3-29-2006

Biochemical characterization of mammalian high mobility group protein A2

Lorraine Katy Edwards
Florida International University

DOI: 10.25148/etd.FI15101270

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

BIOCHEMICAL CHARACTERIZATION OF MAMMALIAN HIGH MOBILITY GROUP PROTEIN A2

A thesis submitted in partial fulfillment of the

requirements for the degree of

MASTER OF SCIENCE

in

FORENSIC SCIENCE

by

Lorraine Katy Edwards

2006

To: Interim Dean Mark Szuchman
College of Arts and Sciences

This thesis, written by Lorraine Katy Edwards, and entitled Biochemical Characterization of Mammalian High Mobility Group Protein A2, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this thesis and recommend that it be approved.

Kalai Mathee

Lidia Kos

Fenfei Leng, Major Professor

Date of Defense: March 29, 2006

The thesis of Lorraine Katy Edwards is approved.

Interim Dean Mark Szuchman
College of Arts and Sciences

Interim Dean Stephan L. Mintz
University Graduate School

Florida International University, 2006

DEDICATION

I dedicate this thesis to my family, who has showered me with encouragement throughout my life. To my extended family who will always be there for me. My adopted family, who has shown me that love, is greater than blood. To my grandparents, who are constantly in my thoughts and whom I still think of in times of trouble and in joy. To my brother, Tim, I still look up to you, I always will. You made me want to do well, to be just as good as you, thank you. To my mum, you are my best friend, you are my world. Without you I could not have done this work, I hope I make you proud. Finally to my dad, I am sorry you did not get to see me finish this work, at least you knew that I started, I did it.

ACKNOWLEDGMENTS

I would like to thank my major professor Dr Fenfei Leng for taking me under his wing. For allowing me to make mistakes along the way, and then teaching me the right way to turn. I would also like to thank my other committee members, Dr Kalai Mathee and Dr Lidia Kos who have given me lots of support along the way and bestowed that playing hard is as important as working hard.

I would like to thank Dr Keith Brew and his lab at FAU for the use of his isothermal titration calorimeter. Also to Dr Watson Lees and Dr DeEtta Mills, both at FIU for use of their lab equipment. I could not have completed this work if not for funding supplied by Dr Leng and my family and also the opportunities provided by teaching under the instruction of Dr Palmer Graves and Dr Doug Heller.

Lab members both past and present have supported me through some rough times and I would especially like to thank them (Gela, Hilda, Roger, Jane, Becca, Danielle, Linda, Mao, Lucy, Suzanne, and Sean). A special mention must go to Roger Cui for developing the purification procedure, Gela Mendoza Baez for the purification of the truncated protein, and Suzanne Joynt for aiding in the ITC experiments. Also to Rebecca Samul for editing this thesis.

Finally I would like to thank my friends and family. I now consider many at FIU an extension of my family. You have made my time in Miami an enjoyable one, thank you. Also to my family, who has always shown encouragement in every adventure I undertake.

ABSTRACT OF THE THESIS

BIOCHEMICAL CHARACTERIZATION OF MAMMALIAN HIGH MOBILITY GROUP PROTEIN A2

by

Lorraine Katy Edwards

Florida International University, 2006

Miami, Florida

Professor Fenfei Leng, Major Professor

The high mobility group protein HMGA2 is an architectural transcription factor, which is expressed during embryogenesis. Aberrant expression causes benign and malignant tumor formation. The protein possesses three "AT hook" domains and an acidic C-terminal. HMGA2 is natively unstructured, however it forms a homodimer. In this study site-directed mutagenesis was used to create single methionine mutants, HMGA2Q37M, HMGA2I71M and HMGA2Q85M. These mutants were cross-linked using EDC and then cleaved using CNBr to determine which domains are involved in homodimer formation. Our results indicate that the second "AT hook" domain may interact with the C-terminal. We then labeled a peptide containing the C-terminal (CTP) with tetramethylrhodamine-5-maleimide (TRM). We found that the CTP-TMR binds to HMGA2Δ95-108, which lacks the C-terminal. These results suggest that the C-terminal is required for homodimer formation. The techniques used within this study can be applied to forensics and with further research HMGA2 may have a forensic application.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

A. Forensics

The root of the word forensic is *forensis* in Latin meaning 'open court, public' (Soanes and Hawker, 2005). Today forensic refers to the use of scientific or technical methods to discern matters in relation to a crime. It means in relation or pertaining to a court of law (Wikipedia, 2006). The job of a forensic scientist is to collect, preserve and examine physical evidence that may be useful in determining the, who, what, when and where a crime took place. It does not include the reason why the crime occurred. Evidence is collected from the crime scene and then is processed by specialized personnel in order to determine certain characteristics of the crime. There are many specialized areas related to this field that include Chemistry, Biology and Entomology as well as forensic accountants, computer specialists and blood spatter analysts to name a few (Saferstein, 2003).

The job of a Forensic Biologist entails the detection, identification and subsequent comparison of biological material found at the scene of a crime (Rudin and Inman, 2002). Deoxyribonucleic acid (DNA) can be used to identify an individual for paternity testing or to exclude/include a suspect as being present at the scene of a crime (Rudin and Inman, 2002). DNA is present in all cells of the body with the exception of mature red blood cells and is unique to every individual aside from identical twins (Rudin and Inman, 2002). Sections of DNA contain inherent repeats of a particular sequence known as short tandem repeats (STRs). The likelihood of having a specific repeat number at

these locations is known. The numbers of repeats at these regions are measured and a profile of an individual can be determined. The greater the number of regions that are tested the more unique the profile becomes and therefore DNA can then be used for identification purposes (Rudin and Inman, 2002).

Specific regions of DNA are known as genes and these are the functional units of hereditary. The gene itself contains an ordered sequence of nucleotides which can be transcribed and then translated to create an amino acid chain. This amino acid chain is capable of folding to create a protein (Lewin, 2003). The expression of many proteins is transient and therefore the presence or absence of that particular protein can be a useful determinant of state of health or age (Patel *et al.*, 1994). Other proteins are only expressed at specific sites in the body and may aid in the forensic detection of body tissues (Castagnoli *et al.*, 1994). DNA and proteins can be used in forensic science in unique ways due to their distinctive qualities. Once tests are performed on evidence, the results can be used in a court of law where it is the job of the analyst to present the findings in layman's terms. The analyst teaches the jury and public the importance of the test performed as well as the results of the test (Saferstein, 2003). This can include the DNA profile of a suspect or victim.

Protein Forensics

A breakthrough in categorizing biological fluids came about in 1901, when Karl Landsteiner determined that blood could be classified into four categories, namely; A, B, AB and O (Saferstein, 2003). Classification was based on differences in the antigens present on the surface of red blood cells. Landsteiner also noted that proteins known as antibodies are found in the blood serum. Each antibody binds to a specific antigen. When

an antibody comes in contact with a specific antigen, agglutination occurs due to the bivalent nature of the antibody attaching to two antigens. This forms a large cross-linked network of antigens and antibodies. The body does not produce an antibody against the antigen, which is present on the red blood cell of that individual. Instead they produce an antibody against the opposite antigen. Someone with antigen A will have anti-B antibodies and a person with antigen B will have anti-A antibodies. An individual who is described as "AB" has both A and B antigens and no anti-A nor anti-B antibodies. Those referred to as "O" have no A, nor B antigens, but possess both anti-A and anti-B antibodies. This prevents agglutination occurring within the body (Saferstein, 2003).

Before Karl Landsteiner's discovery many people died immediately after being transfused with the blood of a different blood type. Since this time many blood typing methods have been discovered that employ proteins found on red blood cells or in plasma. The Rhesus system, found in 1937, was developed to produce an antibody against the red blood cells of the Rhesus monkey. It was additionally found that the blood serum of roughly 85% of the human population agglutinated to this serum (Saferstein, 2003). Therefore blood cells that do agglutinate are known as Rhesus positive (Wikipedia, 2006). The ABO system was used for many years in forensic science as an inclusionary or exclusionary too. This system alone did not provide the level of discrimination required for convicting or exonerating an individual of a crime. New classification systems were discovered based on the ABO system and included phosphoglucomutase and adenylate kinase (Rudin and Inman, 2002). Together these systems could be used to increase the level of discrimination and could be used to narrow the search by excluding more individuals. In this setting, work on proteins has been

replaced by DNA typing, which has more variance at each locus investigated and therefore more discrimination power (Rudin and Inman, 2002). Furthermore with the dawn of the polymerase chain reaction (PCR), minute traces of biological fluids that contain DNA can by typed (Mullis *et al.*, 1986). As a result DNA rather than protein became the preferred tool of forensic scientists for human identification.

However, there are still many uses for proteins in forensic science. DNA does not readily change due to varying environmental conditions such as heat and humidity (Rudin and Inman, 2002). Whereas this characteristic makes DNA an ideal candidate for typing individuals it does not provide information on the persons state of health, age or other changing factor associated with that entity. Proteins are affected due to environmental stresses and many factors can be discerned from the transient nature of these proteins (Rudin and Inman, 2002). The location of proteins can also be specialized, changing depending upon factors affecting a particular area of the body. This can provide more information about the body part present, the environmental stresses that took place and what bodily influences occurred based on where in the body the protein was found.

Proteins are presently used for the presumptive detection of semen (Rudin and Inman, 2002). Semen stains may be seen with the naked eye but this is made more difficult if there are minute traces or the garment has been washed. In 1989, an antigen known as prostate specific antigen (PSA) was first used to aid in the detection of seminal fluid stains (Kamenev *et al.*, 1989). An immunoassay was developed in 1992 for anti-PSA antigens using a radio-labeled Protein A detection system (Rao and Kashyap, 1992). This technique is extremely sensitive and can detect stains up to five years old (Rao and

Kashyap, 1992). In 2002 a "SMITEST" PSA immunochromatographic membrane test card was devised for ease of use to be utilized at the scene of a crime (Sato *et al.*, 2001).

Other proteins, besides PSA, have been found to aid in the identification of semen. The seminal plasma protein semenogelin and p84, a blood group substance present on the sperm plasma membrane help determine the ABO blood-type of the individual through their semen (Sato *et al.*, 1995, 2001).

Protein can also aid in the identification of blood source. Miyaishi *et al.* discovered that by measuring the concentrations of both FDP-D-dimer (fibrin degradation products) and myoglobin, the origin of a bloodstain could be determined as either peripheral or menstrual blood (Miyaishi *et al.*, 1996). High levels of FDP-D-dimer coupled with low levels of myoglobin can identify the stain as menstrual blood (Miyaishi *et al.*, 1996).

Species-specific blood tests also utilize proteins. Tests of this nature are extremely important when dealing with hunting or illegal selling of endangered species, and/or discrimination between animal and human blood at a crime scene. In 1993 Matsuzawa *et al.* used rabbit antisera against human immunoglobulin G (IgG) in a dot blot method to determine identity (Matsuzawa *et al.*, 1993). The test was rapid, completed in less than 5 minutes, but was unable to differentiate between closely related species. In this case, human and monkey would both give a positive result (Matsuzawa *et al.*, 1993). This test could prove useful in forensics at a crime scene where it is unlikely that two similar species would be believed to be present. In the mid 1990's research included the use of albumins, hemoglobin's and hair keratins for species identification using electrophoresis (Miller *et al.*, 1995, Folin and Contiero, 1996). In 1995, Miller *et al.* found differences in

the electrophoretic mobility's of albumin and hemoglobin's using a 1-dimensional electrophoretic technique (Miller *et al.*, 1995). They found differences between blood of the pheasant (*Phasianus colchicus*) and that of a protected species, otter (*Lutra lutra*) (Miller *et al.*, 1995). A year later Folin and Contiero discovered mobility differences between non-human primates. They compared hair keratins on 15 % sodium dodecyl sulfate-polyacrylamide gels (SDS-PAGE) (Folin and Contiero, 1996). Their results also showed that neither sex nor sample age had an impact on the keratin (Folin and Contiero, 1996). Since hair keratins were fairly conserved, the quantity of each keratin was more evident than the absence or presence of a particular type of keratin. They found the highest differences in intensity of polypeptide bands between the non-closely related animals (Folin and Contiero, 1996). If the sensitivity of the test is increased, members of closely related species could be discriminated.

At a crime scene it may be necessary to determine what tissue is present; this is just as important as identifying who the blood or tissue came from. DNA is not a helpful factor in this case, as DNA remains constant between tissues within the body. Specific proteins are found in regionally specialized areas of the body and can therefore help determine the part of the body the tissue derived from. In 1996, Takahama developed a sandwich enzyme immunoassay for use in detecting three organs. These were cardiac troponin I to distinguish heart, liver specific antigen (LSA) for the liver and sucrase-isomaltase to help identify the small intestine (Takahama, 1996). These assays were soon followed by the sandwich enzyme immunoassay developed by Seo *et al.* for detection of brain tissue using S-100 protein in 1997 (Seo *et al.*, 1997). In 2004 another brain sensitive marker was suggested, the neurofilaments NF-H, NF-M and NF-L (Takata *et*

al., 2004). The detection of NF-L and NF-M however, was not always possible in the presence of brains grey matter. It was found that from the neurofilament proteins, NF-H was the best marker for human brain tissue. The detection of NF-H could be detected in tissue that had been subjected to many stresses such as heating and drying. This made NF-H a good candidate for use in a forensic setting as evidence may not be found for several days, months or even years (Takata *et al.*, 2004).

Another important factor in forensic science is the vitality of a wound, whether a wound occurred pre- or post-mortally. Proteins are useful markers of wound healing and determining the time since trauma occurred. Certain specific proteins are present at the site at different stages of wound healing based on the human inflammatory response. These proteins are not present or are reduced in expression at non-healing times (Wikipedia, 2006). In the late 1990s Dressler *et al.* examined the time dependent expression of intercellular adhesion molecule-I (ICAM-I) as an indication of the age of skin wounds (Dressler *et al.*, 1997). This was conducted using immunohistochemistry comparing wounded and non-wounded skin sections. It was concluded that a high expression of ICAM-I indicates that the wound was sustained more than 1.5 hrs and less than 3.5 days before the wound was tested (Dressler *et al.*, 1997). This same group later determined the expression of selectins (L-, P- and E-selectins) and vascular cell adhesion molecule-I (VCAM-I). Both E- and P-selectins are useful markers of skin wound age with E-selectin showing a strong expression between 1 hr and 17 days after injury and P-selectin showing a strong expression between 3 mins and 7 hrs after injury. The results of the VCAM-I expression studies were not as useful however, as only 51 % of the skin wounds showed expression. The strongest expression was between 3 hrs and 3.5 days

after injury and if other markers were taken into account VCAM-I expression could narrow the time estimation further (Dressler *et al.*, 2000). Later studies included pro-inflammatory cytokines, ubiquitin, chemokines and vascular endothelial growth factor (VEGF), as markers to aid in the determination of timing for skin injury (Grellner *et al.*, 2000, Kondo *et al.*, 2002a, 2002b, Hayashi *et al.*, 2004).

The role of the forensic pathologist is to determine postmortem interval (PMI) as well as cause of death. The typical determination of time since death is investigated through the use of body temperature (Henssge *et al.*, 2000). However, this is an inexact science as many factors including temperature, humidity and cause of death have an impact on the body's temperature (al-Alousi *et al.*, 2002). There are some protein markers that have been investigated in an attempt to determine post mortal interval including in 2003 an investigation of degradation of cardiac Troponin I (Sabucedo and Furton, 2003). Western blotting of the denatured protein was visualized. The particular banding pattern of the degrading protein reported the time since death when compared to a known banding pattern. This banding pattern was capable of determining postmortal interval between 0 and 5 days (Sabucedo and Furton, 2003).

