulphenyl (NPS) is an effective leaving group and facilitates the modification of cysteine with *S*-(nitrophenylsulfenyl)-cysteaminyl-EDTA, whereby the cysteaminyl-EDTA moiety is attached via disulphide linkage to the protein (Flaus and Richmond, 1999). The exchange of *tert*-butyl protection for NPS on both *S*-(*tert*-butyl)-cysteaminyl-triethyl-EDTA (novel route) and *S*-(*tert*-butyl)cysteaminyl-EDTA (Flaus and Richmond, 1999) is carried out in acid medium, yielding *S*-(nitrophenylsulphenyl)-cysteaminyl-triethyl-EDTA and *S*-(nitrophenylsulphenyl)-cysteaminyl-EDTA, respectively.

5.6.1. Exchange of *tert*-butyl protection for NPS on *tert*-butylcysteamine-EDTA

In the case of the synthesis of *S*-(nitrophenylsulphenyl)-cysteaminyl-EDTA (Flaus and Richmond, 1999), the reaction products were separated by reverse phase HPLC (Figure 5.7). The compound was detected by monitoring the absorption of eluting fractions at 277 nm. Although *S*-(nitrophenylsulphenyl)-cysteaminyl-EDTA seems to absorb much stronger in the 200 nm – 220 nm range, measuring absorption of the aromatic ring (277 nm) is essential for the detection of *S*-(nitrophenylsulphenyl)-cysteaminyl-EDTA. Candidate fractions were submitted to mass spectrometric analysis in order to confirm the presence of the purified compound (Figure 5.8). A mass peak corresponding to the double-deprotonated compound (502.87 Da) was detected. Two compounds with slightly higher molecular weights (524.451 Da and 552.603 Da, respectively) were also detected. These peaks seem to originate from 2,5-dihydroxybenzoic acid, which was used as matrix in the experiment (see Figure 5.8 B).

5.6.2. Exchange of *tert*-butyl protection for NPS on *S*-(*tert*-butyl)cysteamine-triethyl EDTA and hydrolysis of remaining ethyl groups

In order to facilitate the exchange of *tert*-butyl protection for NPS on *tert*-butylcysteamine-triethyl-EDTA, the compound was reacted with NPS-Cl under identical conditions as described for *tert*-butylcysteamine-EDTA (Flaus and Richmond, 1999).

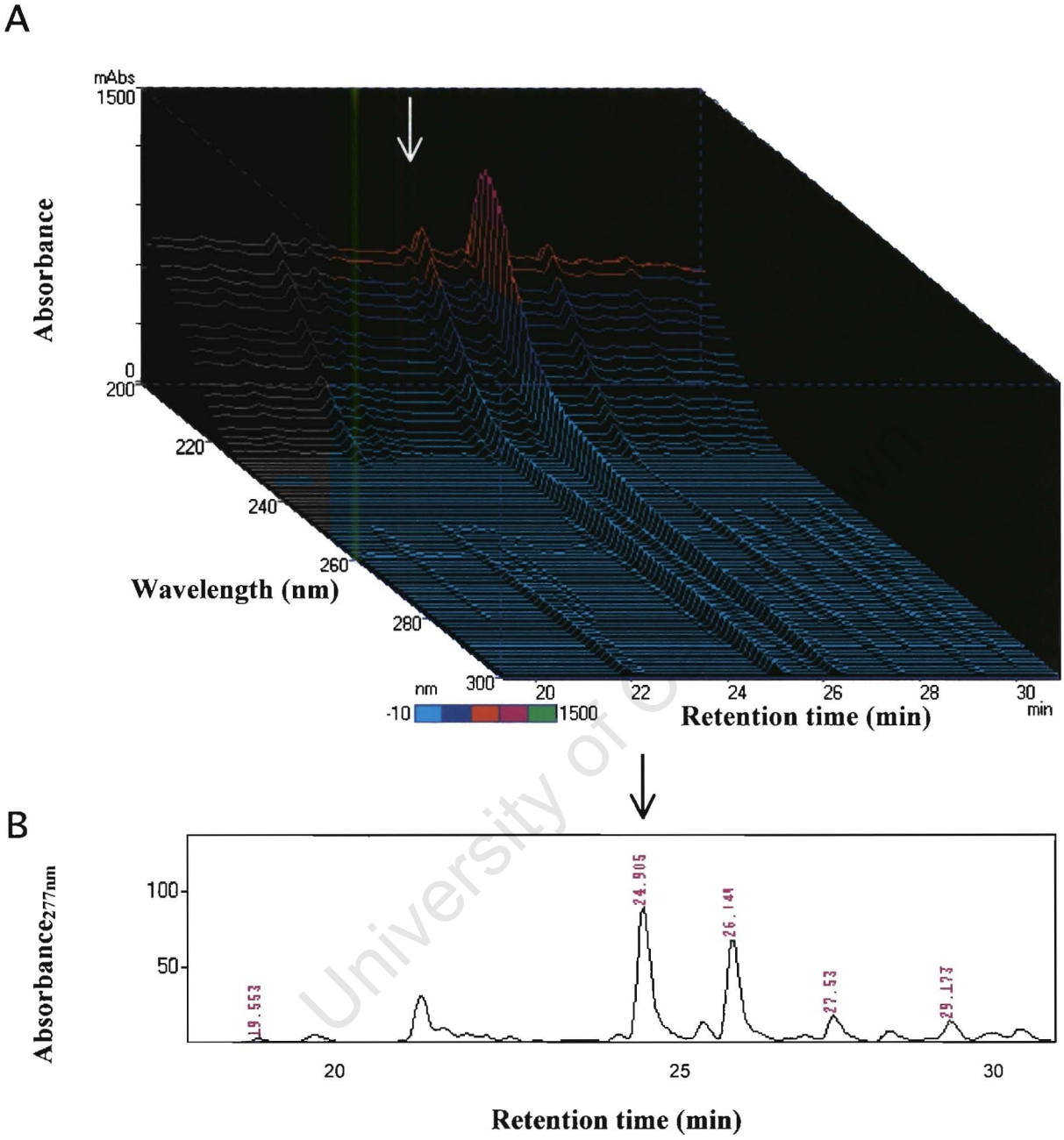


Figure 5.7. HPLC purification of *S*-(nitrophenylsulphenyl)-cysteaminyl-EDTA. Panel **A** shows a three-dimensional chromatograph of the separation products of the exchange reaction (see text). Panel **B** shows the chromatograph with absorption detected only at 277 nm. The peak corresponding to *S*-(nitrophenylsulphenyl)-cysteaminyl-EDTA (24.905 min) is indicated by an arrow. Absorbance is in arbitrary units.