In 2003 Kang *et al.* investigated the use of calmodulin binding proteins in reference to postmortem interval (Kang *et al.*, 2003). They removed both rat lung and skeletal muscle at 0, 24, 48 and 96 hrs after death. By performing immunoblot analysis on lung samples, predictable patterns shown on SDS-polyacrylamide gels of both myristoylated alanine-rich C-kinase substrate (MARCKS) and calcineurin A (CAN) were present. This indicated that, with further analysis on these proteins, calmodulin binding

proteins have a possible forensic use in determining short post-mortal intervals (Kang *et al.*, 2003).

Once the time of death is determined the cause of the death may still be unknown. The oxygen regulated protein 150 kDa (ORP-150) was analyzed as a possible marker for brain ischemia (Ikematsu *et al.*, 2004). Brain tissue sections of 31 patients' whose deaths were known to be; asphyxia, hypothermia, exsanguinations, carbon monoxide (CO) intoxication or sudden cardiac death were taken. The group noted the number of ORP-150 positive cells after introducing a polyclonal antibody to the tissue sections. It was recorded that the number of ORP-150 positive cells was influenced by age. Therefore they used a covariance method of calculating mean ORP-150 level that took into account the age of the deceased to determine brain ischemia. A correlation between ORP-150 levels and degree of brain ischemia after the age factor was removed was discovered. It was assumed that sudden cardiac death was rapid death and therefore no or very little ischemia occurred. This type of death showed the lowest ORP-150 levels. Several minutes of brain ischemia occurs in deaths of exsanguinations, CO intoxication and asphyxiation where they determined increased ORP-150 levels were present. Numerous hours of brain ischemia may occur in deaths of hypothermia and ORP-150 levels were further elevated. In conclusion ORP-150 may be a useful marker in determining the extent of brain ischemia before death (Ikematsu *et al.*, 2004).

Another cause of death, which creates much media attention, is meningitis, a bacterial infection caused by *Neisseria meningitides*. This type of death may also be attributable to other microorganisms such as *Haemophilus influenzae* or *Streptococcus pneumoniae*. Rapid determination of the cause of death in this case is needed to prevent

transmission of the infection. DNA may be used in this instance but a faster method is required to prevent further spread of the infection. In 2005 a screening method involving latex agglutination was devised (Saha *et al.*, 2005). This test would occur prior to the use of PCR on the DNA. The latex agglutination method uses the recognition of *N. meningitides* specific antigens of serogroups A, B, C, Y and W135. Once this screening is completed and *N. meningitides* is suspected, a confirmatory test using PCR could be undertaken (Saha *et al.*, 2005). The latex agglutination test proved a valuable tool as a presumptive test allowing for the correct management to occur to prevent the spread of the infection.

The use and application of proteins in forensic science is expanding. Therefore as much as possible should be understood about the way proteins interact, their functions and structures in order for them to be used as potential markers. In this research the high mobility group protein 2 (HMGA2) will be used as a model protein in order to understand the structure of this type of protein. Cross-linking studies as well as isothermal titration calorimetry, size exclusion chromatography and gel mobility shift assays were performed. The conformation of protein-protein interactions as well as the binding stoichiometry of protein to DNA was determined from this research.

B. High Mobility Group Proteins

The high mobility group proteins are named due to their high mobility on SDS-polyacrylamide gels (Bustin and Reeves, 1996). Three families of high mobility group proteins exist. These are named after the functional domains that are present within each family member. The HMGB (old name HMG-1/2) family has the characteristic functional group of the "HMG-box" (Bustin, 2001) The HMGN (old name HMG-14/-17)

proteins have "nucleosomal binding domains" and the HMGA (old name HMG-I/Y) proteins have conserved "AT hook" DNA binding domains (Bustin, 2001, Manfioletti *et al.*, 1991). All family members are of low molecular weight < 30 kDa and are soluble in 5% perchloric acid (Giancotti, *et al.*, 1985).

High Mobility Group B

The mammalian HMGB protein family possesses two HMG "box" binding domains, which are composed of three α-helices arranged in an L shape as well as an acidic C-terminal (Weir *et al.*, 1993). The HMG "boxes" named A and B have structural and functional differences, but are generally similar (Weir *et al.*, 1993, Hardman *et al.*, 1995). The boxes bind to DNA with no sequence specificity; they instead bind to the minor groove of DNA and induce a bend (Allain *et al.*, 1999). Alternatively they bind to and stabilize pre-bent DNA such as four-way junctions and cruciform structures (Bianchi *et al.*, 1989, Hill *et al.*, 1997, 1999). The C-terminal tail controls the binding affinity for these different DNA structures (Sheflin *et al.*, 1993). The HMGB proteins function as architectural transcription factors (Zwilling *et al.*, 1995). Architectural transcription factors refer to a group of proteins that do not possess transcriptional activation domains yet these proteins still aid in transcriptional control. These architectural transcription factors work by interacting with and changing the conformation of DNA (Wolffe, 1994). The alteration in DNA conformation provides the correct framework for the binding of transcription factors and the polymerase (Bustin and Reeves, 1996). Architectural transcription factors can direct a number of transcription factors, which assemble into a higher order nucleoprotein known as an enhanceosome (Wolffe, 1994). HMGB1 has been shown to bind cooperatively with ZEBRA to the enhancer or promoter of the

Epstein Barr (EB) gene and aid in the formation of the enhanceosome structure (Ellwood *et al.*, 2000).

High Mobility Group N

All members of the HMGN family contain nucleosomal binding domains as well as a C-terminal domain which functions to unfold chromatin (Walker *et al.*, 1977, Abercrombie *et al.*, 1978). The protein family consists of four proteins, HMGN1, HMGN2, HMGN3 and HMGN4. HMGN3 is expressed as two forms due to alternative splicing into proteins HMGN3a and HMGN3b, where HMGN3b lacks most of the C-terminal domain (West *et al.*, 2001). The function of this family of proteins *in vitro* is to increase transcription and replication due to unfolding of chromatin on nucleosomes (West, 2004). Two HMGN protein binding sites exist on each nucleosome allowing for attachment of the nucleosomal binding domains of each protein to the inner side of nucleosomal DNA (Sandeen *et al.*, 1980). The *in vivo* functions of the protein remain less understood as the protein family is only expressed in higher eukaryotes (West, 2004). However, knockout studies in mice involving the *Hmgn1* gene suggest a role in DNA repair as these *Hmgn1*$^{-/-}$ mice are more sensitive to UV-irradiation (Birger *et al.*, 2003). This suggests that HMGN1 may unfold chromatin surrounding the DNA damage allowing for DNA repair machinery to enter (Birger *et al.*, 2003).

High Mobility Group A

The mammalian HMGA family consists of three proteins coded by *HMGA1* and *HMGA2*. The *HMGA1* gene expresses the HMGA1a and HMGA1b proteins, which occur through alternative splicing (Johnson *et al.*, 1988). The *HMGA2* gene expresses the HMGA2 protein only (Manfioletti *et al.*, 1991). A non-functional intron-less pseudogene

has been identified in mice. However, no functional alternatively spliced variant of HMGA2 has been found (Manfioletti *et al.*, 1995).The mammalian HMGA family of proteins contain three conserved "AT hook" DNA binding domains and an acidic C-terminal domain (Manfioletti *et al.*, 1991, Reeves and Nissen, 1990). The "AT hook" domains bind to the minor groove of AT rich DNA and alter the DNA conformation (Reeves and Nissen, 1990). The HMGA family of proteins function as architectural transcription factors (Wolffe, 1994). These proteins have been identified in relation to a large number of benign and malignant tumors, mainly of mesenchymal origin (reviewed in Fedele *et al.*, 2001).

HMGA1

The mammalian HMGA1 proteins possess three "AT hook" DNA binding domains (Reeves and Nissen, 1990). HMGA1b is a truncated version of HMGA1a missing 11 amino acids from the linker region between the third "AT hook" domain and the acidic C-terminal (Johnson *et al.*, 1988). These proteins function in regulating the transcription of a number of genes such as recombination activating gene 2 (RAG2) and interferon-β (Battista *et al.*, 2005, Thanos and Maniatis, 1992). HMGA1 is normally expressed in rapidly dividing embryonic cells, and its expression is negligible in normal adult tissues (Johnson *et al.*, 1988). Overexpression of HMGA1 results in the formation of a number of malignant tumors such as breast carcinomas and epithelial ovarian carcinomas (Baldassarre *et al.*, 2003, Masciullo *et al.*, 2003). Cells overexpressing HMGA1a protein were observed to inhibit nuclear excision repair pathways upon exposure to UV-irradiation (Adair *et al.*, 2005). Despite this knowledge little is understood about the molecular mechanism by which the HMGA1 proteins cause the

malignant phenotype. Loss of the *Hmga1* gene in mice was shown to result in reduced insulin secretion and weakened insulin signaling pathways so that the phenotype resembled that of type II diabetes (Foti *et al.*, 2005). Re-establishment of the protein expression resulted in restored insulin receptor protein expression and insulin binding capacity (Foti *et al.*, 2005).

HMGA2

HMGA2 is involved in the transcriptional control of a number of genes. HMGA2 upregulates the expression of cyclin A that is involved in cell cycle control (Minshull *et al.*, 1990). Cyclin A is a crucial factor; along with cyclin dependent kinase 2 (cdk2), for S phase entry (Lees *et al.*, 1992). Cyclin A also functions with cdc2 (cell division control) for the G_2/M transition (Minshull *et al.*, 1990). Cyclin A is repressed during the remaining cell cycle by the binding of $p120^{E4F}$ to the cyclic AMP (adenosine monophosphate)-responsive element (CRE) at the cyclin A promoter (Fajas *et al.*, 2001) HMGA2 works by binding, via the second "AT hook" DNA binding domain to $p120^{E4F}$ and preventing it from binding to the CRE. This prevents the repression of transcription by $p120^{E4F}$. Secondly, HMGA2 binds to the CRE and improves access of activating transcription factor-2 (ATF-2) to the enhancer. This interaction can be achieved due to protein-protein interactions and possibly DNA conformational changes (Tessari *et al.*, 2003).

HMGA2 is also involved in the prevention of apoptosis (Edelstein *et al.*, 2003). The protein has been shown to enhance expression of the apoptotic inhibitor *Bfl-1/A1* (Edelstein *et al*, 2003). HMGA2 is required for the correct recruitment of a large number of transcription factors to the *Bfl-1/A1* regulatory region (Edelstein *et al*, 2003). T-cell

activation triggers the recruitment of HMGA2, nuclear factor-κB (NF-κB), activator protein-1 (AP-1) and CCAAT enhancer binding protein β (C/EBPβ) to the region (Edelstein *et al.*, 2003) This is followed by binding of the co-factors TAFII250 (TBP-associated factor 250 kDa) and p300, and by a SWI/SNF (mating-type switch/sucrose nonfermenting) chromatin-remodeling complex. These factors drive the acetylation of histones H3 and H4 and hence activate transcription (Edelstein *et al.*, 2003).

The HMGA2 protein is involved in down-regulating some genes. It has been found that both the wild type and truncated HMGA2 (C-terminal removed) proteins down-regulated the DNA repair gene, *excision repair cross-complimenting rodent repair deficiency, complimentary group 1* (*ERCC1*) (Borrmann *et al.*, 2003). ERCC1 has been shown to be vital in the nuclear excision repair pathway to reverse DNA damage caused by UV-light and chemical mutagens (Araujo and Wood, 1999). The HMGA2 protein in this context worked through reversing bends in the DNA making assembly of a nucleoprotein complex at this promoter site more difficult (Borrmann *et al.*, 2003).

Cyclin A, Bfl-1/A1 and ERCC1 are involved in such processes as cell cycle control, apoptosis and DNA repair, respectively (Tessari *et al.*, 2003, Edelstein *et al.*, 2003, Borrmann *et al.*, 2003). Aberrant expression of HMGA2 thus involves the deregulation of these pathways and leads to incorrect functioning of the cells. In embryonic tissue, cells are rapidly dividing and the expressions of cell cycle control genes as well as apoptosis inhibitor genes are necessary. In adult tissues, most cells have become differentiated and as a result the expressions of cell cycle control and apoptosis inhibitor genes need to be more tightly regulated. DNA repair genes are also required,

thus the aberrant expression of HMGA2 deregulating these pathways cause diseases such as cancer.

Overexpression and Null Expression of *HMGA2*

Disruption of the mouse *Hmga2* gene results in a *pygmy* phenotype, in which there is a decline in adult body weight that mostly affects adipose tissue and embryonic fibroblasts (Zhou *et al.*, 1995). The brain region is least affected in the pygmy mouse as this region in the developing wild-type embryo shows the least *Hmga2* expression (Zhou *et al.*, 1996). It has also been shown that heterozygous null mice ($Hmga2^{+/-}$) have a 50 % reduction in expression compared to wild type mice (Zhou *et al.*, 1995). This signifies that the wild-type allele does not compensate for loss of function from the deleted allele (Zhou *et al.*, 1995). This information provided a link between *Hmga2* expression and growth. This link was further exemplified by a measurable *Hmga2* expression in fat deposits of normal mice after a weeklong high fat diet, whereas those on a standard diet did not express the *Hmga2* gene (Anand and Chada, 2000). Furthermore, genetically altered mice that lacked leptin (Lep^{ob}/Lep^{ob}) or its receptor ($Lepr^{db}/Lepr^{db}$) and are therefore genetically obese, showed a detectable level of *Hmga2* expression (Anand and Chada, 2000). Expression was not detected in any other tissues besides fat deposits. Further analysis using homozygous and heterozygous *Hmga2*-null mice on a high fat and standard diet demonstrated a haploinsufficiency effect between wild-type and $Hmga2^{+/-}$ mice upon weight gain at 30 weeks of age (Anand and Chada, 2000). This demonstrated that mice could avoid obesity after being fed a high fat diet if one or both *Hmga2* alleles are absent. The same mutations were also expressed in Lep^{ob}/Lep^{ob} mice (Anand and Chada, 2000). The results showed that there is a reduction in weight gain of the $Hmga2^{-/-}$,

Lep^{ob}/*Lep*^{ob} mice when compared to *Lep*^{ob}/*Lep*^{ob} mice alone. These mice were nevertheless heavier than their *Hmga2*$^{-/-}$ counterparts indicating another pathway in fat proliferation that is independent of *Hmga2* (Anand and Chada, 2000).

Overexpression of *HMGA2* can result in malignant neoplasias in human tissue. Röijer *et al.* described a carcinoma ex pleomorphic adenoma (CexPA) in which *HMGA2* was amplified along with a centromeric *murine double minute 2* (*MDM2*) gene (Roijer *et al.*, 2002). Pleomorphic adenoma is usually a benign tumor, which affects the salivary glands (Wesylik *et al.*, 2001).The benign neoplasm can undergo malignant transformation in select cases, with the frequency of occurrences dependent on duration before treatment (Roijer *et al.*, 2002, Bradley, 2005) The characteristics of CexPA involve a high level of recurrence and risk of metastases. It has been identified that 8 % of benign PA karyotypes involve rearrangements of 12q14-15 representing a region that includes both *HMGA2* and *MDM2* genes (Roijer *et al.*, 2002). In the case of CexPA, the breakpoint occurred centromeric to *HMGA2* and the entire gene was translocated to the der (10) marker. In some cases both the *HMGA2* and *MDM2* genes were deleted from the der (10) and formed visible extrachromasomal dmin. Further integration of these genes into additional chromosomal sites created homogenously stained regions (hsr). Southern blot analysis confirmed that in all of these cases the *HMGA2* and *MDM2* genes were amplified without any gross rearrangements and the entire gene was therefore amplified. Further analysis of the tumor using fluorescence *in situ* hybridization revealed a high level of amplification, mainly within the carcinomatous region of the tumor. Röijer *et al* therefore concluded that overexpression of *HMGA2* was linked to a malignant phenotype (Roijer *et al.*, 2002).