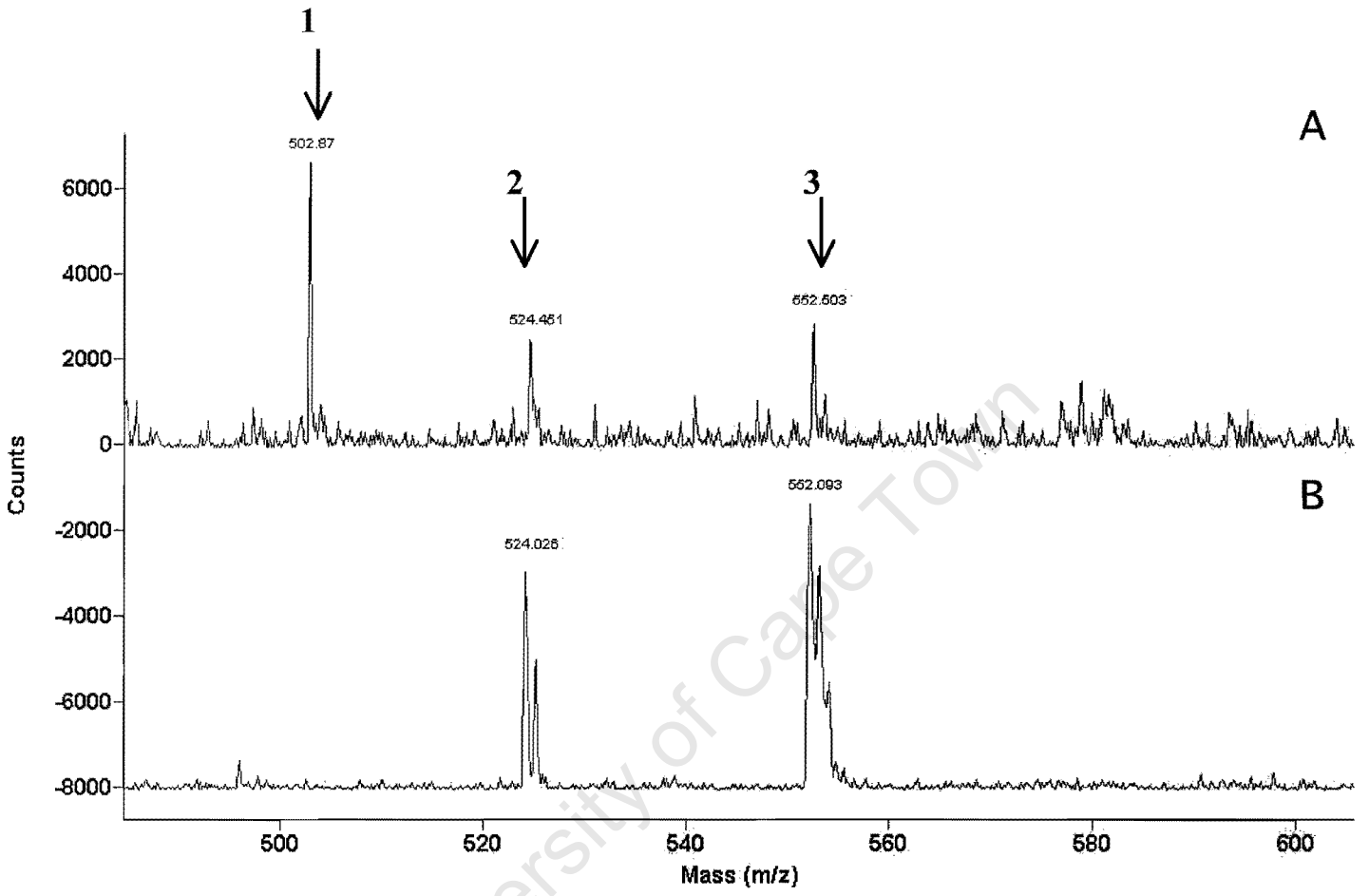


Figure 5.8. MALDI-TOF analysis of *S*-(nitrophenylsulfenyl)-cysteaminyl-EDTA. The compound was mixed with 2,5-dihydroxybenzoic acid and dried to completion on a golden grid. The mass spectrum of purified *S*-(nitrophenylsulfenyl-cysteaminyl-EDTA) is shown in panel **A**, overlaid with the mass spectrum generated by 2,5-dihydroxybenzoic acid, shown in panel **B**. Arrow **1** indicates a 502.87 Da peak, not observed in the matrix, corresponding to the calculated mass of twofold deprotonated *S*-(nitrophenylsulfenyl)-cysteaminyl-EDTA (505 Da). Arrows **2** and **3** indicate peaks also present in the 2,5-dihydroxybenzoic acid spectrum (**B**).

The product was extracted as described in Chapter 2, without use of liquid chromatography. Reaction products were analysed by TLC (Figure 5.4 C) and were found to contain contaminants. Impure *S*-(nitrophenylsulphenyl)-cysteaminyl-triethyl-EDTA crystals were dissolved in methanol and reacted with solid potassium carbonate for three hours at room temperature to hydrolyse the three remaining ethyl groups attached to the EDTA moiety. The product was extracted and dried as described (see Chapter 2) and was determined to be pure by TLC (Figure 5.4 D). Therefore, the purity of *S*-(nitrophenylsulphenyl)-cysteaminyl-triethyl-EDTA crystals is not a requirement for successful removal of ethyl groups by K_2CO_3. The structure of the compound was confirmed by ^{1}H-NMR and the compound was stable when stored dry at 4°C. The ^{1}H-NMR spectrum displayed all the requisite peaks. In particular the multiplet at δ = 2.9 ppm is characteristic of methylene protons alpha to a disulphide moiety while the peaks in the region δ = 7.4 - 8.4 ppm were assigned to the aromatic protons of the NPS-group.

5.7. Molar yields

The molar yields of the reactions described above are reported in Table 5.1. As can be seen from Table 5.1, similar yields were obtained for the synthesis of *S*-(*tert*-butyl)-cysteaminyl-EDTA (method of Flaus and Richmond (1999); 9%) and *S*-(*tert*-butyl)-cysteaminyl-triethyl-EDTA (this study; 8%). This finding indicates that the presence of ethyl groups does not significantly interfere with the DCC-mediated reaction. Since *S*-(nitrophenylsulphenyl)-cysteaminyl-triethyl-EDTA was not purified prior to reaction

with K_2CO_3, the yield reported for the synthesis of this compound, as well as for *S*-(nitrophenylsulphenyl)-cysteaminyl-EDTA were estimated. This was done by assuming a yield similar to the reaction signifying the exchange of the *tert*-butyl group of *S*-(*tert*-butyl)-cysteaminyl-EDTA for NPS (71%, see Table 5.1), since the ethyl groups on *S*-(*tert*-butyl)-cysteaminyl-triethyl-EDTA is not expected to influence this exchange. Although the overall yields of both methods are low (see Table 5.1), the method reported in this study has the advantage that no HPLC chromatography is required. Therefore, the process can be scaled up significantly to overcome the low yields.

Table 5.1. Molar yields obtained for the reactions reported in this study.

Reaction Product	*Molar yield*[1] *(%)*	
	Flaus and Richmond[2]	**This study**
S-(*tert*-butyl)-cysteamine	85	85
S-(*tert*-butyl)-cysteaminyl-EDTA	9	-
triethyl-EDTA	-	85
S-(*tert*-butyl)-cysteaminyl-triethyl-EDTA	-	8
S-(nitrophenylsulphenyl)-cysteaminyl-triethyl-EDTA	-	71*
S-(nitrophenylsulphenyl)-cysteaminyl-EDTA	71	83*
Overall yield	5	4*

1) Molar yields were calculated as moles of recovered product divided by moles of starting reagent

2) Flaus and Richmond (1999)

* *Estimated value, since S-(nitrophenylsulphenyl)-cysteaminyl-triethyl-EDTA was not purified prior to removal of ethyl groups*

CHAPTER 6: DISCUSSION

6.1. Introduction

We have generated a model for the binding of the globular domain of chicken linker histone H5 (GH5) in a nucleosome (see Chapter 3). It was pointed out that this model could be experimentally verified by site-directed metal-affinity cleavage experiments employing specific GH5 mutants. These mutants were designed and overexpressed in *E. coli* (Chapter 4). In addition, we have synthesised an EDTA derivative to modify GH5 mutants with the aim of mapping the secondary DNA-binding site of these mutants in a nucleosome (Chapter 5).

6.2. The position and orientation of GH5 on the nucleosome

6.2.1. Primary recognition of DNA by GH5

It is widely accepted that GH5 recognises linker DNA via helix III through the experimental data of Zhou *et al.* (1998). In our model (see Chapter 3), helix III of GH5 makes contacts with the major groove of DNA, with subsidiary recognition by the N-termini of helix I and helix II.

According to the model proposed in this study, His 62, Gln 67, Lys 69, Arg 73 and Arg 74 all contact DNA from the surface of helix III, while only one residue, Lys 85,

binds to DNA from the β-wing. The cross-linking data of Mirzabekov *et al.* (1990) as well as the mutagenesis studies of Goytisolo *et al.,* (1996) demonstrated that His 62 is important in DNA-binding. In addition, Buckle *et al.,* (1992) and Thomas and Wilson (1986) established the importance of Lys 69 in DNA binding. In the model proposed in this study, Lys 69 binds to DNA with a contact distance of 2.58 Å and is inaccessible in a hydrophilic environment. Clearly, this would prohibit reaction of this residue with an alkylating reagent, as observed by Thomas and Wilson (1986). Moreover, Lys 85, located in the β-wing motif, is protected from chemical modification when bound in chromatin (Thomas and Wilson, 1986). Although Lys 85 donates 2 strong H-bonds to the DNA-backbone in our model, it is solvent-accessible. However, the intimate association of this residue with DNA might affect the reactivity of its ε-amino group, thereby preventing its modification. Gajiwala *et al.* (2000) also reported 2 H-bonds between Lys 85 and the DNA-backbone in β-wing recognition. Based on its conservation in linker histones, Arg 73 was implicated in DNA-binding by Ramakrishnan *et al.* (1993). In the model discussed in chapter three, Arg 73 contacts the DNA backbone via three strong H-bonds, at angles close to 180° (see Table 3.3), while Arg 74 binds to the DNA-backbone via a single H-bond. Since both these residues are located in helix III, their conservation might reflect an essential role in primary DNA-recognition. Therefore, at the level of individual residues, our model of primary recognition by helix III is consistent with the experimental findings mentioned above.

Segers *et al.* (1991) proposed a model for GH5 interaction based on homology with other helix-turn-helix proteins where helix II recognises the major groove of DNA, while helix

III makes contacts with the minor groove, making contact with a full turn of DNA. In contrast, the crystal structure of hRFX-DBD in complex with its cognate X-box DNA (Gajiwala *et al.,* 2000) revealed a new mode of binding of a winged helix motif to DNA, where DNA is primarily recognised by the β-wing bound to the phosphodiester backbone. This work was published only after the modelling of the primary DNA-binding site of GH5 presented in this study was completed, and was therefore not considered, although it would have interesting consequences for GH5-binding. Gajiwala *et al.* (2000) base their argument for β-wing recognition by GH5 on the fact that this motif is more basic than helix III. However, in our model, three of the four basic residues in the β-wing (Lys 82, Lys 85, Arg 94 and Arg 97) are involved in DNA binding. Of these, only Lys 85 is associated with the primary DNA-binding site. Arg 94 and Arg 97 are both part of the secondary DNA binding site. Moreover, Thr 84 donates a H-bond to the DNA associated with the secondary DNA-binding site. Only Arg 94 has been experimentally shown to be involved in DNA-binding by Goytisolo *et al.* (1996). Gajiwala *et al.* (2000) modelled the binding of GH5 to a single DNA duplex and it is conceivable that in such a setting, most of the basic residues in the β-wing motif are in contact with DNA. However, our model opens the possibility that the β-wing plays an important role in the secondary DNA-binding site when GH5 binds to a nucleosome.

The orientation of the protein on the nucleosome as predicted by cross-linking (Mirzabekov *et al.,* 1990) is inconsistent with DNA-recognition by helix III. According to this study, His 25 interacts with the terminal regions of nucleosomal DNA, whereas His 62 is associated with more distal segments of DNA. In our model, His 25 interacts

with the terminus of the chromatosomal DNA, while His 62 interacts with DNA 2 bp away and proximal to the entry/exit point of the nucleosome. It is therefore possible that the protein binds to a nucleosome in the orientation demonstrated by Mirzabekov *et al.* (1990) by β-wing recognition. The possibility that β-wing recognition might have an effect on the placement of the secondary DNA-binding site also cannot be excluded. Clearly, the question of whether GH5 recognises linker DNA mainly via the β-wing motif or helix III remains unresolved, and high-resolution mapping of the primary DNA-binding site is required in order to resolve this dispute.

6.2.2. The secondary DNA-binding site of GH5

Although Crane-Robinson and Ptitsyn (1989) argued for more than one DNA-binding site on the face of GH5, the crystal structure of this protein (Ramakrishnan *et al.,* 1993) suggested that only two sites existed. It could be argued that the primary DNA-binding site consists of two sub-sites, constituted by helix III and the β-wing, respectively. However, there would be no functional differentiation between such sites.

Goytisolo *et al.,* (1996) abolished the charge on a putative secondary DNA-binding site of GH5 comprising four residues, Lys 40, Arg 42, Lys 52 and Arg 94. This resulted in the loss of binding to synthetic four-way junctions, abolished the formation of cooperative tramline complexes of juxtaposed DNA duplexes and abolished the canonical 166 bp protection against MNase digestion, suggesting the importance of these residues in conferring the ability of GH5 to bind to two DNA-duplexes. The model proposed in

Chapter 3 successfully explains the involvement of Lys 40, Arg 42 and Arg 94. Contrary to this model, Ramakrishnan *et al.* (1993) and Goytisolo *et al.* (1996) assigned Lys 52 to the secondary DNA-binding site of GH5. However, Ramakrishnan *et al.* (1993) only speculated on the role of Lys 52 being part of a basic cluster of residues now thought to be the secondary DNA-binding site. Goytisolo *et al.* (1996), replaced GH5 residues with alanine or glutamic acid. This yielded a net charge at the primary DNA-binding site between +3 and 0, and between +4 and –2 at the secondary DNA-binding site (to which Lys 52 had been assigned). Goytisolo *et al.* (1996) showed that this neutralization of either basic cluster either impaired or abolished the ability of GH5 to form cooperative tramline complexes with DNA. However, the formation of a residual tram-line complex was observed when the primary cluster alone was mutated, suggesting that His 25 and His 62, not mutated in the primary cluster (Goytisolo *et al.*, 1996), conferred this ability. Because the model discussed here assigns Lys 52 to the primary DNA-binding site and Lys 52 was not mutated as part of the primary cluster by Goytisolo *et al.* (1996), it is likely that the residual DNA-binding observed might also be attributed to the interaction of Lys 52 with the DNA backbone. The fact that neither His 25 nor His 62 is conserved among linker histones, and that Lys 52 is highly conserved, supports this view.

The secondary DNA-binding site proposed in this study, is of a diffuse nature and involves several unrelated secondary structural features of the protein. Recently, Duggan and Thomas (2000) showed that Lys 40, Arg 42, Lys 52 and Arg 94 act as a cluster in binding to the nucleosome and that individual mutations in either of these residues are not sufficient in abolishing DNA-binding. The model proposed in this study is consistent

with these observations, although Lys 52 is predicted to be part of the secondary DNA-binding site. Indeed, this assignment correlates with the data of Duggan and Thomas (2000), since a mutation of a primary site residue is not expected to affect DNA-binding at the secondary site.

This study also revealed the possible involvement of non-conserved residues in secondary DNA-binding by GH5. These residues include Gly 43, which binds to the same DNA backbone phosphate as the conserved Gly 44. Thr 84 donates a H-bond to the DNA backbone and although not conserved, occurs more than once at this position in aligned sequences of linker histones. Lastly, Arg 37 also binds to the DNA-backbone. The additional interactions of these residues, not conserved in the H1 family, may explain why GH5 binds to four-way junction DNA with higher affinity compared to GH1 (Varga-Weiss *et al.*, 1994).

6.2.3. Orientation of GH5 in the nucleosome

The study of Zhou *et al.* (1998) employed *N*-azidophenacyl cross-linking of specific Ser to Cys substituted GH5 mutants to investigate the location of GH5 on nucleosomes isolated from bulk chromatin. The results suggested an orientation where helix II of the globular domain faces the interior of the chromatosome, placing helix I on the outside. A location of the putative secondary DNA-binding site near the dyad axis of symmetry was also demonstrated. The model proposed in this study is consistent with this setting. It should however be noted that the resulting location of two residues, Ser 29 and Ser 71

does not correlate with the experimental findings of Zhou *et al.* (1998), since these residues, when replaced with cysteine, are not close enough to the DNA to be cross-linked by *N*-azidophenacyl. Although a rotation of GH5 in the major groove away from the dyad axis would resolve this matter, such a rotation would result in a clash between the extended loop connecting helices I and II and nucleosomal DNA. This discrepancy will be addressed by mapping the location of GH5 on the nucleosome using mutant S41C, of which the expression and purification was described in Chapter 5.