HMGA2 is also amplified and overexpressed in cases of human prolactinomas (Finelli *et al.*, 2002). This relationship was evaluated using transgenic mice carrying either the wild-type or the truncated version (lacks the C-terminal domain) of the *Hmga2* gene (Finelli *et al.*, 2002). These genes were put under the control of the cytomegalovirus promoter. It was found that a high percentage (85 %) of the female mice acquired pituitary adenomas which secreted growth hormones and prolactin within the first 6 months (Finelli *et al.*, 2002). The males developed tumors at a lower rate (40 %) and these tumors did not arise until later in development (before 18 months). These findings led to a possible connection between *Hmga2* and pituitary adenomas (Finelli *et al.*, 2002). In the case of human prolactin-secreting pituitary adenomas, there is a high incidence of trisomy of chromosome 12, which contains the *HMGA2* gene. Indeed, further investigation of such prolactinomas demonstrated that there is an increased quantity of HMGA2 mRNA and protein (Finelli *et al.*, 2002). The entire gene was overexpressed and no gross rearrangement was evident. A high degree of resistance to dopaminergic drugs was also developed, which is linked to a more aggressive tumor (Finelli *et al.*, 2002). The temporal and spatial expression of *HMGA2* is therefore important for the normal functioning of the cell.

Truncation of HMGA2

A truncated version of the *HMGA2* gene involves a breakpoint within the third intron, separating the three "AT hook" DNA binding domains from the acidic C-terminal domain. Two independent research groups generated transgenic mice expressing the truncated version. The first group used a novel embryonic stem cell approach where the transcript was transfected into the embryonic stem cell AB2.2 under the control of the

cytomegalovirus promoter (Battista *et al.*, 1999). G418-resistant clones were selected and clones expressing the highest level of truncated HMGA2 were microinjected into C57BL6/J mouse blastocysts and subsequently transferred to pseudopregnant females. The resultant offspring were then crossed with wild-type mice and the offspring generated contained a high expression level of truncated HMGA2. These mice showed a giant phenotype that exhibited increased levels of retroperitoneal and subcutaneous white adipose tissue (Battista *et al.*, 1999).

Another group created transgenic mice by microinjecting the truncated HMGA2 mRNA into a fertilized mouse embryo under the control of the H-2K^b (class I major histocompatibility complex promoter) (Arlotta *et al.*, 2000). Again, the mice generated showed a giant phenotype that manifested with an early increase in adipose tissue, inflammation of the adipose tissue and a high incidence of lipomas. These findings demonstrated that despite the ubiquitous expression of the truncated HMGA2 transcript, the phenotype exhibits only within the adipose tissue (Arlotta *et al.*, 2000). This demonstrated that HMGA2 facilitates growth of adipose tissue.

These findings, using a mouse model have also been demonstrated in human subjects. A recent study of an 8-year-old boy showed some of the same phenotypic observations as in the murine model (Ligon *et al.* 2005). This subject had a *de novo* pericentric inversion of chromosome 12, inv (12) (p11.22q14.3). The phenotypic observations included: multiple subcutaneous lipomas, advanced endochondral bone and dental ages, postnatal onset of extreme somatic overgrowth, persistent thrombocytopenia, arthritis, a stable cerebellar tumor, brachydactyly and facial dysmorphism. Breakpoint mapping studies revealed a breakpoint at 12q14.3 resulting in truncation of the *HMGA2*

gene within the third intron which separates the three "AT hook" regions from the C terminal (Ligon *et al.*, 2005).

Fusion Products Involving HMGA2

The human *HMGA2* gene maps to chromosomal location 12q14-q15 (Ashar *et al.*, 1996). Rearrangements involving *HMGA2* normally occur within the third intron separating the three "AT hook" domains from the C-terminal domain and the 3' untranslated region (UTR). Many fusion partners have been identified for *HMGA2* and these protein rearrangements have been identified in a variety of benign tumors, mainly of mesenchymal origin.

In lipomas the vast majority of rearrangements involving *HMGA2* also involve the gene at location 3q27-q28 (Petit *et al.*, 1996). This gene is known as the lipoma preferred partner (*LPP*) and represents a proline rich protein containing a leucine zipper and three LIM domains (named from the Lin-11, Isl-1 and Mec-3 genes) (Petit *et al.*, 1996, Freyd *et al.*, 1990) The HMGA2/LPP fusion protein is a result of a truncation of the HMGA2 protein through deletion of the C-terminal domain and the 3'UTR. This product was fused to the three LIM domains of the LPP protein due to deletion up to either intron 6 (all three LIM domains with part of proline rich domain attached) or intron 8 (two LIM domains attached) (Petit *et al.*, 1996). The second protein fusion partner to be identified for lipomas involving *HMGA2* was that of the lipoma HMGA2 fusion partner (*LHFP*) (Petit *et al.*, 1999). *LHFP* is located at chromosomal region 13q12 and represents a 200 amino acid long protein whose functions remain unknown. The fusion involves the first three "AT hook" binding domains of HMGA2 and the last 69 amino acids of LHFP (Petit *et al.*, 1999).

The truncation of HMGA2 rather than the fusion of functional domains from other proteins have been shown to be important in tumor formation. In lipomas a rearrangement involving *HMGA2* and the G-protein coupled receptor gene *RDC1* occurred at t(2;12)(q35-37;q13-15) (Broberg *et al.*, 2002). The fusion entailed the joining of the first three "AT hook" domains of HMGA2 with only one amino acid from RDC1 (Broberg *et al.*, 2002). This one amino acid was due to a stop codon downstream of the RDC1 breakpoint. The resulting fusion protein brought about mainly a truncation of the HMGA2 protein, no additional function was gained from the fusion as the HMGA2 protein acquired only one additional amino acid (Broberg *et al.*, 2002). This demonstrated that the truncation of the HMGA2 alone could cause the same phenotype as seen when the protein is fused with another protein partner.

Other tumors have arisen from rearrangements involving HMGA2 and include lung harmatomas, soft tissue chondromas and uterine leimyomas (Stenman, 2005). No common functions have been identified to link the fusion partner with the type of tumor formed. Lipomas, for example are caused by fusion of different protein partners to HMGA2 but cause a single phenotype (Petit *et al.*, 1996, 1999, Broberg *et al.*, 2002). The protein partner LPP fused to HMGA2 has also been discovered in other benign tumors as well as lipomas (reviewed in Stenman, 2005). Fusions with LPP have been recognized in pulmonary chondroid harmatomas and in soft tissue chondromas (reviewed in Stenman, 2005). These findings again point to the truncation of HMGA2 rather than the fusion to another protein as the functionally significant event in the formation of a particular tumor. The tissue type that the truncated protein is expressed in may then result in the variety of tumors noted.

C. High Mobility Group Protein 2 – Structure and Expression

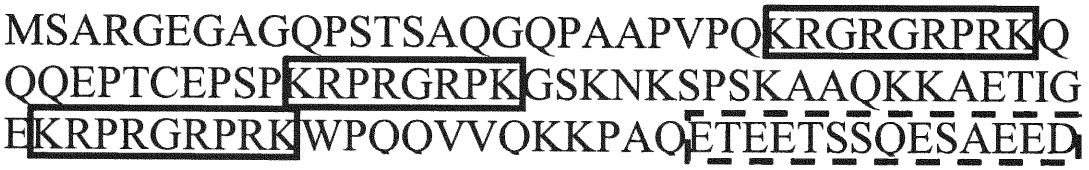
HMGA2 Gene in Humans and Mice and its Expression Pattern

The murine *Hmga2* gene resides on chromosome 10 at the pygmy locus, named due to a null-mutation resulting in a pygmy phenotype (Zhou *et al.*, 1995). The human *HMGA2* gene is located at 12q14-q15, discovered due to multiple rearrangements at this locus in human benign lipomas (Ashar *et al.*, 1996).

The murine HMGA2 mRNA was first evaluated in 1991 (Manfioletti *et al.*, 1991). At this time only 3 kb of the possible 4.1 kb transcript was found. The remaining 1.1 kb fragment was thought to be present at the very far end of the 3' UTR (Manfioletti *et al.*, 1995). The transcript revealed an open reading frame (ORF) that encoded a protein of 108 amino acids, high in levels of alanine, arginine, glutamic acid, glutamine, lysine, proline and serine. The protein also contained a single tryptophan residue allowing for spectroscopic analysis. Using this technique the molecular weight of the protein was found to be 11,977 Da (Manfioletti *et al.*, 1991) (Fig. 1). The cDNA was then cloned and exposed a gene over 50 kb in length. The gene was found to contain 5 exons and 4 introns, each intron/exon junction followed the splice junction GT/AG rule (Manfioletti *et al.*, 1995, Breathnach *et al.*, 1978). Each of the exons was discovered to encode for a separate functional domain. Exon 1 encodes the 5' UTR, the ATG transcription start codon as well as amino acids 1 through 37, which includes a single DNA binding domain. The second exon encodes amino acids 38 to 66 and includes another DNA binding domain. Exon three gives rise to a third DNA binding domain (amino acids 67-83) and the fourth exon transcribes a short linker domain (amino acids 84-94). Finally the fifth exon encodes the C-terminal domain (amino acids 95-108) and the 3' UTR

(Manfioletti *et al.*, 1995). The total length of cDNA is 4655 bp in length, where the 5' UTR is 1351 bp (Manfioletti *et al.*, 1995). This pattern of exons is also observed in the human cDNA, which was first cloned from a human hepatoma cell line, PLC/PRF/5 (Chau *et al.*, 1995, Giancotti *et al.*, 1991). The cDNA contains 5 exons over a span of more than 60 kb that gives rise to a 4.1 kb transcript; the transcript size is found to be identical in size to the murine transcript (Ashar *et al.*, 1996). The human cDNA was first found to contain an ORF of 330 bp, a 5'UTR of 812 bp and a 58 bp 3' UTR (Patel *et al.*, 1994). This was later increased to 854 bp of 5' UTR and 2919 bp of 3' UTR (Ashar *et al.*, 1996). Intron 3 for both human and mice is unusually long (>25 kb in humans), and as a result is the site for many translocation events (Ashar *et al.*, 1996).

Figure 1: Amino acid sequence of murine HMGA2.



The three "AT hook" DNA binding domains are boxed with solid lines. The C-terminal domain is boxed with dashed lines.

Murine *Hmga2* 5'UTR

Promoter analysis of the 5' UTR for both the murine and human genes revealed no TATA or CAAAT box (Ashar *et al.*, 1996, Rustighi *et al.*, 1999). The murine 5' UTR revealed two transcription initiation sites as well as binding sites for many transcription factors. The downstream initiation site provides the basal activity for the gene. It is characterized by a polypyrimidine/polypurine (ppyr/ppur) tract and is located around 83 bp upstream of the start site (Rustighi *et al.*, 1999, Schiltz *et al.*, 2003). Within this tract

are binding sites for the transcription factors specificity protein 1 (Sp1) and Sp3. Near the Sp1 site is the sequence TGGC that is a half site for CTF/NF-1 (CCAAT-binding transcription factor/nuclear factor 1). Sp3 has been shown to enhance the activating ability of Sp1 but cannot activate transcription itself (Rustighi *et al.*, 1999). The ppyr/ppur tract has been shown to adopt non-B-DNA conformations resulting in melting of the DNA in this area (Kohwi and Kohwi-Shigematsu, 1993) This allows the binding of the single strand binding protein, polypyrimidine tract binding protein (PTB) (Rustighi *et al.*, 2002). The binding of the PTB maintains the open conformation of DNA aiding the entry of the RNA polymerase. A balance between Sp1 and PTB binding occurs depending on the conformation of the DNA and the stresses, such as supercoiling (Rustighi *et al.*, 2002). Sites for the binding of other transcription factors such as ATF, Ets (epithelium specific) and E2F (family of transcription factors in higher eukaryotes) are also present in this minimal promoter. There is in addition a conserved site for Huntingtin interactin protein 1 (HIP1) binding, 3 bp from the start site (Zhou *et al.*, 1996). An upstream initiation site is also present and is located -8.1 to -3.7 kb upstream of the start site (Schiltz *et al.* 2003). This upstream site is thought to regulate the spatial and temporal expression of the gene (Schiltz *et al.*, 2003).

Human *HMGA2* 5' UTR

The human *HMGA2* 5' UTR is similar to the murine 5' UTR as there are two separate transcription initiation sites (Patel *et al.*, 1994, 1999, Chau *et al.*, 1995). The start site is located within a GC rich region next to an E box (CACGTG) that cooperatively binds the transcription factors transcription factor II-1 (TFII-1) and upstream stimulatory factor (USF) (Ashar *et al.*, 1996). A continuous tract of $(CT)_{28}$

located at -252 is interrupted by a single GT and omission of a single C. There is another tract, $(CA)_{19}$ and a GC box that is located 200 bp upstream of the start site (Patel *et al.*, 1994, Chau *et al.*, 1995). The GC box is a site for binding of transcription factor Sp1 and is the positive regulatory element for the transcription initiation site (Chau *et al.*, 1999). Many other binding sites for transcription factors are present throughout the 5' UTR including AP2, NFκB, Myc, still life protein (SIF) and polyoma virus enhancer A-binding protein (PEA) (Ashar *et al.*, 1996). Another promoter is thought to exist between -1125 and -868 upstream of the start site (Chau *et al.*, 1999).

Human and Murine 3' UTR

The 3' UTR has been shown to be important in regulatory control. The fact that the truncation of the gene, including the 3' UTR results in benign tumors, suggests that the gene may be under regulatory control within the 3' UTR (Borrmann *et al.*, 2001) Upon truncation of the gene at the 3' UTR, activity using a luciferase assay increases, which suggests that the gene may be under negative control via this region (Borrmann *et al.*, 2001). In fact four-tandem repeats of GGGGT followed by nine G's (one A interruption) and two more tandem GGGGT repeats, have been found in the human 3' UTR which are important in regulation (Chau *et al.*, 1995).

In humans three possible polyadenylation sequences have been identified (Chau *et al.*, 1995). In the mouse two possible polyadenylation sequences have been found, one 18 bp downstream of the poly(A) tail and the second 75 bp further downstream of the first. There is a preference for the first signal due to a TGTGTTCA sequence 31bp downstream of this first signal. Together these two sequences are important for an effective 3' terminus of the mRNA (Manfioletti *et al.*, 1995).

Transcription of the gene is downregulated by histone deactylation inhibition using the inhibitor trichostatin A (TSA). TSA decreases the expression of protein as well as mRNA and therefore the regulation of the gene was determined to occur at the post-transcription level (Ferguson *et al.*, 2003). *Hmga2* is also reduced by regulation of alpha-fetoprotein/mitogen activated or extracellular signal-regulated protein kinase/extracellular signal-related kinase (Raf/MEK.ERK) signaling pathway (Li *et al.*, 1997). This pathway is involved in cellular proliferation and oncogenesis. Persistent activation of *Raf-1* produces a delayed increase in *Hmga2* transcription of roughly 2 hours. This is in keeping with *Hmga2* being a delayed early response gene (Li *et al.*, 1997).