6.2.4. The position of GH5 relative to the core histone tails

The binding site of GH5 to the nucleosome was determined and modelled using the nucleosomal DNA only. The role of the core histone N-terminal tails in chromatin condensation has been demonstrated (Allan *et al.,* 1982; de la Barre *et al.,* 2000). By reconstructing the core histones from the crystal structure of the nucleosome (Luger *et al.,* 1997), a model of the GH5 chromatosome was generated. This model was examined for any possible interactions between GH5 and parts of the core histone tails that are observed in the crystal structure of Luger *et al.* (1997). Because the globular domain of the linker histone is seen to be located far from the exit points of the core histone tails in this model, it seems unlikely that there is direct cooperativity between the linker histone and the core histone tails in chromatin compaction. However, it is known that the tail of H2A competes for a binding site at the nucleosomal DNA in the position of the dyad (Lee and Hayes, 1998). This suggests that the structure of the segments of core histone tails

that have been solved in the crystal structure of Luger *et al.* (1997) might not present sufficient information to sustain arguments of this nature regarding the core histone tails.

6.2.5. Asymmetry of the chromatosome

The length of the nucleosomal DNA in the crystal structure of the nucleosome (Luger *et al.*, 1997) is 146 bp. It has been shown that GH5 confers protection to an additional 20 bp of nucleosomal DNA from micrococcal nuclease (MNase) digestion (Simpson *et al.*, 1978). Because of the bulkiness of the MNase enzyme, this is consistent with the length of DNA contacted by GH5 being shorter than 20 bp. The DNA in the primary DNA-binding site model reported here could safely be truncated to 13 bp without significant loss of GH5-DNA interactions, consistent with an asymmetry in the extension of nucleosomal DNA-protection upon linker histone binding.

6.3. Overexpression of GH5 in *E. coli*

Several factors could affect the stable expression of a eukaryotic protein in a prokaryotic system. One of these are differences in the frequencies of codon usage for a particular amino acid, in which case it is possible that low levels of the heterologous protein accumulate in response to tRNA depletion (Zahn, 1996). These differences can also affect cell growth. The AGG codon for Arg is the rarest codon in *E. coli.* Moreover, frame-shifting was observed in proteins expressed under conditions of $tRNA^{Lys}$ depletion (Gallant and Lindsley, 1992). Since the GH5 coding sequence is rich in both rare Arg

(AGG) and Lys codons, it is possible that ribosomes stall on mRNA transcripts of GH5 as a result of depletion of the tRNA molecules charged with both these codons.

Gerchman *et al.* (1993) investigated the expression of linker histones H1 and H5 in *E. coli* and proposed several reasons for failure of expression of GH5 in a prokaryotic expression system. The authors suggested that the expression of GH5 constructs in *E. coli* increases as the GC content of the first 18 bp downstream of the ATG start decreases. Although not explicitly mentioned by the authors, it is clear from an analysis of their data that a GC content below 56% seems to be required for the expression of GH5 constructs in *E. coli.* Indeed, the first 18 bp of the coding sequence for the globular domain of H5 corresponding to residues 20 - 109 exhibits a GC-content of 72%, while this value is 55% for the domain (spanning residues 23 - 106) preceded by the coding sequence for an MTE-motif. Therefore, our results are consistent with the observation of Gerchman *et al.* (1993). In addition, it was postulated that this might be due to complementarity of this sequence to the region corresponding to position 1469 – 1483 of the 16S RNA sequence of *E. coli* (Gerchman *et al.,* 1993). Indeed, we observed 60% complementarity to this area of the 16S rRNA sequence for the domain preceded by an MTE-motif, while the broader domain lacking this motif showed only 40% complementarity. Although there is no obvious rationale for this observation, it is possible that complementarity to this area of the 16S rRNA sequence performs a similar role as complementarity to the Shine-Dalgarno sequence located upstream of the start codon.

The C-terminal location of Cys 96 in mutant A96C suggests that the incorporation of a cysteine at this position is not likely to affect the folding of the protein. However, this manipulation renders a highly reactive thiol group accessible and may stimulate cross-reaction with other peptides spontaneously, or via bacterial disulphide-isomerase activity. However, this is unlikely to affect the expression of these constructs, since the equally accessible cysteine residues of mutants S41C and R37C do not seem to affect the expression of these proteins. In addition, since the coding sequence for the expressed globular domain of GH5 (MTE-GH5) and that for mutant A96C differ at only two base pairs, the fact that this mutant cannot be expressed in *E. coli* is not likely to be explained at the level of DNA sequence alone. However, it is possible that the Ala to Cys mutation might well affect the tertiary structure of the protein in such a way that the protein is targeted for degradation by bacterial proteases. It is therefore not clear why mutant A96C cannot be expressed in *E. coli.*

6.4. Synthesis of *S*-(nitrophenylsulphenyl)-cysteaminyl-EDTA

In order to use GH5 mutants R37C and S41C in site-directed metal affinity cleavage experiments, they should first be modified with the chelating agent *S*-(nitrophenylsulphenyl)-cysteaminyl-EDTA. The successful synthesis of this compound was reported by Flaus and Richmond (1999). However, we identified certain drawbacks in the synthesis procedure. The availability of preparative scale HPLC is required, since mono-substituted EDTA should be separated from additional EDTA substitutes as well as polar reaction components. Only milligram amounts of mono-substituted product can be

purified in this way per HPLC run. We therefore eliminated the need for an expensive HPLC purification by coupling the tri-ethyl ester of EDTA instead of EDTA to *tert*-butyl-cysteaminyl. The tri-ethyl ester can be enzymatically synthesised and purified in one day using a crude ammonium sulphate precipitate of Pig Liver Esterase (PLE). The resulting coupled product, *S*-(*tert*-butyl)-cysteaminyl-triethyl-EDTA, is nonpolar and can be crystallised in pure form from *N,N*-dimethylformamide, the reaction solvent, eliminating the need for purification by chromatography. Therefore, the procedure can be scaled up, allowing the concurrent synthesis of large quantities of the product. Although modification of mutant peptides requires small amounts of *S*-(nitrophenylsulphenyl)-cysteaminyl-EDTA (50 mM; Flaus and Richmond, 1999), the compound is unstable in solution and should be prepared fresh daily.

In addition to the new synthesis scheme proposed in Chapter 4, we have introduced minor alterations to the procedure of Flaus and Richmond (1999). Firstly, instead of using ninhydrin or iodine vapour (Kerstin Bysticki, personal communication) in the detection of products following TLC, we found that cerium ammonium sulphate was more effective. Ninhydrin is toxic and iodine vapour constitutes a temporary stain, which sublimes with time. Staining with cerium ammonium sulphate circumvents both these difficulties. Finally, the detection of *S*-(*tert*-butyl)-cysteaminyl-EDTA eluting from the HPLC column is typically performed at 220 nm, since this is the absorbance maximum for the compound (Kerstin Bysticki, personal communication). However, we observed that this compound could be separated with higher purity by detection at 237 nm.

6.5. Future work

We plan to determine the location of GH5 on the nucleosome by site-directed DNA cleavage. The nucleosomal location of a specific engineered cysteine residue can be mapped at single-base pair resolution by this procedure, which involves three steps. Firstly, mutant proteins into which individual cysteine residues were incorporated are expressed and purified. Next, mutant proteins are modified with *S*-(nitrophenylsulphenyl)-cysteaminyl-EDTA, a chelating adduct. Proteins modified in this way are charged with Fe^{2+} and reconstituted into chromatosomes. Lastly, the position of mutant cysteine residues is determined by regulated strand scission as a result of hydroxyl radicals released from protein-chelated Fe^{2+}.

Since the exact location of the secondary DNA-binding site of GH5 to the nucleosome remains unclear, we have targeted three residues predicted to be associated with the secondary DNA-binding site of GH5, Arg 37, Ser 41 and Ala 96, for future investigation. Two of these mutants, R37C and S41C, were stably expressed in *E. coli.*

In order to use these mutants in site-directed metal-affinity cleavage experiments, it is essential to determine whether the mutants bind to nucleosomes in the same way as the native peptide. We therefore propose to investigate the folding of mutants by comparing their circular dichroism spectra to those of both recombinant GH5 and tryptic GH5. A 166 bp kinetic pause is observed when chromatosomes containing properly bound GH5 are digested with MNase (Simpson, 1978). Therefore, this feature can be used to

determine whether mutation and modification of GH5 affect proper binding to the nucleosome.

To facilitate their purification, both native and mutant recombinant GH5 peptides were expressed in-frame with 6 C-terminal His residues. It is conceivable that these residues might slightly alter the binding of GH5 to the nucleosome. Our model places the C-terminal Lys 96 in close contact with DNA (see Chapter 3), suggesting that the 6 × His tract might be in close proximity to DNA. Therefore, it is important that the effect of these residues be addressed. This question could be investigated by comparing the UV-laser induced cross-links of both native and His-tagged GH5 bound to identical nucleosomes. In addition, it is known that His tracts chelate divalent metal ions and might therefore generate a background signal in metal affinity cleavage experiments. However, this signal would be identical to the signal generated by native His-tagged GH5 and could be subtracted by performing a control metal affinity-cleavage experiment using this protein.

As mentioned in Section 1.7.7.3., there is a discrepancy between the location of GH5 on the 5S rDNA sequence from *Xenopus* and its location in bulk chromatin. In order to address this, we plan to map the location of GH5 on nucleosomes reconstituted on the 5S rDNA of *Lytechinus variegatis* and a histone gene spacer from *Psammechinus miliaris* by site-directed metal-affinity cleavage. Although not included in this dissertation, the efforts discussed in this section are currently underway.

REFERENCES

Allan, J., Harborne, N., Rau, D.C., Gould, H. (1982). Participation of core histone "tails" in the stabilization of the chromatin solenoid. J. Cell. Biol. *93*, 285-97.

Allan, J., Mitchell, T., Harborne, N., Bohm, L., Crane-Robinson, C. (1986). Roles of H1 domains in determining higher order chromatin structure and H1 location. J. Mol. Biol. *187*, 591-601.

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

An, W., Van Holde, K., Zlatanova, J. (1998). Linker histone protection of chromatosomes reconstituted on 5S rDNA from *Xenopus borealis*: a reinvestigation. Nucl. Acids Res. *26*, 4042-4046.

An, W., Zlatanova, J., Leuba, S.H., van Holde, K. (1999). The site of binding of linker histone to the nucleosome does not depend upon the amino termini of core histones. Biochimie *81*, 727-732.

Ausubel, F., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (1995). Short protocols in Molecular Biology, 3rd Edition (New York: John Wiley & Sons Inc.).

Aviles, F. J., Chapman, G.E., Kneale, G.G., Crane-Robinson, C., Bradbury, E.M. (1978). The conformation of histone H5. Isolation and characterisation of the globular segment. Eur. J. Biochem. *88*, 363-371.

Banchev, T. B., Srebreva, L.N., Zlatanova, J.S. (1990). Accessibility of histone H1(0) and its structural domains to antibody binding in extended and folded chromatin. Mol. Cell. Biochem. *95*, 167-175.

Baneyx, F. (1999). Recombinant protein expression in *Escherichia coli*. Curr. Opin. Biotechnol. *10*, 411-21.

Barnes, W. M. (1992). The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion. Gene *112*, 29-35.

Bates, D. L., Thomas, J.O. (1981). Histones H1 and H5: one or two molecules per nucleosome? Nucl. Acids Res. *9*, 5883-5895.

Beavis, R. C., Chaudhary, T. , Chait, B. T. (1992). α-Cyano-4-hydroxycinnamic acid as a matrix for matrix-assisted laser desorption mass spectrometry. Org. Mass. Spectrom. *27*, 156-158.

Belikov, S., Karpov, V. (1998). Linker histones: paradigm lost but questions remain. FEBS Letters *441*, 161-164.

Belikov, S., Karpov, V. (1998). Localization of histone binding sites within the nucleosome by UV-induced H1-crosslinking *in vivo*. Journal of Biolmolecular Structure and Dynamics *16*, 35-39.

Boulikas, T., Wiseman, J.M., Garrard, W.T. (1980). Points of contact between histone H1 and the histone octamer. Proc. Natl. Acad. Sci. USA *77*, 127-131.

Bradbury, E. M., Inglis, R.J., Matthews, H.R., Sarner, N. (1973). Phosphorylation of very-lysine-rich histone in *Physarum polycephalum*. Correlation with chromosome condensation. Eur. J. Biochem. *33*, 131-139.

Bucci, L. R., Brock, W.A., Meistrich, M.L. (1982). Distribution and synthesis of histone 1 subfractions during spermatogenesis in the rat. Exp. Cell. Res. *140*, 111-118.

Buckle, R. S., Maman, J.D. and Allan, J. (1992). Site-directed mutagenesis on the binding of the globular domain of linker histone H5 to the nucleosome. J. Mol. Biol. *223*, 651-659.

Burks, E., Koshti, N., Jacobs, H., Gopalan, A. (1998). Selective monohydrolysis of esters of polyaminocarboxylic acids using Pig Liver Esterase. Synlett., 1285-1287.

Butler, P. J., Thomas, J.O. (1998). Dinucleosomes show compaction by ionic strength, consistent with bending of linker DNA. J. Mol. Biol. *281*, 401-407.

Carruthers, L. M., Hansen, J.C. (2000). The core histone N termini function independently of linker histones during chromatin condensation. J. Biol. Chem. *275*, 37285-37290.

Carter, G. J., van Holde, K. (1998). Self-association of linker histone H5 and of its globular domain: evidence for specific self-contacts. Biochemistry *37*, 12477-12488.

Cerf, C., Lippens, G., Ramakrishnan, V., Muyldermans, S., Segers, A., Wyns,L., Wodak, S.J., Hallenga, K. (1994). Homo- and heteronuclear two-dimensional NMR studies of the globular domain of histone H1: full assignment, tertiary structure, and comparison with the globular domain of histone H5. Biochemistry *33*, 11079-11086.

Cirillo, L. A., McPherson, C.E., Bossard, P., Stevens, K., Cherian, S., Shim, E.Y., Clark, K.L., Burley, S.K., Zaret, K.S. (1998). Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. EMBO J. *17*, 244-254.

Clark, D. J., Hill, C.S., Martin, S.R., Thomas, J.O. (1988). Alpha-helix in the carboxy-terminal domains of histones H1 and H5. EMBO. J. *7*, 69-75.

Clark, D. J., Thomas, J.O. (1986). Salt-dependent co-operative interaction of histone H1 with linear DNA. J. Mol. Biol. *187*, 569-80.

Clark, J. M. (1988). Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. Nucl. Acids Res. *16*, 9677-9686.

Clark, K. L., Halay, E.D., Lai, E., Burley, S.K. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature *364*, 412-420.

Cline, J., Braman, J., Hogrefe, H.H. (1996). PCR fidelity of *Pfu* DNA polymerase and other thermostable DNA polymerases. Nucl. Acids Res. *24*, 3546-3551.