Hmga2 expression in mice has been observed in embryos from 10.5 to 15.5 days postcoitum (d.p.c) (Zhou *et al.*, 1996). Expression after this time-point dramatically decreases, and in newborns is observed only within the stomach and large intestine. This expression is drastically reduced compared to fetal expression levels (Zhou *et al.*, 1996). In mature adult mice the protein was undetectable in all non-tumor tissues (Zhou *et al.*, 1996). Expression in the brain and spinal cord are limited and faint at all time points (Hirning-Folz *et al.*, 1998). The expression in brain is observed only in the ventricularus lateralis of the telencephalon at stages 12.5 and 14.5 d. p. c. and the ventricular lining of the cerebellum at stage 12.5 d.p.c. (Hirning-Folz *et al.*, 1998). This limited expression in the brain points to a reason why the pygmy mice (null *Hmga2* expression) are 50 % reduced in body weight yet the brain region is largely unaffected (Zhou *et al.*, 1995). In humans the expression is similar to that of the mouse embryo. *HMGA2* is not expressed in most adult human tissues (Rogalla *et al.*, 1996). The expression has been seen in tumor tissues such as those of the hepatoma cell lines PLC/PRF/5, Hep G2 and Hep 3B (Patel *et*

al., 1994). This expression indicated that the aberrant expression of *Hmga2* is involved in tumorigenesis.

HMGA2 Protein Structure

As previously described HMGA2 belongs to the high mobility group protein A family. HMGA2 itself has a molecular weight of around 12 kDa (Manfioletti *et al.*, 1991). The protein contains 23 % basic (arginine and lysine) and 11 % acidic (aspartic acid and glutamic acid) residues. However, when examining the individual domains the charge distribution is uneven. The "AT hook" domains contain 68 % basic amino acids and the C-terminal domain contains 58 % acidic amino acids (Fig. 1) (Manfioletti *et al.*, 1991). This charge distribution allows the protein to adopt unique structures upon binding to DNA or other proteins.

The protein contains three "AT hook" DNA binding domains that bind to the minor groove of AT rich regions of DNA (Reeves and Nissen, 1990). All "AT hook" DNA binding domains have the consensus sequence PRGRP and are then characterized into three types due to the flanking regions surrounding the consensus sequence (Aravind and Landsman, 1998, Dragon *et al.* 2003). Type I "AT hook" DNA binding domains are characterized by predominantly basic residues C-terminal to the core consensus sequence. A glycine residue is also the most likely amino acid to be present two positions downstream of this consensus sequence (Aravind and Landsman, 1998). Type II domains characteristically have a high probability of a lysine residue rather than a glycine two positions downstream of the consensus sequence (Aravind and Landsman, 1998). Type III domains have features of both Types I and II. It is similar to Type II in that there is a high probability of the presence of a lysine residue two positions downstream of the

PRGRP core sequence. However, the Type III "AT hook" DNA binding domains also have a high presence of basic residues much like the Type II domains (Aravind and Landsman, 1998). For the "AT hook" DNA binding domains of HMGA2 the second "AT hook" domain would therefore be Type I and the first and third "AT hook" domains Type II.

The HMGA2 protein itself has been shown to have no discernable structure (Lehn *et al.*, 1988, Huth *et al.*, 1997) The protein is mainly an extended structure with roughly 20% β-sheet content (Baez and Leng, unpublished results). Nuclear magnetic resonance (NMR) analysis has shown that upon binding to the minor groove of AT rich DNA, the "AT hook" domain forms a structured conformation (Huth *et al.*, 1997). When the "AT hook" DNA binding domain interacts with the minor groove of AT DNA it forms a C-shaped structure (Huth *et al.*, 1997). The central three amino acids of the consensus PRGRP adopt an extended structure with the arginine residues interacting with the DNA bases deep in the helix. The flanking proline residues are considered to provide support to maintain the structure (Huth *et al.*, 1997). Other residues flanking the consensus sequence could also have loose contacts with the DNA, adding further support to the binding (Geierstanger *et al.* 1994).

Maher and Nathans reported that high affinity binding requires multivalent attachment using two or three of the "AT hook" DNA binding domains spaced suitably apart (Maher and Nathans, 1996). They found that two AT rich DNA sites spaced less than 8 bp apart could be utilized as a high affinity-binding site for one HMGA protein. Spacing between the tracts any more than 8 bp would result in two low affinity-binding sites. Although HMGA proteins are flexible and able to bind to AT tracts with different

length intervals the maximum length of flexibility is 8 bp (Maher and Nathans, 1996). Two models have been proposed for the binding of a single protein to multiple AT tracts. The first model, termed the flexible model, involves the use of the flexible polypeptide chain between the "AT hook" domains. In this model the polypeptide chain allows the "AT hooks" to bind to tracts of different spacing up until the maximum flexibility of the polypeptide chain (Maher and Nathans, 1996). The second model, the alternate model proposed that combinations of any two of the three "AT hooks" are involved in the binding of differently spaced AT tracts (Maher and Nathans, 1996).

The binding site length of a single "AT hook" is 5 bp long and as the protein contains three such domains, the binding site length for the entire protein was determined to be 15 bp long (Cui *et al.*, 2005). The HMGA2 protein has been shown to bind with high affinity to both poly(dA-dT)$_2$ and poly(dA)poly(dT) (Cui *et al.*, 2005). The binding energetics of these interactions is dissimilar. The binding of both the AT hook and the entire HMGA2 protein to poly(dA-dT)$_2$ is enthalpy driven, whereas the binding to poly(dA)poly(dT) is entropy driven (Cui *et al.*, 2005). This difference is due to enthalpy/entropy compensation where HMGA2 binding causes release of more water molecules upon binding to poly(dA)poly(dT) than upon binding to poly(dA-dT)$_2$ (Cui *et al.*, 2005). The large change in heat capacity suggests that changes in hydration, as well as protein folding due to DNA binding or charge-charge interactions could be important factors (Cui *et al.*, 2005).

These "AT hook" domains have also been found in other species such as *Chironomus tentans* (Claus *et al.*, 1994) and in other types of proteins, for example the D1 protein in *Drosophila melanogaster* (Ashley *et al.* 1989) and muNTS proteins in mice

(Wegner *et al.*, 1989). Proteins can artificially be formed by combining multiple copies of the "AT hook" DNA binding domains into a single protein; these proteins are termed multiple "AT" hook (MATH) proteins (Strick and Laemmli, 1995). MATH proteins have been shown to attach to scaffold attachment regions (SARs), which are frequently associated with enhancer elements (Girard *et al.*, 1998) These MATH proteins were shown to displace scaffold proteins and ultimately prevent chromosome assembly (Stick and Laemmli, 1995).

The earliest discovered HMGA protein known as α-protein was demonstrated as binding to the minor groove of 5-6 bp of AT rich DNA with wide sequence variations (Solomon *et al.*, 1986). Solomon *et al.* indicated that α-protein may therefore recognize some conformational aspect of the minor groove of AT DNA rather than its exact sequence (Solomon *et al.*, 1986). Further studies comparing binding of free DNA against that of DNA wrapped around a nucleosome indicated that the preferred binding site for HMGA proteins altered when the DNA conformation changed (Reeves and Wolffe, 1996). In fact, when α-satellite DNA is wrapped around a nucleosome, the AT tracts line up in phase to allow for HMGA protein binding. Satellite DNA is composed of long tandem repeats and proper spacing of AT tracts can phase to be ideal high affinity binding sites for HMGA proteins (Strauss and Varshavsky, 1984). Binding of HMGA proteins on DNA wrapped around a nucleosome results in an alteration of the number of base pairs per turn of the DNA in the area of protein binding (Reeves and Wolffe, 1996). Other structures on which HMGA proteins bind to include, four way junctions, such as those of Holliday junctions (Reeves and Wolffe, 1996). Competition for binding to these junctions occurs between HMGA proteins, HMGB1 and the histone H1. All three

proteins bind to the center of the junction; however, HMGA has the highest affinity for the junction and can therefore displace both H1 and HMGB1 from the DNA (Reeves and Wolffe, 1996). This "AT hook" is also able to confer its binding ability on chimeric proteins. Domain swapping experiments using HMGB1 and the "AT hook" of HMGA1 have shown that the binding abilities of the HMGB1 protein differs from that of the HMGA1 protein upon domain swapping of the DNA binding domains (Banks *et al.*, 1999). This experiment revealed that the "AT hook" domain could function in a protein with very little structure (HMGA) and also in a structurally rigid protein (HMGB1) (Banks *et al.*, 1999).

Binding of the "AT hook" to DNA can induce conformational changes in the DNA double helix. As already discussed, binding of the protein to DNA wrapped around a core particle of a nucleosome can induce conformational changes in DNA at the area of protein binding (Reeves and Wolffe, 1996). Changes in DNA conformation have been shown to be influenced by protein stoichiometry (Slama-Schwok *et al.*, 2000). Binding of HMGA1 to α-satellite DNA at a 1:1 ratio does not introduce any major changes into the DNA structure. However, by increasing the protein to DNA binding ratio to 3:1, the natural bend of the DNA can be increased or reduced depending on sequence (Slama-Schwok *et al.*, 2000). Alterations in DNA supercoiling can be introduced by protein binding, which is dependent upon molar ratios (Nissen and Reeves, 1995). HMGA binding to closed circular DNA at low molar ratios introduce positive supercoils, whereas at higher molar ratios negative supercoils are established. It is believed that at low molar ratios the protein introduces a bend in the DNA, and when the molar ratio increases, the binding causes both bending and unwinding of the DNA resulting in negative supercoil

formation (Nissen and Reeves, 1995). The negative C-terminal domain influences the topology, as the truncated protein (without the C terminal domain) binds with a higher affinity and only introduces negative supercoils into the DNA (Nissen and Reeves, 1995).

Previous Research

Procedures for the purification of both the wild type HMGA2 (wt HMGA2) and the truncated mutant protein HMGA2 (HMGA2Δ95-108) were developed and implemented previously in this lab (Cui *et al.*, 2005). Three regions composed mostly of positively charged amino acids are present (the three "AT hook" regions) towards the N-terminus of the protein while the C-terminus of the protein is mainly negatively charged. These regions can be exploited in the purification process, as the protein is capable of binding to both anion and cation exchange columns such as SP-Sepharose and Q-Sepharose respectively (Cui *et al.*, 2005). The basic procedure involves using 1 mM IPTG to induce protein expression from BLR (DE3) cells with the appropriate plasmid containing the *Hmga2* gene. Protein expression was evaluated using 15% SDS PAGE. Cells were then lyzed and sonicated before being subjected to SP-Sepharose ion exchange chromatography, followed by Q-Sepharose ion exchange chromatography.

The truncated protein HMGA2Δ95-108 lacks the C-terminal domain and would therefore not bind to an anion exchange column such as Q-Sepharose. A DEAE-Sepharose column was used as a replacement for the Q-Sepharose column to remove nucleic acid contaminants. The procedure is rapid and can be completed within three days, producing milligrams of protein. The purity of the protein can be as high as 98% for HMGA2 and 95% for HMGA2Δ95-108 evaluated using 15% SDS PAGE (Baez and Leng, unpublished results). The same procedure used for wtHMGA2, was used to purify

the mutant proteins HMGA2Q37M, HMGA2C41G, HMGA2I71M and HMGA2Q85M in this current research.

Optical properties of both the wild type and truncated proteins were also evaluated. Both proteins contain a single tryptophan residue at position 70 and therefore give a maximum absorbance under ultra-violet light at 280 nm (Mach *et al.*, 1992). The extinction coefficients of both proteins were determined to be 5810 $cm^{-1}M^{-1}$ for both proteins (Gill and von Hippel, 1989). Ionic strength was found to have no effect on the extinction coefficient for either protein (Baez and Leng, unpublished results).

Circular dichroism (CD) analysis was used to study HMGA2's secondary structure (Beaz and Leng, unpublished results). A strong peak at 200 nm was observed and suggests an unordered structure. In addition, this data was analyzed using three CD analysis programs, namely CONTIN, CDSSTR, and SELCON3 (Sreerama and Woody, 2000). The cumulative data from these programs suggests that HMGA2 may contain about 15-20% β-sheet conformation (Beaz and Leng, unpublished results).

Sedimentation velocity analysis was implemented to analyze the sedimentation velocity, s, and the molecular weight of the HMGA2 protein (Grievink *et al.*, 1974) The sedimentation velocity was found to be 1.711 s (95% confidence interval: 1.708 s-1.715 s) and the molecular weight was calculated as 23.5 kDa (Baez and Leng, unpublished results). The molecular weight determined by this analysis is roughly double the expected molecular weight of 12 kDa, calculated using the HMGA2 amino acid sequence (Beaz and Leng, unpublished results). Therefore this suggests that the protein exists as a homodimer when free in solution.

Gel filtration studies were performed in order to calculate the Stokes radius (R_s) of both the wild type and truncated HMGA2 proteins (Horiike *et al.*, 1983). When comparing the calculated Stokes radius for the wild type HMGA2 protein to a monomeric globular protein, ribonuclease A, the Stokes radius is much larger. Ribonuclease A is 13.7 kDa and therefore understood to be similar to HMGA2 for comparison (Smyth *et al.*, 1963). The Stokes radius of 30.2 Å for HMGA2 compared with 16.4 Å for ribonuclease A suggests that the wtHMGA2 protein exists as a homodimer, a non-globular protein or both. This is due to the increased size when compared to a monomeric, globular protein. The estimated molecular weights are 45 kDa for wtHMGA2 and 33.9 kDa for the truncated HMGA2 protein, again suggesting a non-monomeric protein (Beaz and Leng, unpublished results).

Preliminary experiments were also performed using the chemical cross-linkers EDC and DMS. EDC is a zero-length cross-linker while DMS is a homobifunctional imidoester cross-linker that only allows binding between primary amines. Experiments involving wtHMGA2 and EDC show a homodimer formation as well as higher homo-oligomers within an SDS polyacrylamide gel. The monomer also appears to run at a faster rate than the control protein, which suggests that the C-terminal domain may interact with other regions within the same protein. Even though the molecular weight would be the same, the interaction would serve to alter the shape of the protein, thus allowing it to run faster on an SDS polyacrylamide gel (Beaz and Leng, unpublished results). Experiments involving the truncated HMGA2 protein and EDC did not show homodimer or higher oligomer formation, suggesting that the C-terminal tail may play an important role in homodimer formation (Beaz and Leng, unpublished results). The

proteins were also tested using the cross-linker DMS. The same basic pattern as found with EDC was again observed with DMS. The cross-linking ability of DMS with HMGA2 was lower than that found with EDC. The truncated HMGA2 protein again showed no homodimer formation. This once more suggests a role for the C-terminal in homodimer formation (Beaz and Leng, unpublished results).

Hypothesis and Objectives

In this thesis, our central hypothesis is that HMGA2, an intrinsically unstructured protein, is a homodimer both as a free protein and upon binding to DNA. The electrostatic interaction between the positive charged "AT hooks" and the negatively charged C-terminus is the mechanism of the HMGA2 homodimer formation.

The specific aims for this project are to:

- Determine that the HMGA2 protein exists as a homodimer when free in solution
- Identify the structural elements involved in the HMGA2 homodimer formation
- Determine the stoichiometry of binding between HMGA2 and duplex DNA containing either a single or multiple AT-rich sequences.