Clore, G. M., Gronenborn, A.M., Nilges, M., Sukumaran, D.K., Zarbock, J. (1987). The polypeptide fold of the globular domain of histone H5 in solution. A study using nuclear magnetic resonance, distance geometry and restrained molecular dynamics. EMBO J. *6*, 1833-1842.

Cornell, W. D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M. Jr., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., Kollman, P.A. (1995). A second generation force field for the simulation of proteins and nucleic acids. J. Am. Chem. Soc. *117*, 5179-5197.

Crane-Robinson, C., Ptitsyn, O.B. (1989). Binding of the globular domain of linker histone H5/H1 to the nucleosome: a hypothesis. Protein Engineering *2*, 577-582.

Crane-Robinson, C. (1997). Where is the globular domain of the linker histone located on the nucleosome? TIBS, 75-77.

D'Anna, J. A., Thayer, M.M., Tobey, R.A., Gurley, L.R. (1985). G1- and S-phase syntheses of histones H1 and H1o in mitotically selected CHO cells: utilization of high-performance liquid chromatography. Biochemistry *24*, 2005-2010.

Drabent, B., Saftig, P., Bode, C., Doenecke, D. (2000). Spermatogenesis proceeds normally in mice without linker histone H1t. Histochem. Cell. Biol. *113*, 433-442.

Draves, P. H., Lowary, P.T., Widom, J. (1992). Co-operative binding of the globular domain of histone H5 to DNA. J. Mol. Biol. *225*, 1105-1121.

Duggan, M. M., Thomas, J.O. (2000). Two DNA-binding sites on the globular domain of histone H5 are required for binding to both bulk and 5 S reconstituted nucleosomes. J. Mol. Biol. *304*, 21-33.

Ebright, Y. W., Chen, Y., Pendergrast, P.S., Ebright, R.H. (1992). Incorporation of an EDTA-metal complex at a rationally selected site within a protein: application to EDTA-iron DNA affinity cleaving with catabolite gene activator protein (CAP) and Cro. Biochemistry *31*, 10664-10670.

Ebright, Y. W., Chen, Y., Ludescher, R.D., Ebright, R.H. (1993). N-(iodoacetyl)-p-phenylenediamine-EDTA: a reagent for high-efficiency incorporation of an EDTA-metal complex at a rationally selected site within a protein. Bioconjug. Chem. *4*, 219-225.

Eckert, K. A., Kunkel, T.A. (1991). DNA polymerase fidelity and the polymerase chain reaction. PCR Methods Appl. *1*, 17-24.

Ermacora, M. R., Delfino, J.M., Cuenoud, B., Schepartz, A., Fox, R.O. (1992). Conformation-dependent cleavage of staphylococcal nuclease with a disulfide-linked iron chelate. Proc. Natl. Acad. Sci. USA. *89*, 6383-6387.

Finch, J. T., Klug, A. (1976). Solenoidal model for superstructure in chromatin. Proc. Natl. Acad. Sci. USA *73*, 1897-901.

Finch, J. T., Lutter, L.C., Rhodes, D., Brown, R.S., Rushton, B., Levitt, M., Klug, A. (1977). Structure of nucleosome core particles of chromatin. Nature *269*, 29-36.

Flaus, A., Richmond, T.J. (1999). Base-pair resolution mapping of nucleosome positions using site-directed hydroxy radicals. Methods Enzymol. *304*, 251-63.

Flaus, A., Luger, K., Tan, S., Richmond, T.J. (1996). Mapping nucleosome position at single base-pair resolution by using site-directed hydroxyl radicals. Proc. Natl. Acad. Sci. USA. *93*, 1370-1375.

Flenniken, A. M., Newrock, K.M. (1987). H1 histone subtypes and subtype synthesis switches of normal and delobed embryos of *Ilyanassa obsoleta.* Dev. Biol. *124*, 457-68.

Franke, K., Drabent, B., Doenecke, D. (1998). Expression of murine H1 histone genes during postnatal development. Biochim. Biophys. Acta *1398*, 232-242.

Furrer, P., Bednar, J., Dubochet, J., Hamiche, A., Prunell, A. (1995). DNA at the entry-exit of the nucleosome observed by cryoelectron microscopy. J. Struct. Biol. *114*, 177-83.

Gajiwala, K., Chen, H., Cornille, F., Roques, B.P., Reith, W., Mach, B., Burley, S.K. (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. Nature *403*, 916-921.

Gallant, J. A., Lindsley, D. (1992). Leftward ribosome frameshifting at a hungry codon. J. Mol. Biol. *223*, 31-40.

Gerchman, E. E., Graziano, V., Ramakrishnan, V. (1993). Expression of linker histones in *E. coli*: sources of problems and methods for overcoming some of the difficulties. Protein Expression and Purification *5*, 242-251.

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

Goytisolo, F. A., Packman, L.C., Thomas, J.O. (1996). Photoaffinity labelling of a DNA-binding site on the globular domain of histone H5. Eur. J. Biochem. *242*, 619-626.

Graziano, V., Gerchman, S.E., Schneider, D.K., Ramakrishnan, V. (1994). Histone H1 is located in the interior of the chromatin 30-nm filament. Nature *24*, 351-354.

Graziano, V., Gerchman, S.E., Ramakrishnan, V. (1988). Reconstitution of chromatin higher-order structure from histone H5 and depleted chromatin. J. Mol. Biol. *203*, 997-1007.

Green, G. R., Poccia, D.L. (1985). Phosphorylation of sea urchin sperm H1 and H2B histones precedes chromatin decondensation and H1 exchange during pronuclear formation. Dev. Biol. *108*, 235-245.

Grunstein, M. (1990). Nucleosomes: regulators of transcription. Trends Genet. *6*, 395-400.

Guex, N., Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis *18*, 2714-2723.

Guschin, D., Chandler, S., Wolffe, A.P. (1998). Asymmetric linker histone association directs the asymmetric rearrangement of core histone interactions in a positioned

nucleosome containing a thyroid hormone response element. Biochemistry *37*, 8629-8636.

Hamiche, A., Schultz, P., Ramakrishnan, V., Oudet, P., Prunell, A. (1996). Linker histone-dependent DNA structure in linear mononucleosomes. J. Mol. Biol. *257*, 30-42.

Harp, J. M., Hanson, B.L., Timm, D.E., Bunick, G.J. (2000). Asymmetries in the nucleosome core particle at 2.5 A resolution. Acta Crystallogr. D. Biol. Crystallogr. *56*, 1513-1534.

Hayashi, T., Hayashi, H., Iwai, K. (1987). Tetrahymena histone H1. Isolation and amino acid sequence lacking the central hydrophobic domain conserved in other H1 histones. J. Biochem. (Tokyo) *102*, 369-376.

Hayes, J. J., Pruss, D., Wolffe, A.P. (1994). Contacts of the globular domain of histone H5 and core histones with DNA in a "chromatosome". Proc. Natl. Acad. Sci. USA *91*, 7818-7821.

Hayes, J. J., Wolffe, A.P. (1993). Preferential and asymmetric interaction of linker histones with 5S DNA in the nucleosome. Proc. Natl. Acad. Sci. USA *90*, 6415-6419.

Hayes, J. J. (1996). Site-directed cleavage by a linker histone-Fe(II) EDTA conjugate: localization of a globular domain binding site within a nucleosome. Biochemisry *35*, 11931-11937.

Hendrickson, F. M., Cole, R.D. (1994). Selectivity in the interaction of various DNA sequences with H1 histone. Biochemistry *33*, 2297-3006.

Hendrickson, F. M., Cole, R.D. (1994). Selectivity in the interaction of various DNA sequences with H1 histone. Biochemistry *33*, 2997-3006.

Herrera, J. E., West, K.L., Schiltz, R.L., Nakatani, Y., Bustin, M. (2000). Histone H1 is a specific repressor of core histone acetylation in chromatin. Mol. Cell. Biol. *20*, 523-529.

Hertzberg, R. P., Dervan, P.B. (1984). Cleavage of DNA with methidiumpropyl-EDTA-iron(II): reaction conditions and product analyses. Biochemistry *23*, 3934-3945.

Hill, C. S., Rimmer, J.M., Green, B.N., Finch, J.T., Thomas, J.O. (1991). Histone-DNA interactions and their modulation by phosphorylation of -Ser-Pro-X-Lys/Arg- motifs. EMBO J. *10*, 1939-1948.

Holm, L., Sander, C. (1995). Dali: a network tool for protein structure comparison. TIBS *20*, 478-480.

Jenuwein, T., Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Kadowaki, H., Kadowaki, T., Wondisford, F.E., Taylor, S.I. (1989). Use of polymerase chain reaction catalyzed by Taq DNA polymerase for site-specific mutagenesis. Gene *76*, 161-166.

Kasinsky, H. E., Lewis, J.D., Dacks, J.B., Ausio, J. (2001). Origin of H1 linker histones. FASEB J. *15*, 34-42.

Khochbin, S. (2001). Histone H1 diversity: bridging regulatory signals to linker histone function. Gene *271*, 1-12.

Krylov, D., Leuba, S., Van Holde, K., Zlatanova, J. (1993). Histones H1 and H5 interact preferentially with crossovers of double-helical DNA. Proc. Natl. Acad. Sci. USA *90*, 5052-5056.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T7. Nature *227*, 680-685.

Lambert, S., Muyldermans, S., Baldwin, J., Kilner, J., Ibel, K., Wijns, L. (1991). Neutron scattering studies of chromatosomes. Biochem. Biophys. Res. Commun. *179*, 810-816.

Lauderdale, J. D., Stein, A. (1993). Effects of plasmid length and positioned nucleosomes on chromatin assembly in vitro. Biochemistry *32*, 489-499.

Lee, K. M., Hayes, J.J. (1998). Linker DNA and H1-dependent reorganization of histone-DNA interactions within the nucleosome. Biochemistry *37*, 8622-8628.

Lennard, A. C., Thomas, J.O. (1985). The arrangement of H5 molecules in extended and condensed chicken erythrocyte chromatin. EMBO. J. *4*, 3455-3462.

Leuba, S. H., Zlatanova, J., van Holde, K. (1993). On the location of histones H1 and H5 in the chromatin fiber. Studies with immobilized trypsin and chymotrypsin. J. Mol. Biol. *229*, 917-929.

Lewin, B. (1995). Genes VI (Oxford: Oxford University Press).

Lin, Q., Sirotkin, A., Skoultchi, A.I. (2000). Normal spermatogenesis in mice lacking the testis-specific linker histone H1t. Mol. Cell. Biol. *20*, 2122-2128.

Linder, C., Thoma, F. (1994). Histone H1 expressed in Saccharomyces cerevisiae binds to chromatin and affects survival, growth, transcription, and plasmid stability but does not change nucleosomal spacing. Mol. Cell. Biol. *14*, 2822-2835.

Ling, L. L., Keohavong, P., Dias, C., Thilly, W.G. (1991). Optimization of the polymerase chain reaction with regard to fidelity: modified T7, *Taq*, and vent DNA polymerases. PCR Methods Appl. *1*, 63-69.

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

Mackay, D. H. J., Cross, A.J., Hagler, A.T. (1989). The role of energy minimisation in simulation strategies of biomolecular systems. *In* Prediction of protein structure and principles the principles of protein conformation. (New York: Plenum Press), pp. 317-358.

McArthur, M., Thomas, J.O. (1996). A preference of histone H1 for methylated DNA. EMBO J. *15*, 1705-1714.

McGhee, J. D., Nickol, J.M., Felsenfeld, G., Rau, D.C. (1983). Higher order structure of chromatin: orientation of nucleosomes within the 30 nm chromatin solenoid is independent of species and spacer length. Cell *33*, 831-841.

Mirzabekov, A. D., Pruss, D.V., Ebralidse, K.K. (1990). Chromatin superstructure-dependent crosslinking with DNA of the histone H5 residues Thr1, His25 and His62. J. Mol. Biol. *211*, 479-491.

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

Mitra, S., Sen, D., Crothers, D.M. (1984). Orientation of nucleosomes and linker DNA in calf thymus chromatin determined by photochemical dichroism. Nature *308*, 247-250.

Nelson, H. C., Finch, J.T., Luisi, B.F., Klug, A. (1987). The structure of an oligo(dA)·oligo(dT) tract and its biological implications. Nature *330*, 221-226.

Nightingale, K., Wolffe, A.P. (1995). Methylation at CpG sequences does not influence histone H1 binding to a nucleosome including a *Xenopus borealis* 5S rRNA gene. J. Biol. Chem. *270*, 4197-4200.

Nightingale, K. P., Pruss, D., Wolffe, A.P. (1995). A single high-affinity binding site for histone H1 in a nucleosome containing the *Xenopus borealis* 5S ribosomal RNA gene. J. Biol. Chem. *271*, 7090-7094.

Noll, M., Kornberg, R.D. (1977). Action of micrococcal nuclease on chromatin and the location of histone H1. J. Mol. Biol. *109*, 393-404.

Panetta, G., Buttinelli, M., Flaus, A., Richmond, T.J., Rhodes, D. (1998). Differential nucleosome positioning on *Xenopus* oocyte and somatic 5 S RNA genes determines both TFIIIA and H1 binding: a mechanism for selective H1 repression. J. Mol. Biol. *282*, 683-697.

Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y.W., Ebright, R.E., Berman, H.M. (1996). Structure of the CAP-DNA complex at 2.5 angstroms resolution: a complete picture of the protein-DNA interface. J. Mol. Biol. *260*, 395-408.

Patterton, H.-G., Landel, C.C., Landsman, D., Peterson, C.L., Simpson, R.T. (1998). The biochemical and phenotypic characterization of Hho1p, the putative linker histone H1 of Saccharomyces cerevisiae. J. Biol. Chem. *273*, 7268-7276.

Pehrson, J. R. (1996). Probing the conformation of nucleosome linker DNA in situ with pyrimidine dimer formation. J. Biol. Chem. *270*, 22440-22444.

Polak, E., and Ribiere, G. (1969). Note sur la Convergence de Methodes de Directions Conjugees. Revue Francaise d'Informatique et de Recherche Operationnelle *3*, 35-43.

Pruss, D., Bartholomew, B., Persinger, J., Hayes, J., Arents, G., Moudrianakis, E.N., Wolffe, A.P. (1996). An asymmetric model for the nucleosome: a binding site for linker histones inside the DNA gyres. Science *274*, 614-617.

Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L., Sweet, R.M. (1993). Crystal structure of the globular domain of histone H5 and its implications for nucleosome binding. Nature *362*, 219-233.

Ramakrishnan, V. (1997). Histone structure and the organization of the nucleosome. Annu. Rev. Biophys. Biomol. Struct. *26*, 83-112.

Ring, D., Cole, R.D. (1983). Close contacts between H1 histone molecules in nuclei. J. Biol. Chem. *258*, 15361-15364.

Roth, S. Y., Schulman, I.G., Richman, R., Cook, R.G., Allis, C.D. (1988). Characterization of phosphorylation sites in histone H1 in the amitotic macronucleus of *Tetrahymena* during different physiological states. J. Cell. Biol. *107*, 2473-2482.

Roth, S. Y., Allis, C.D. (1992). Chromatin condensation: does histone H1 dephosphorylation play a role? TIBS *17*, 93-98.

Rousseau, S., Renaud, J., Ruiz-Carrillo, A. (1989). Basal expression of the histone H5 gene is controlled by positive and negative *cis*-acting sequences. Nucleic Acids Res. *17*, 7495-7511.

Rousseau, S., Asselin, M., Renaud, J., Ruiz-Carrillo, A. (1993). Transcription of the histone H5 gene is regulated by three differentiation-specific enhancers. Mol. Cell. Biol. *13*, 4904-4917.