Significance of Research

HMGA2 is a medically important protein. Aberrant expression of the full-length, truncated or fusion protein causes benign or malignant tumors. This study involves determining the biochemical and biophysical properties of the protein. The research is significant since we discovered that HMGA2 is a homodimer and the homodimer formation requires the negatively charged C-terminus. This may explain why over-expression of the truncated HMGA2 protein without the C-terminus causes tumor formation. These findings may be utilized in order to develop anticancer drugs, which

target the functionally important structural properties of the protein. The HMGA2 protein may have a potential forensic application in determining fetal age for the legal implications of abortion or for identifying fetal tissue. The techniques used in this research also have a role in Forensic Science in areas which include Forensic Biology and Forensic Chemistry.

CHAPTER II

MATERIALS AND METHODS

Materials

Phenylmethylsulphonylfluoride (PMSF), dithiothreitol (DTT), 2-Morpholinoethanesulfonic acid (MES), sodium dibasic, sodium monobasic cyanogens bromide, formic acid, tricine, acetic acid, and ethidium bromide were commercially obtained from Sigma-Aldrich. Kanamycin, lysozyme, dimethyl sulfoxide (DMSO), Magnesium chloride, agarose, Tris (hydroxymethyl) aminomethane, ethylenediaminetetraacetic acid (EDTA), Bis acrylamide, acrylamide, ammonium phosphate, potassium monobasic, sodium dodecyl sulphate (SDS) and N,N,N,N - Tetramethyl-Ethylenediamine (TEMED) were commercially obtained from Omnipure, EM Science, EMD Chemicals Inc. (Gibbstown, NJ). The QIAquick® gel extraction kit, QIAquick® PCR purification kit and QIAprep® spin miniprep kit were purchased from Qiagen (Valencia, CA). New England biolabs (Beverly, MA) commercially provided all restriction enzymes, ligase, bovine serum albumin (BSA), and all ladders and buffers provided for these enzymes. Q-Sepharose FF, SP-Sepharose FF, dNTPs, G-50 Superfine Sephadex, Sephacryl S-100 HR, low molecular weight gel filtration calibration kit and poly(dA-dT)$_2$ were purchased from Amersham Biosciences (Piscataway, NJ). Pierce (Rockford, IL) commercially supplied cross-linkers disuccinimidyl suberate (DSS) and 1-ethyl-3-(3-dimethylaminopropyl)-car-bodiimide (EDC) as well as sulpho-N-hydroxysuccinimide (Sulpho-NHS). Glycerol, potassium dibasic, sodium chloride and isopropyl-b-D-thiogalactopyranoside (IPTG) were purchased from Fisher (Fairlawn, NJ).

MWB Biotech, Inc. (High Point, NC) commercially provided all primers and oligos, and pET30a was purchased by Novagen (Madison, WI). Pfu DNA polymerase was purchased from Stratagene Corporate (La Jolla, CA) while yeast extract and tryptone were purchased from BD (Sparks, MD). A peptide containing the negatively charged C-terminus (H-CETEETSSQESAEED-OH)), was custom-synthesized by Advanced ChemTech, Inc. All dialysis tubing was commercially obtained from BioDesign Inc. of New York (Carmel, NY) and the DNA stain Sybr Gold® and Tetramethylrhodamine-5-maleimide (TMR) was commercially obtained from Molecular Probes (Eugene, OR)

Site-Directed Mutagenesis

The plasmid pMGM1, which contains a full length murine *Hmga2* gene, was created previously in this lab using PCR based site-directed mutagenesis and was, described previously (Cui *et al.*, 2005).

A brief outline of the PCR site-directed mutagenesis used in this study is as follows. For each mutant created, a total of three PCR reactions are required in addition to four primers. A 27mer oligonucleotide, namely FL#20 was composed and created. This oligonucleotide is complimentary to the first 20 bases at the 5' end of the murine *Hmga2* gene. Also, at the 5' ending of this oligonucleotide are 7 bases, which contain an *Nde*I restriction enzyme site. In addition, another 27mer oligonucleotide was composed and created, namely FL#21. This oligonucleotide is complimentary to the final 15 bases at the 3' end of the murine *Hmga2* gene. Also at the 5' end of this oligonucleotide is the final 15 bases that contain a *Xho*I restriction enzyme site along with a TAG stop codon. The final two primers contain the required mutation in addition to complimentary bases either side of the mutation (Table 1).

The DNA template pMGM1 was used for the first 2 PCR reactions. To create pLKE4, primers FL#20 and FL#102 were used in PCR 1 and primers FL#21 and FL#101 were used in PCR 2. The reactions were carried out using an MWG-Biotech Thermocycler primus96. The polymerase used was *Pfu* Turbo DNA polymerase, which is thermostable and has proofreading capabilities to avoid unnecessary mutations. A total reaction volume of 50 µl was used for each PCR reaction that contained 100 ng of template DNA (pMGM1), 20 pmol of each primer, 50 µM of each dNTP, 1× BSA, 2.5 U of *Pfu* polymerase and 1× *Pfu* reaction buffer supplied by the vendor. The reaction was initially heated to 95°C for 3 minutes. This was followed by a denaturing step at 95°C for 30 seconds, annealing at 55°C for 1 minute and finally an extension step at 72°C for 1 minute, these three steps were cycled 25 times. A further extension at 72°C for 10 minutes was performed at the end of the procedure to ensure extension of the entire region. The PCR products from both reactions were analyzed by running both a 2 % agarose gel and using 6 % polyacrylamide gel electrophoresis (PAGE). The products were purified from the 2 % agarose gel using a Qiagen gel purification kit. A third PCR was performed using the products of the first two PCR reactions as templates and primers FL#20 and FL#21 that provide the complimentary sequence to the ends of the *Hmga2* gene. The conditions for the third PCR reaction were as described previously. The product of the third PCR reaction was purified using Qiagen PCR purification kit. The purified product was digested using the restriction enzymes *Nde*I and *Xho*I and the digested product was ligated into the vector pET30a at the *Nde*I and *Xho*I sites. The plasmid was sequenced at MWG Biotech lab to confirm the desired mutation. This plasmid was constructed in order to over-express the mutant protein HMGA2Q85M

where the glutamine amino acid residue at position 85 was replaced with a methionine residue.

Three additional mutant plasmids were constructed, pLKE5, pLKE6 and pLKE7, which produced proteins, HMGA2I71M, HMGA2Q85M and HMGA2C41G respectively. The template DNA pMGM1 was used for each mutant along with specific primers for each mutation (Table 2). The PCR conditions used were the same as described above. All PCR products were digested using the restriction enzymes *Nde*I and *Xho*I and were subsequently ligated into the *Nde*I and *Xho*I sites of vector pET30a.

Table 1: PCR Based Site Directed Mutagenesis Primers

FL#	Sequence
20	5'-CAGGATCCATATGAGCGCACGCGG TGAGGGC-3'
21	5'-CATCTCGAGCTAATCCTCCTCTGCGGA-3'
101	5'-CCCAGGAAGCAGCAGATGGAGCCAACCTGTGAGC-3'
102	5'-GCTCACAGGTTGGCTCCATCTGCTGCTTCCTGGG-3'
103	5'-GAAGAAAGCAGAGACCATGGGAGAAAAACGGCCAAG-3
104	5'-CTTGGCCGTTTTTCTCCCATGGTCTCTGCTTTCTTC-3'
105	5'-CAGACCTAGGAAATGGCCAATGCAAGTCGTTCAGAAGAAGCC-3'
106	5'-GGCTTCTTCTGAACGACTTGCATTGGCCATTTCCTAGGTCTG-3'
206	5'-CAGCAGCAAGAGCCAACCGGTGAGCCCTCTCCTAAGAGAC-3'
207	5'GTCTCTTAGGAGAGGGCTCACCGGTTGGCTCTTGCTGCTG-3'

Table 2: Mutants created using PCR site-directed mutagenesis. The plasmid name, mutation, primers and protein name are given for each mutant produced.

Plasmid Name	Mutation	Primers	Protein Name
pLKE4	Q37 to M37	FL#20, 21, 101, 102	HMGA2Q37M
pLKE5	I71 to M71	FL#20, 21, 103, 104	HMGA2I71M
pLKE6	Q85 to M85	FL#20, 21, 105, 106	HMGA2Q85M
pLKE7	C41 to G41	FL#20, 21, 206, 207	HMGA2C41G

Purification of HMGA2 Mutants

The HMGA2 protein has positively charged 'AT-hook' regions and a negatively charged C-terminal domain. These properties allow the protein to be purified by a simple procedure. A combination of using the cation exchange resin, SP-Sepharose (binding the positively charged region) and the anion exchange resin, Q-Sepharose (binding the negatively charged C-terminus) results in producing pure HMGA2 protein (Cui *et al.*, 2005).

The mutant plasmids described above were transformed into *Escherichia coli (E. coli)* host strain BLR (DE3) in order to over-express the protein. The transformed BLR (DE3) strain was then grown overnight in Luria-Bertani (LB) broth containing 50 μg/ml of the antibiotic kanamycin. The resultant cell stock was transferred to Terrific Broth (TB) containing 50 μg/ml of kanamycin. Cells were grown by shaking at 37°C and monitored by measuring OD_{595} using an Amersham Ultraspec 2000 UV-VIS spectrophotometer. Measurements were recorded every hour until the OD_{595} reached 0.6-0.7 and the protein was then expressed by adding 1 mM of isopropyl-β-D-

thiogalactopyranoside (IPTG). The cells were incubated by shaking at 37°C for a further 3 hours and the OD_{595} was monitored every hour. The cell stock solution was centrifuged at 4°C, 4,000 rpm for 25 minutes. The supernatant was discarded and the pellet air-dried. The cell pellet was subsequently resuspended in 5 ml per gram of ice-cold lysis buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 0.5 mM PMSF, 0.1 mM DTT) and 1 mg/ml lysozyme was added. The cell solution was incubated on ice for 60 minutes before snap-freezing in liquid nitrogen and storing in a –80°C freezer overnight. The frozen cells were thawed on ice and the salt concentration was altered to 1 M by the addition of solid NaCl. The solution was sonicated on ice at 300 W, 8 times for 10 seconds with a 10 second interval between each sonication or until the viscosity had significantly decreased. The resultant solution was centrifuged at 4°C, 16,000 rpm for 20 minutes. The supernatant was saved while the pellet was discarded. The supernatant was dialyzed against buffer 1 (50 mM sodium phosphate, pH 8.0, 10 % glycerol, 0.5 mM PMSF, 0.1 mM DTT) plus 200 mM NaCl overnight at 4°C. A 40 ml SP-Sepharose column was equilibrated with 500 ml buffer 1 plus 200 mM NaCl and the dialyzed solution was loaded onto the column. The column was washed with 120 ml buffer 1 plus 300 mM NaCl. Applying a salt gradient using 300 ml buffer 1 plus a gradient of 300 mM to 800 mM NaCl was then used to elute the HMGA2 protein. Peak fractions were first identified by measuring OD_{280} and were confirmed electrophoretically using 15 % SDS-PAGE. Peak fractions were pooled and dialyzed against buffer 2 (50 mM sodium phosphate pH 8.0, 10 % glycerol, 0.5 mM PMSF, 0.1 mM DTT) as well as 20 mM NaCl, overnight at 4°C. A 40 ml Q-Sepharose column was equilibrated using 500 ml buffer 2 plus 20 mM NaCl. The dialyzed solution was then loaded onto the Q-Sepharose column. A salt

gradient of 20 mM to 300 mM NaCl in 300 ml of buffer 2 was used to elute the HMGA2 protein. Peak fractions were identified by measuring OD_{280} and were confirmed electrophoretically using 15 % SDS-PAGE. Peak fractions were pooled and dialyzed against buffer 1 plus 200 mM NaCl overnight at 4°C. A small 5 ml SP-Sepharose column was used to concentrate the protein. The column was equilibrated with 100 ml of buffer 1 as well as 200 mM NaCl. The dialyzed fractions were loaded onto the column. The protein was eluted using buffer 1 plus 800 mM NaCl collecting fractions of approximately 400 µl. The protein was identified and the concentration determined by measuring OD_{280}. Confirmation of the purity of the protein was determined using 15 % SDS-PAGE.

Chemical Cross-linking

A zero-length cross-linker 1-ethyl-3-(-3-dimethylaminopropyl) carbodiimide HCl (EDC) and an 11.4 Å length cross-linker disuccinimidyl suberate (DSS) were used to cross-link HMGA2 and all the mutant proteins.

1-Ethyl-3-(-Dimethylaminopropyl) Carbodiimide HCl (EDC)

HMGA2 proteins were incubated against MES buffer (100 mM MES, pH 5.5, 50 mM NaCl) overnight. A stock of 100 mM EDC was freshly prepared; 10 mM final concentration of EDC was used along with approximately 50 µM protein in MES buffer. The protein/EDC solution was incubated at room temperature for 2 hours. The reaction was quenched using 100 mM Tris (pH 7.5) and then filtered by G-50 Sephadex filtration equilibrated with MES buffer pH 5.5 to remove excess cross-linker. The resultant cross-linked protein was evaluated using 15 % SDS PAGE. Monomers and dimers of each protein were separated using 15 % SDS PAGE. The dimers were extracted from the gel

by excising the band and placing in dialysis tubing. The protein was removed into 1× SDS PAGE running buffer by the use of 100 V of electricity for a period of 2 hours.

A duplex was formed using oligos manufactured from MWG biotech Inc. The oligo FL123 (5'$G_5CA_{15}CG_5$3') was duplexed with oligo FL124 (5'$C_5GT_{15}GC_5$3') along with 50 mM NaCl by heating a waterbath to 95°C and allowing to cool overnight. The duplex formation was analyzed using 20 % PAGE.

The protein HMGA2C41G and the DNA duplex were dialyzed against a 20 mM sodium phosphate buffer (pH 7.0) plus 180 mM NaCl. A 2:0, 2:1 or 1:1 protein to DNA ratio was used with a 20 μM protein concentration. The reaction was incubated for 30 minutes at room temperature before adding 10 mM EDC and 5 mM Sulpho-NHS. The reaction was then continued for a further 2 hours at room temperature and was stopped by adding 100 mM Tris (pH 7.5) followed by filtration through a Sephadex G-50 column equilibrated with 20 mM phosphate buffer (pH 7.0) and 180 mM NaCl. The reaction was analyzed using 15 % SDS PAGE.

The reaction was proceeded as described above except oligo FL123-124 were replaced by poly(dA-dT)$_2$. The cross-linking reaction was examined using 15 % SDS PAGE

Disuccinimidyl Suberate (DSS)

HMGA2 proteins were incubated overnight in 10 mM sodium phosphate buffer pH 9.0 with 50 mM NaCl. A stock solution of 100 mM DSS was freshly prepared dissolved in DMSO. A final concentration of 10 mM DSS was used to cross-link approximately 40 μM protein in phosphate buffer. The reaction was incubated at room

temperature for 2 hours and quenched by the addition of 120 mM Tris (pH 7.5). The resultant cross-linked protein was evaluated using 15 % SDS PAGE.

The protein HMGA2C41G and the DNA duplex were dialyzed against a 10 mM sodium phosphate buffer (pH 9.0) plus 50 mM NaCl. A 2:0, 2:1 or 1:1 protein to DNA ratio was used with a 20 μM protein concentration. The reaction was incubated for 30 minutes at room temperature before the addition of 10 mM DSS. The reaction then proceeded for a further 2 hours at room temperature and was quenched by the addition of 120 mM Tris (pH7.5). The reaction was analyzed using 15 % SDS PAGE.