Ruiz-Carrillo, A., Wangh, L.J., Allfrey, V.G. (1975). Processing of newly synthesized histone molecules. Science *190*, 117-128.

Sander, C., Schneider, R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins *9*, 56-68.

Satchwell, S. C., Travers, A.A. (1989). Asymmetry and polarity of nucleosomes in chicken erythrocyte chromatin. EMBO J. *8*, 229-238.

Schlegel, R. A., Haye, K.R., Litwack, A.H., Phelps, B.M. (1980). Nucleosome repeat lengths in the definitive erythroid series of the adult chicken. Biochim. Biophys. Acta *606*, 316-330.

Schwarz, P. M., Hansen, J.C. (1994). Formation and stability of higher order chromatin structures: Contributions of the histone octamer. J. Biol. Chem. *269*, 16284-16289.

Segers, A., Muyldermans, S., Wyns, L. (1991). The interaction if histone H5 and its globular domain with core particles, depleted chromatosomes, polynucleosomes, and a DNA decamer. J. Biol. Chem. *266*, 1502-1508.

Segers, A., Wyns, L., Lasters, I. (1991). A model for histone H5-DNA interaction: simultaneous minor and major groove binding. Biochem. Biophys. Res. Commun. *174*, 898-902.

Segers, A., Wyns, L., Laster, I. (1991). A model for histone H5-DNA interaction: simultaneous minor and major groove binding. Biochem. Biophys. Res. Commun. *174*, 898-902.

Shimada, T., Okihama, Y., Murata, C., Shukuya, R. (1981). Occurrence of H1o-like protein and protein A24 in the chromatin of bullfrog erythrocytes lacking histone 5. J. Biol. Chem. *256*, 10577-10582.

Simpson, R. T. (1978). Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. Biochemistry *17*, 5524-5531.

Staynov, D. Z. (2000). DNase I digestion reveals alternating asymmetrical protection of the nucleosome by the higher order chromatin structure. Nucl. Acids Res. *28*, 3092-3099.

Staynov, D. Z., Crane-Robinson, C. (1988). Footprinting of linker histones H5 and H1 on the nucleosome. EMBO. J. *7*, 3685-3691.

Strupat, K., Karas, M., Hillenkamp, F. (1991). 2,5-Dihydroxybenzoic acid: a new matrix for laser desorption-ionization mass spectrometry. Int. J. Mass. Spectrom. Ion Processes *111*, 89.

Sulkowski, E. (1989). The saga of IMAC and MIT. Bioessays *10*, 170-175.

Sung, M. T., Harford, J., Bundman, M., Vidalakas, G. (1977). Metabolism of histones in avian erythroid cells. Biochemistry *16*, 279-285.

Suzuki, M. (1989). SPKK, a new nucleic acid-binding unit of protein found in histone. EMBO J. *8*, 797-804.

Sweet, M. T., Allis, C.D. (1993). Phosphorylation of linker histones by cAMP-dependent protein kinase in mitotic micronuclei of *Tetrahymena*. Chromosoma *102*, 637-647.

Tanigawa, Y., Tsuchiya, M., Imai, Y., Shimoyama, M. (1983). ADP-ribosylation regulates the phosphorylation of histones by the catalytic subunit of cyclic AMP-dependent protein kinase. FEBS Lett. *160*, 217-220.

Tarkka, T., Oikarinen, J., Grundstrom, T. (1997). Nucleotide and calcium-induced conformational changes in histone H1. FEBS. Lett. *406*, 56-60.

Thoma, F., Koller, T., Klug, A. (1979). Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. J. Cell. Biol. *83*, 203-227.

Thomas, J. O., Rees, C., Finch, J.T. (1992). Cooperative binding of the globular domains of histones H1 and H5 to DNA. Nucl. Acids Res. *20*, 187-194.

Thomas, J. O., Khabaza, A.J. (1980). Cross-linking of histone H1 in chromatin. Eur. J. Bioch. *112*, 501-511.

Thomas, J. O. (1999). Histone H1: location and role. Curr. Opin. Cell. Biol. *11*, 312-317.

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

Tomaszewski, R., Jerzmanowski, A. (1997). The AT-rich flanks of the oocyte-type 5S RNA gene of *Xenopus laevis* act as a strong local signal for histone H1-mediated chromatn reorganization *in vitro*. Nucl. Acids Res. *25*, 458-465.

Travers, A. (1999). The location of the linker histone on the nucleosome. TIBS *24*, 4-7.

Travers, A. A., Muyldermans, S.V. (1996). A DNA sequence for positioning chromatosomes. J. Mol. Biol. *257*, 486-491.

Trieschmann, L., Schulze, E., Schulze, B., Grossbach, U. (1997). The histone H1 genes of the dipteran insect, *Chironomus thummi*, fall under two divergent classes and encode proteins with distinct intranuclear distribution and potentially different functions. Eur. J. Biochem. *250*, 184-196.

Tullius, T. D. (1988). DNA footprinting with hydroxyl radical. Nature *332*, 663-664.

Ura, K., Wolffe, A.P., Hayes, J.J. (1994). Core histone acetylation does not block linker histone binding to a nucleosome inclusing a *Xenopus borealis* 5S rRNA gene. J. Biol. Chem. *269*, 27171-27174.

Van Holde, K. E. (1989). Chromatin, 1st Edition, A. Rich, ed. (New York: Springer-Verlag).

Varga-Weiss, P., Zlatanova, J., Leuba, S., Schroth, G.P., Van Holde, K. (1994). Binding of histones H1 and H5 and their globular domains to four-way junction DNA. Proc. Natl. Acad. Sci. USA *91*, 3525-3529.

Vriend, G. (1990). WHAT IF: A molecular modeling and drug design program. J. Mol. Graph. *8*, 52-56.

Wang, Z. F., Sirotkin, A.M., Buchold, G.M., Skoultchi, A.I., Marzluff, W.F. (1997). The mouse histone H1 genes: gene organization and differential regulation. J. Mol. Biol. *271*, 124-38.

Wellman, S. E., Sittman, D.B., Chaires, J.B. (1994). Preferential binding of H1e to GC-rich DNA. Biochemistry *33*, 384-388.

Williams, S. P., Athey, B.D., Muglia, L.J., Schappe, R.S., Gough, A.H., Langmore, J.P. (1986). Chromatin fibers are left-handed double helices with diameter and mass per unit length that depend on linker length. Biophys. J. *49*, 233-248.

Wolffe, A. P. (1998). Chromatin: Structure and Function, 3rd Edition (Academic Press).

Woodcock, C. L. (1994). Chromatin fibers observed *in situ* in frozen hydrated sections. Native fiber diameter is not correlated with nucleosome repeat length. J. Cell. Biol. *125*, 11-19.

Woodcock, C. L., Frado, L.L., Rattner, J.B. (1984). The higher-order structure of chromatin: evidence for a helical ribbon arrangement. J. Cell. Biol. *99*, 42-52.

Zahn (1996). Overexpression of an mRNA dependent on rare codons inhibits protein synthesis and cell growth. J. Bacteriol. *178*, 2926-2933.

Zentgraf, H., Franke, W.W. (1984). Differences of supranucleosomal organization in different kinds of chromatin: cell type-specific globular subunits containing different numbers of nucleosomes. J. Cell. Biol. *99*, 272-286.

Zhou, Y.-B., Gerchman, S.E., Ramakrishnan, V., Travers, A., Muyldermans, S. (1998). Position and orientation of the globular domain of linker histone H5 on the nucleosome. Nature *395*, 402-405.

Zlatanova, J., Doenecke, D. (1994). Histone H1 zero: a major player in cell differentiation? FASEB J. *8*, 1260-1268.

Zlatanova, J., van Holde, K. (1996). The linker histones and chromatin structure: new twists. Prog. Nucl. Acids Res. Mol. Biol. *52*, 217-259.