The reaction was proceeded as described above except oligo FL123-124 were replaced by poly(dA-dT)$_2$. Also a protein concentration of 23 μM was used instead of 20 μM. The reaction was examined using 15 % SDS PAGE

Cyanogens Bromide Cleavage

All cross-linked and non-cross-linked proteins were dried using a speed-vac at 30°C. One ml of 500 mM cyanogens bromide (CNBr) in 70 % formic acid was added to the dried proteins, the reaction was incubated in a waterbath at 37°C in the dark overnight. The formic acid was evaporated off using a speed-vac at 30°C and the dried pellet was washed 3 times with 200 μl of water. The resultant cleaved proteins were evaluated using 15-20 % Tris tricine SDS-gradient gel electrophoresis (Scagger and von Jagow, 1987).

Labeling of HMGA2 and C-terminal Peptide with Tetramethylrhodamine-5-maleimide (TMR) and Gel Filtration

Tetramethylrhodamine-5-maleimide was used to label both the HMGA2 protein and the C-terminal peptide by tagging the cysteine residue. A concentration of 100 μM of

HMGA2 or the C-terminal peptide was incubated with 200 μM of TMR, 400 μM TCEP, and 50 mM phosphate buffer (pH 7.2) plus 20 mM NaCl for 2 hours at 24°C. Labeled HMGA2 was subjected to purification by running twice through a pre-equilibrated Sephadex G-50 Spin column, equilibrated with 50 mM phosphate buffer (pH 7.2) plus 20 mM NaCl. To purify the CTP-TMR, the labeling mixture was loaded onto a pre-equilibrated SP-Sepharose column (1 ml) and eluted with 50 mM phosphate buffer (pH 7.2) plus 500 mM NaCl. An extinction coefficient of 95,000 $cm^{-1}M^{-1}$ at 541 nm in methanol was used to determine the TMR concentration. Binding between labeled C-terminal and HMGA2Δ95-108 occurred at 24°C for 30 minutes. The resulting mixture was resolved using a Sephacryl S-100 HR filtration column (1 × 50 cm). The column was equilibrated with BPES buffer and gravitational force was used to elute the proteins. The column was pre-calibrated using the protein standards, ribonuclease A (M_r, 13,000), chymotrypsin A (M_r, 25,000), ovalbumin (M_r 43,000), albumin (M_r 67,000) and Blue Dextran 2000. Fractions of volume 534 μl were collected and TMR concentration was determined by UV absorbance.

Gel Mobility Shift Assay

A 0.4 μM DNA duplex FL123-124 was utilized along with the specified concentration of protein; HMGA2, HMGA2Δ95-108 (a version of HMGA2 which lacks the C-terminal domain) or HMGA2C41G. The reaction proceeded in a buffer containing 10 mM Tris (pH 7.5), 0.5 mM $MgCl_2$, 0.1 mM EDTA, 150 μg/μl BSA, 5 % glycerol and either 50 mM (low salt) or 200 mM (high salt) KCl and either with or without 1 μM DTT. The reactions were incubated at room temperature for 30 minutes before being

analyzed on 15 % PAGE (pre-run at 100 V for 2 hours and run at 100 V). The DNA was stained using Sybr Gold® DNA stain and photographed under UV light.

Isothermal Titration Calorimetry

ITC experiments were carried out using a VP-ITC titration calorimeter (Microcal Inc.). Samples were extensively dialyzed against BPE buffer containing 4 mM NaCl. Typically, the titration was set up so that 15 μl of a 75 μM HMGA2 sample (dimer concentration) was injected every 200 seconds, up to a total of 18 injections, into a DNA sample (1.7 ml of 10 μM) in the sample cell. The heat liberated or absorbed with each injection of ligand is observed as a peak that corresponds to the power required to keep the sample and reference cells at identical temperatures. The peaks produced over the course of a titration are converted to heat output per injection by integration and corrected for cell volume and sample concentration. Control experiments were carried out to determine the contribution to the measurement by the heats of dilution arising from (1) protein into buffer, and (2) buffer into DNA. The net enthalpy for each protein-DNA interaction was determined by subtraction of the component heats of dilution.

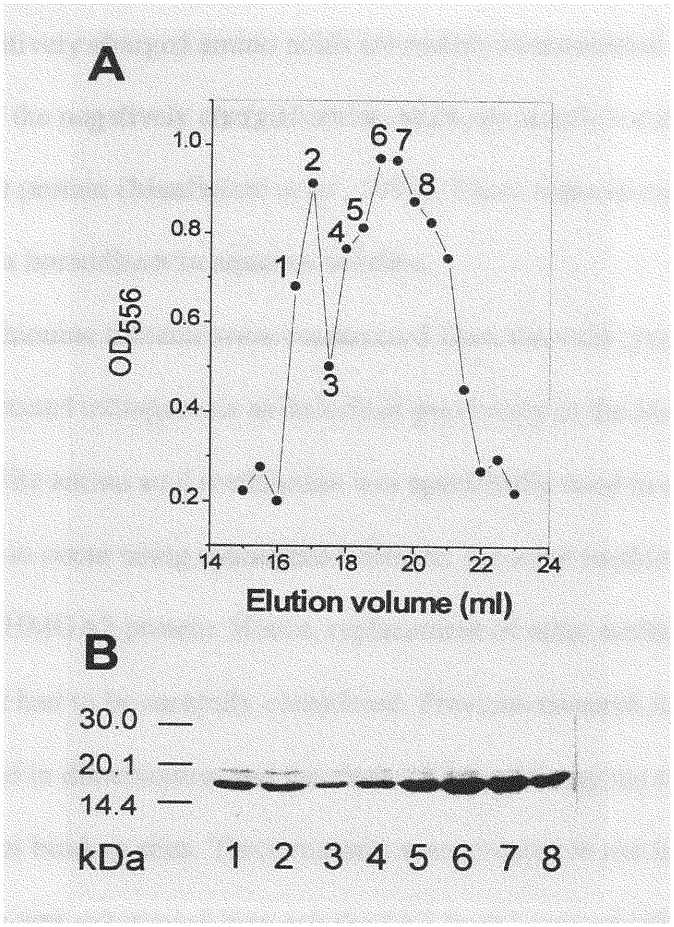
CHAPTER IV

RESULTS

Size Exclusion Chromatography

Preliminary experiments using the chemical cross-linkers EDC and DMS revealed that the wtHMGA2 protein could form homodimers whereas the truncated protein (HMGA2Δ95-108) could not form homodimers (Baez and Leng, unpublished results). We therefore assumed that the C-terminal domain may be involved in the formation of the homodimer. We used tetramethylrhodamine-5-maleimide (TMR), to label a 14 amino acid residue C-terminal peptide (H-CETEETSSQSAEED-OH) (the CTP) to produce the CTP-TMR. The CTP-TMR was incubated with HMGA2Δ95-108 and subjected to a pre-equilibrated gel filtration column. Figure 2 shows the elution profile of the gel filtration experiment. Our results demonstrated that the CTP-TMR was co-eluted with HMGA2Δ95-108. This is shown as the peaks in the graph representing the concentration of the CTP-TMR (Fig. 2A) correspond to increases in concentration in the SDS-polyacrylamide gel representing the concentration of the HMGA2Δ95-108 (Fig. 2B). Interestingly, there are two co-elution peaks (Fig. 2). Possibly, the first peak represents two CTP-TMR molecules binding to one HMGA2Δ95-108 and the second peak represents one CTP-TMR molecule binding to one HMGA2Δ95-108. An alternative possibility would be that the first peak contains one molecule of the CTP-TMR binding to two molecules of HMGA2Δ95-108 and the second corresponds to one molecule of the CTP-TMR binding to one molecule of HMGA2Δ95-108. Further studies are required to determine the binding stoichiometry between the CTP and HMGA2Δ95-108 protein.

Figure 2: The CTP-TMR and HMGA2Δ95-108 co-elution in gel-filtration chromatography.



The CTP-TMR was prepared as described under "Materials and Methods" and incubated with HMGA2Δ95-108 at 24 °C for 30 min in BPES buffer. The CTP-TMR and HMGA2Δ95-108 mixture was then subjected to a Sephacryl S-100 HR filtration column (1×50 cm) equilibrated with BPES buffer. Gel filtration profile of the CTP-TMR binding to HMGA2Δ95-108 was monitored by a graph of OD_{556} versus elution volume (A) and a 15% SDS PAGE gel (B). Lanes 1 to 8 of the SDS-PAGE gel (B) correspond to the fractions 1 to 8 labeled in panel A. Free HMGA2Δ95-108 and the CTP-TMR were eluted at 22 and 30 ml respectively in the column.

Single Methionine Mutants

One unique characteristic of HMGA2 is the charge distribution over the primary structure. The positively charged amino acids are mainly concentrated in the three "AT hook" regions and the negatively charged amino acids are largely located at the C-terminal end of the protein (Manfioletti *et al.*, 1991). These features may allow the protein to exist as a homodimer in aqueous solution.

Single methionine mutants were constructed from the wild-type HMGA2 using PCR based site-directed mutagenesis as described previously in the Materials and Methods section. The amino acid methionine was specifically used in order for a chemical cleavage to occur using cyanogens bromide. No other methionine residues are present within the HMGA2 protein. Hence, replacement of other amino acids to a methionine residue had to be carefully considered. Previous research indicated that the C-terminal is involved in dimerization and the three "AT hook" regions were considered as potential C-terminal binding sites. Three mutants were created in our lab where a methionine residue was substituted between the "AT hook" regions where homodimer formation could occur. In the first mutant, HMGA2Q37M, a glutamine was replaced with a methionine residue as position 37; this is located between the first and the second "AT hook" regions. In the second mutant, HMGA2I71M, an isoleucine is replaced with a methionine at position 71, located between the second and third "AT hook" regions. In the third mutant, HMGA2Q85M, a glutamine is replaced with a methionine residue at position 85, positioned between the third "AT hook" and the C terminal domain.

Three possible models were evaluated in this study as potential homodimer formation scenarios (Fig. 3A, 3B and 3C). These represented the C-terminal interacting

with the first "AT hook" (Fig. 3A(iii), 3B(iii), 3C(iii)), the second "AT hook" (Fig. 3A(ii), 3B(ii), 3C(ii)) or the third "AT hook" (Fig. 3A(i), 3B(i), 3C(i)). All methionine residues could be cleaved using cyanogens bromide. However, methionine residues that are between areas of cross-linking, would not be seen to separate as expected on a polyacrylamide gel. This is because the cross-linking bond would hold the cut fragments together so they would not separate as anticipated. Figures 3A(i), 3B(i) and 3C(i) demonstrate the C-terminal domain interacting with the third "AT hook" region. In this model all three single methionine mutants would appear on an SDS polyacrylamide gel as being cleaved using cyanogens bromide. This is due to the location of the methionine residues in mutants HMGA2Q37M and HMGA2I71M situated away from the areas of chemical cross-linking (Fig. 3A(i) and 3B(i)). For the mutant HMGA2Q85M the methionine residues are located so that when cleaved the homodimer would separate in half as there is no cross-linking between the methionine residues of the two monomers (Fig 3C(i)). Figure 3A(ii), 3B(ii) and 3C(ii) display the homodimer formation occurring between the C-terminal domain and the second "AT hook" region. If this occurred, only the mutant HMGA2Q37M would appear to be cleaved and show two bands on an SDS polyacrylamide gel (Fig. 3A(ii)). One band would represent the majority of the homodimer including the second and third "AT hook" regions of both monomers as well as the C-terminal domains of the monomers. The second band would represent the very N-terminals of the monomers including the first "AT hook" regions and would therefore be of a smaller size (Fig. 3A(ii)). Even though HMGA2I71M and HMGA2Q85M would also be cleaved, the cleavage site within these mutants is surrounded by areas of chemical cross-linking and therefore would appear on a gel as one band (Fig. 3B(ii), 3C(ii)) The

Figure 3: Models of potential homodimer formation.

Figure 3A: Models of potential homodimer formation with HMGA2Q37M mutant protein.

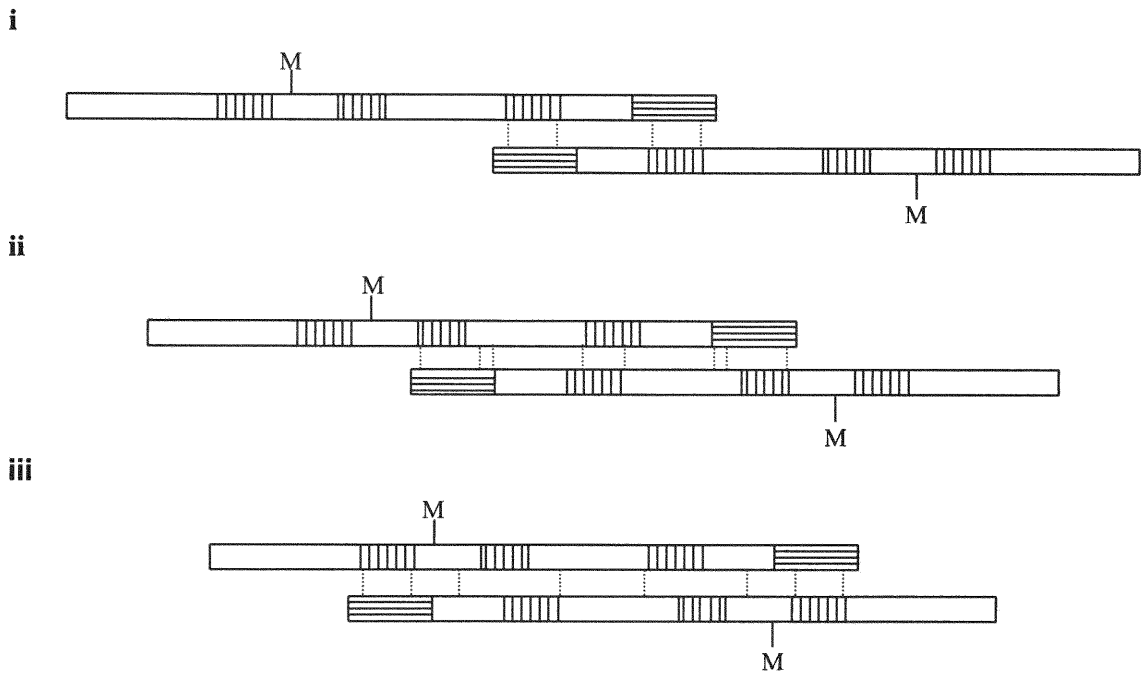


Figure shows C-terminal domain interacting with the "AT hook" DNA binding domains between two HMGA2Q37M monomers. Panel i shows the C-terminal interacting with the third "AT hook" DNA binding domain. Panel ii illustrates the C-terminal interacting with the second "AT hook" DNA binding domain and panel iii shows the C-terminal interacting with the first "AT hook" DNA binding domain. The dotted lines show potential EDC cross-linking. Vertical striped areas represent the three "AT hook" DNA binding domains. Horizontal striped areas represent the C-terminal domain.

Figure 3B: Models of potential homodimer formation with HMGA2I71M mutant protein.

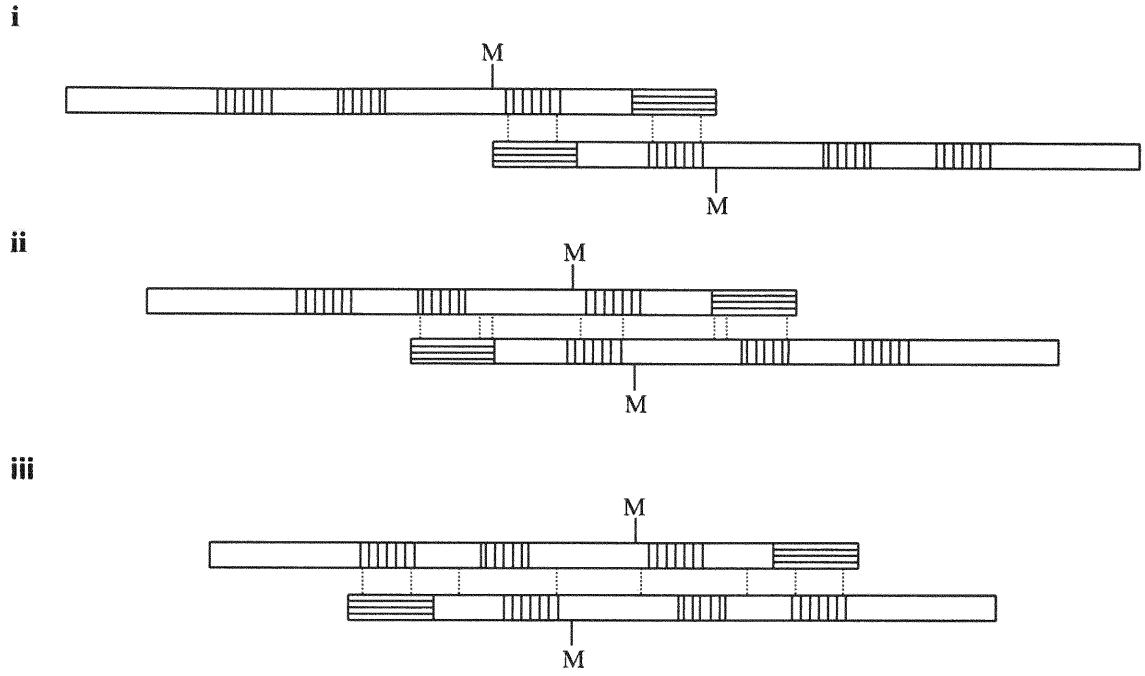


Figure shows C-terminal domain interacting with the "AT hook" DNA binding domains between two HMGA2I71M monomers. Panel i shows the C-terminal interacting with the third "AT hook" DNA binding domain. Panel ii illustrates the C-terminal interacting with the second "AT hook" DNA binding domain and panel iii shows the C-terminal interacting with the first "AT hook" DNA binding domain. The dotted lines show potential EDC cross-linking between the two monomers. Vertical striped areas represent the three "AT hook" DNA binding domains. Horizontal striped areas represent the C-terminal domain.

Figure 3B: Models of potential homodimer formation with HMGA2I71M mutant protein.

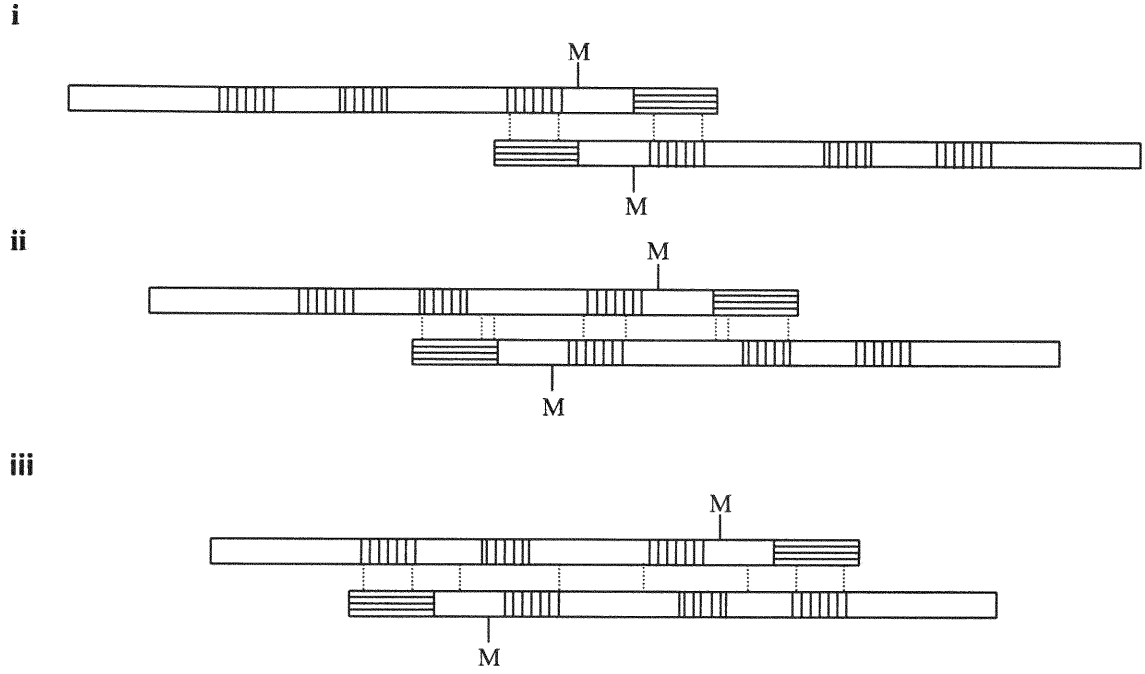
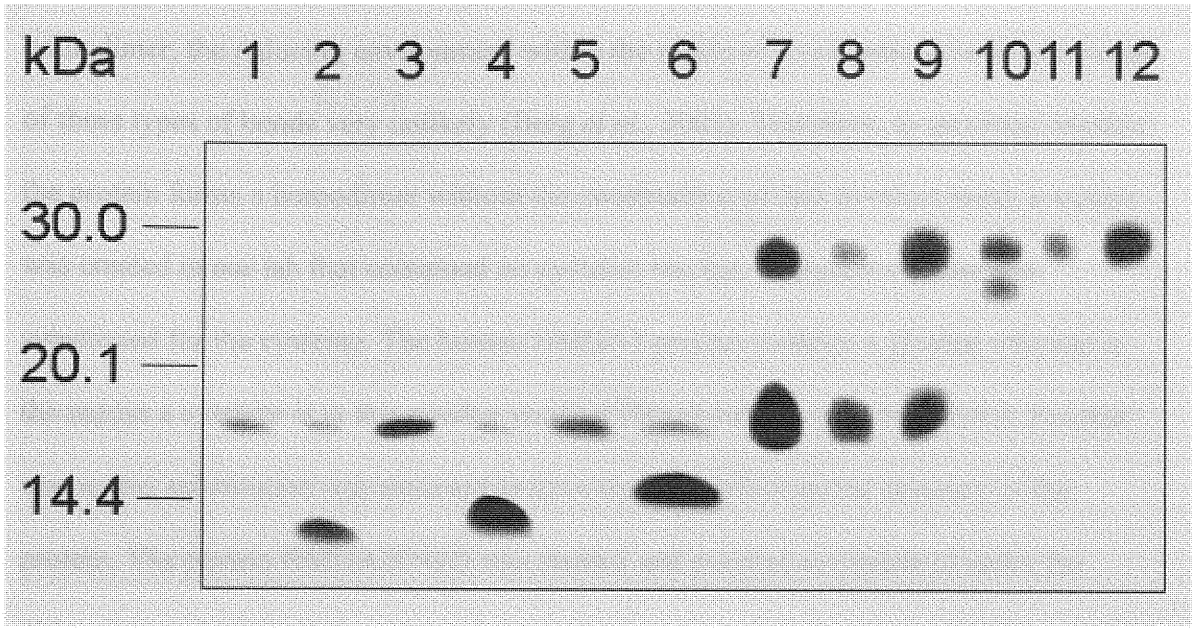


Figure shows C-terminal domain interacting with the "AT hook" DNA binding domains between two HMGA2Q85M monomers. Panel i shows the C-terminal interacting with the third "AT hook" DNA binding domain. Panel ii illustrates the C-terminal interacting with the second "AT hook" DNA binding domain and panel iii shows the C-terminal interacting with the first "AT hook" DNA binding domain. The dotted lines show potential EDC cross-linking between the two monomers. Vertical striped areas represent the three "AT hook" DNA binding domains. Horizontal striped areas represent the C-terminal domain.

third model, Figure 3A(iii), 3B(iii) and 3C(iii) show the homodimer formation between the C terminal domain and the first "AT hook" region. The mutant HMGA2I71M would be cleaved in this scenario as the methionine residues are located such that after cleavage the homodimer would be separated in half and would appear on an SDS-polyacrylamide gel as a monomer (Fig. 3B(iii)). Both mutants HMGA2Q37M and HMGA2Q85M would appear to not be cleaved as the methionine residues are surrounded by areas of cross-linking (Fig. 3A(iii), 3C(iii)). If the C terminal domain interacted with itself, all the proteins, HMGA2Q37M, HMGA2I71M and HMGA2Q85M would appear to be cleaved.

These three models were tested in this study. First, the proteins were cleaved using cyanogens bromide to confirm the cleavage was successful. All three mutants were cleaved as expected, and only the larger of the fragments are shown within this gel (Fig 4. HMGA2Q37M compare Lanes 1 and 2; HMGA2I71M compare Lanes 3 and 4; HMGA2Q85M compare Lanes 5 and 6). All three mutants were then cross-linked using the chemical EDC; all mutants had the same efficiency in cross-linking (Fig 4. Lanes 7, 8 and 9). The homodimer was purified from the gel and chemically cleaved using cyanogens bromide as discussed previously in the "Materials and Methods" section. Only the mutant HMGA2Q37M homodimer can be cleaved by CNBr, only the larger cleavage product is shown representing the second and third "AT hook" domains of both monomers as well as both C-terminal domains, the larger band represents uncleaved homodimer (Fig 4. Lane 10). There was no visible cleavage of mutants HMGA2I71M or HMGA2Q85M (Fig 4. Lanes 11 and 12). The most probable scenario for this protein, inferred from the cleavage products is the second model where the C-terminal interacts with the second "AT hook" region.

Figure 4: CNBr cleavage analyses of the EDC cross-linked HMGA mutant proteins.



HMGA2Q37M, HMGAI71M and HMGAQ85M EDC cross-linked protein samples were prepared and cleaved by CNBr as described under "Materials and Methods". Lanes 1, 3 and 5 contained untreated HMGA2Q37M, HMGAI71M and HMGAQ85M respectively; Lanes 2, 4 and 6 respectively contained, uncross-linked HMGA2Q37M, HMGAI71M and HMGAQ85M CNBr-cleaved; Lanes 7-9 respectively contained, the EDC cross-linked HMGA2Q37M, HMGAI71M and HMGAQ85M proteins; Lanes 10-12 contain EDC cross-linked HMGA2Q37M, HMGA2I71M and HMGA2Q85M dimers cleaved by CNBr, respectively.

Chemical Cross-Linking of HMGA2C41G Mutant Protein

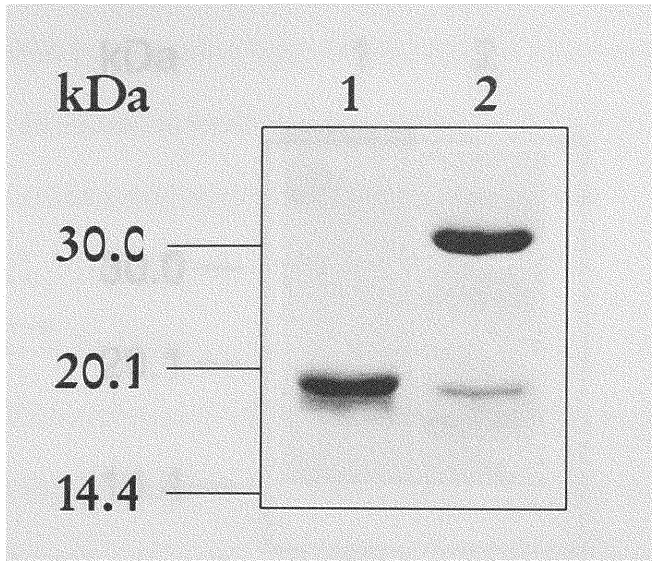
HMGA2 has a cysteine residue at position 41 (Manfioletti *et al.*, 1995). The cysteine residue can form disulphide bonds by joining sulfur atoms together to make this covalent bond. This can be observed in Figure 5 showing the HMGA2 protein with (Lane 1) and without (Lane 2) β-mercaptoethanol. The chemical β-mercaptoethanol reduces disulphide bonds and so without the addition of this chemical the homodimer is observed

(Fig. 5, Lane 2) (Berg *et al.*, 2002). Even though the disulphide bond creates a homodimer, the reducing environment within the inside of a cell makes the development of these types of bonds very unlikely (Berg *et al.*, 2002). Therefore, to ascertain whether the protein forms a homodimer without the construction of this covalent bond, a mutant was created in our lab that contained no cysteine residue. A glycine residue was substituted for the cysteine. The human HMGA2 protein contains a glycine whereas in the mouse a cysteine is at this residue location (Manfioletti *et al.*, 1991, 1995). For that reason this substitution was assumed to have the least affect on the structure of the protein. The protein HMGA2C41G was created and purified as described under the "Materials and Methods" section. The protein was cross-linked using both EDC and DSS. Homodimers as well as other higher order oligomers were formed when HMGA2C41G was cross-linked with EDC. This can be seen when comparing Lanes 1 and 2 of Figure 6. DSS is a homobifunctional cross-linker with a spacer arm of 11.4 Å that reacts using amine-reactive *N*-hydroxysuccinimide (NHS) esters located on each arm (Partis *et al.*, 1983). The reaction involving HMGA2C41G with DSS produced dimers as well as other higher order oligomers. This is shown when comparing Lanes 1 and 2 of Figure 7. The ability of both EDC and DSS to cross-link the HMGA2 protein lacking a cysteine residue, demonstrates that a homodimer between two HMGA2 proteins can occur independent of the formation of a disulphide bond.

Isothermal Titration Calorimetry

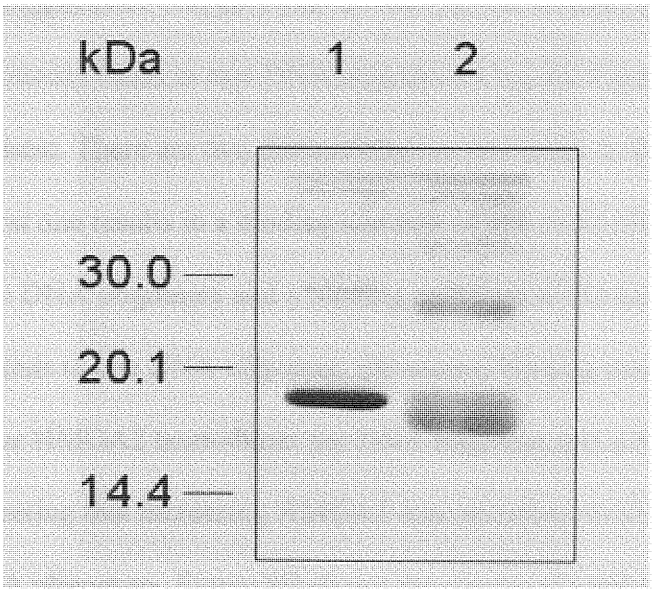
In this study, we used a DNA oligomer, 5'$G_5C(AT)_7ACG_5$-3'(top strand) that contains a single 15 bp AT site in the middle, to determine how the HMGA2 homodimer binds to the isolated AT site. Interestingly, our results from isothermal titration

Figure 5: Disulphide bond formation.



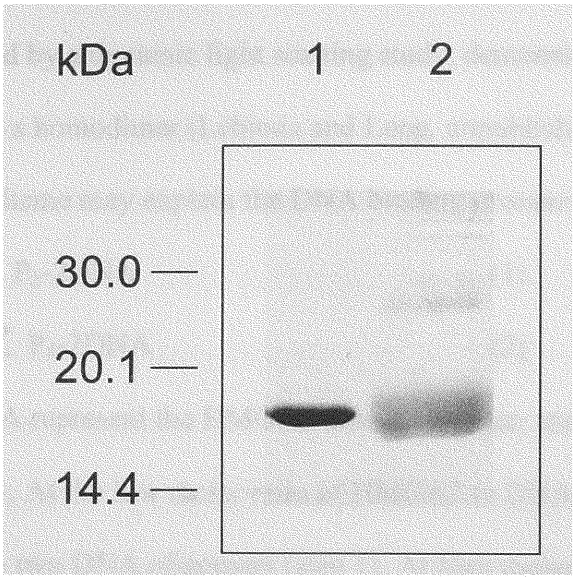
15% SDS PAGE showing disulphide bond formation. Lane 1 HMGA2 with β-mercaptoethanol; Lane 2 HMGA2 without β-mercaptoethanol showing disulphide bond

Figure 6: Mutant HMGA2C41G cross-linked using EDC.



15% SDS-PAGE of HMGA2C41G cross-linked with EDC as described under "Materials and Methods" section. Lane 1, HMGA2C41G mutant; Lane 2, HMGA2C41G cross-linked with EDC

Figure 7: Mutant HMGA2C41G cross-linked using DSS.



15% SDS-PAGE of HMGA2C41G cross-linked with DSS as described under "Materials and Methods" section. Lane 1, HMGA2C41G mutant; lane 2, HMGA2C41G cross-linked with DSS

calorimetry (ITC) experiments and gel mobility shift assays showed that each HMGA2 homodimer cooperatively binds to two DNA oligomers (Fig. 8). Figure 8A shows a typical ITC experiment. There are two distinct binding processes in this titration experiment. The first one results in a binding site size of one molecule of HMGA2 homodimer per two DNA oligomers and a binding enthalpy of -52 kcal/mol, suggesting that each subunit binds to one DNA oligomer; the second binding reaction has a binding stoichiometry of one HMGA2 homodimer per DNA oligomer and a binding enthalpy of -38 kcal/mol, suggesting that increasing HMGA2 concentration results in a tetramer binding to two DNA oligomers. Figure 8B shows the results of a gel mobility-shift assay. At the low molar ratio of HMGA2 to DNA, the HMGA2 homodimer binds to two DNA oligomers (Lanes 2 and 3; the first shift, D). At the high molar ratio of HMGA2 to DNA,

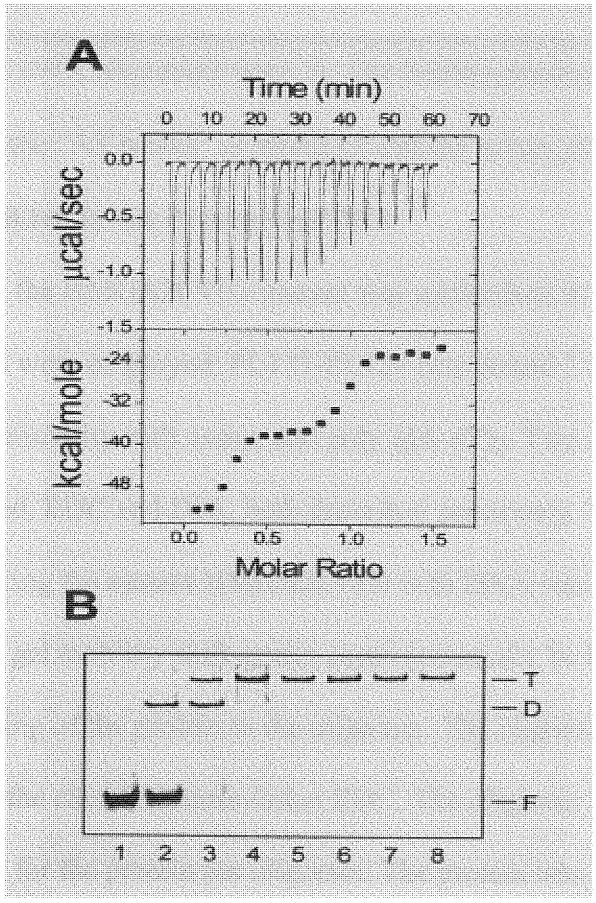
the HMGA2 tetramer binds to the DNA oligomer (Lanes 4-8; the second shift, T). Our results were confirmed by a dynamic light scatting study, demonstrating that HMGA2 binds to AT DNAs as a homodimer (Lebioda and Leng, unpublished results). The following two-step scheme may explain the DNA binding process:

$$P_2 + 2DNA \rightleftharpoons P_2\text{-}2DNA \qquad (1)$$

$$P_2 + P_2\text{-}2DNA \rightleftharpoons P_4\text{-}2DNA \qquad (2)$$

where P_2, P_4, and DNA represent the HMGA2 dimer, tetramer, and the AT DNA oligomer, respectively. At the low molar ratio of HMGA2 to DNA, the homodimer cooperatively binds to two DNA oligomers (step 1). At high molar ratio, the HMGA2

Figure 8: HMGA2 binding to a DNA oligonucleotide containing a single HMGA2 DNA binding site.



(A) Sample raw data for the titration of HMGA2 into the deoxyoligonucleotide, 5'-GGGGGCATATATATATATATACGGGGG-3' (top strand) at 25 °C in BPE buffer plus 4 mM NaCl (total 20 mM Na^+). Top, each peak shows the heat produced by injection of an aliquot of 15 μl of HMGA2 (75 μM, dimer concentration) into DNA solution (1.7 ml of 10 μM the oligonucleotide). Bottom, the binding isotherm resulting from integration with respect to time. The X-axis represents the protein to DNA molar ratio. (B) Gel mobility shift assay of HMGA2 binding to the DNA oligonucleotide. Binding reactions of HMGA2 to DNA were carried out as described under "Materials and Methods." 0.4 μM DNA was used in the experiment. Lanes 1 to 8 contain, respectively, 0, 0.2, 0.4, 0.8, 1.0, 2.0, 5.0, and 10 μM of HMGA2. F, D, and T represent free DNA, HMGA2 dimer binding to DNA, and HMGA2 tetramer binding to DNA respectively.

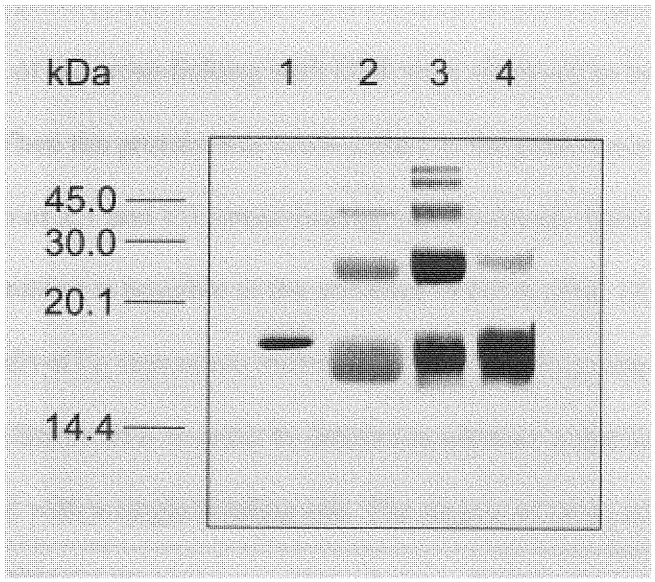
---

tetramer binds to two DNA oligomers (step 2). The binding processes are achieved by the interactions between the "AT hooks" of HMGA2 and the minor groove of the AT DNAs.

Chemical Cross-Linking in the Presence of DNA

Cross-linking the HMGA2C41G mutant was undertaken in the presence of DNA. Two DNA oligomers were employed; FL123-FL124 (top strand 5'$G_5CA_{15}CG_5$3'), used previously in the gel mobility shift assay, and poly (dA-dT)$_2$. Two chemical cross-linkers were individually used, EDC and DSS, both described previously. The purpose of chemically cross-linking the mutant HMGA2C41G with DNA was to observe whether the band shifts in the gel mobility shift assays could potentially be homodimeric or homotetrameric binding of the protein to the oligomer. The oligomer FL123-FL124 was first used along with the chemical cross-linker EDC. The oligomer represents one potential binding site for the protein per oligomer. The protein was observed to cross-link

Figure 9: Mutant HMGA2C41G cross-linked using EDC in the presence of DNA oligomer FL123-124.



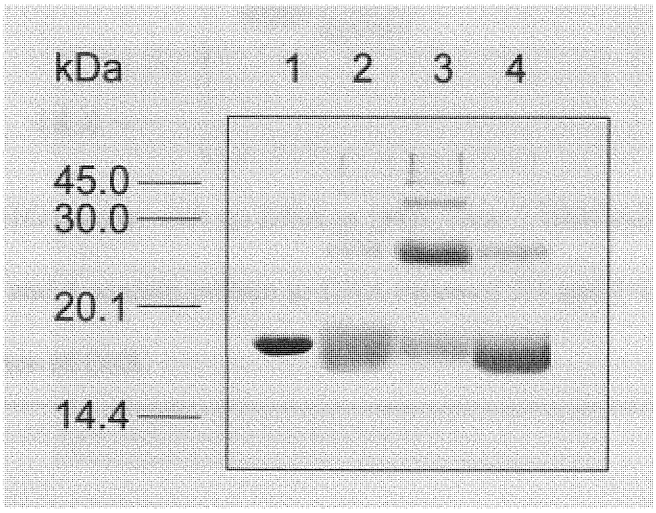
Lane 1 HMGA2C41G protein with no DNA or EDC, lane 2 HMGA2C41G cross-linked with EDC, Lanes 3 and 4 contain protein to DNA ratios of 2:1 and 1:1 respectively cross-linked in the presence of EDC.

in the presence of no DNA as expected (Fig. 9, Lane 2). When the protein to DNA ratio was altered to a 2:1 or 1:1 ratio the cross-linking ability of EDC remained (Fig. 9, Lanes 3 and 4). This demonstrates that in the presence of DNA the HMGA2 protein is still able to form homodimers and higher homo-oligomers.

The oligomers FL123-FL124 were then cross-linked along with HMGA2C41G and the chemical cross-linker DSS. The cross-linker EDC reacts with closely associated carboxyl (Glu and Asp or unmodified C-terminus) and amino groups (Lys or unmodified N-terminus), while DSS is able to cross-link between two primary amines. Both cross-linkers were tested in order to observe the difference in efficiency between the two and

therefore establish which amino acids are most likely to contribute to the cross-linking ability. A similar result was observed using the cross-linker DSS as was seen using EDC. Cross-linking of the protein was witnessed with no DNA present as predicted (Fig. 10, Lane 2). The results of the cross-linking with DSS were observed to be the same as that shown with EDC. When the protein was cross-linked in the presence of DNA at both a protein to DNA ratio of 2:1 and 1:1, cross-linking was observed. This again demonstrates that homodimer formation is possible in the presence of DNA (Fig. 10, Lanes 3 and 4). The oligomer poly $(dA\text{-}dT)_2$ represents multiple HMGA2 binding sites. It was utilized along with the cross-linker DSS in order to determine whether a similar pattern can be seen as was observed with a single binding site. The degree of cross-linking ability

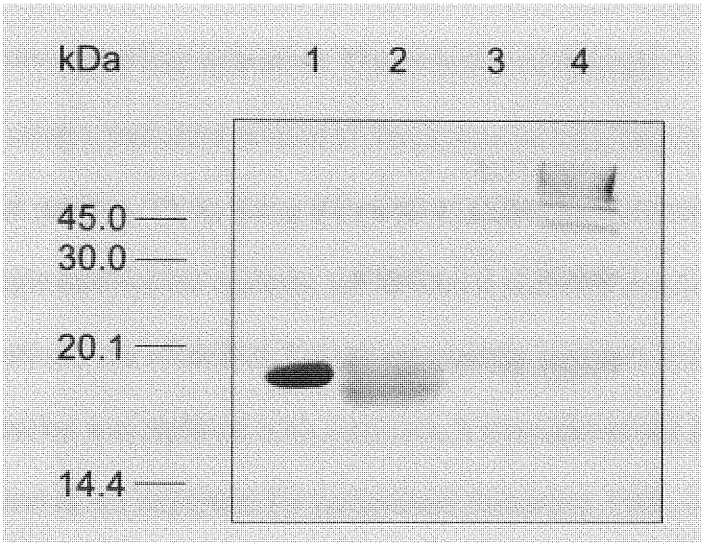
Figure 10: Mutant HMGA2C41G cross-linked using DSS in the presence of DNA oligomer FL123-124.



Lane 1 HMGA2C41G protein with no DNA or DSS, Lane 2 HMGA2C41G cross-linked with DSS, Lanes 3 and 4 contain protein to DNA ratios of 2:1 and 1:1 respectively cross-linked in the presence of DSS

appeared to be reduced for poly $(dA\text{-}dT)_2$ compared to that seen for the single binding site (FL123-FL124). Cross-linking in the absence of DNA occurred as expected (Fig. 11, lane 2). When the protein was incubated with DNA containing multiple protein binding sites cross-linking with DSS was observed. This may imply that the protein prefers to bind to multiple DNA binding sites as a homodimer or higher homo-oligomer.

Figure 11: Mutant HMGA2C41G cross-linked using DSS in the presence of DNA oligomer poly$(dA\text{-}dT)_2$.



Lane 1 HMGA2C41G protein with no DNA or DSS, Lane 2 HMGA2C41G cross-linked with DSS, Lanes 3 and 4 contain protein to DNA ratios of 2:1 and 1:1, respectively cross-linked in the presence of DSS.

CHAPTER V

DISCUSSION

Earlier studies demonstrated that HMGA2, an intrinsically unstructured protein, exists as a homodimer when free in solution (Baez and Leng unpublished results). The current research further supports this hypothesis and provides additional insight into how the protein behaves when bound to DNA. Five conclusions can be gained from this research. First, based upon chemical cross-linking experiments using EDC and DSS, the HMGA2 protein was shown to exist as a homodimer when free in solution. Second, the C-terminal is involved in dimer formation. This was shown through size exclusion chromatography using a truncated version of the protein lacking the C-terminal, and a labeled C-terminal peptide. Third, using information gained by chemical cross-linking and cutting of single methionine mutants, the structural elements involved in the homodimer formation were identified. Fourth, by creating a mutant protein where the single cysteine residue was replaced with a glycine, it was determined that the formation of a disulphide bond is not important for homodimer formation. Fifth, using a combination of isothermal titration calorimetry and gel mobility shift assays the binding stoichiometry of HMGA2 to a single or multiple AT-rich DNA sites was observed. The forensic relevance of HMGA2 and the techniques used in this study is also discussed.

HMGA2 is a Homodimer When Free in Solution

HMGA2 has an asymmetrical charge distribution. Positively charged amino acids are located mainly within the center, concentrated mostly within the "AT hook" domains. The negatively charged amino acids are positioned towards the C-terminus of the protein

Manfioletti *et al.*, 1991). This arrangement led us to propose that HMGA2 could potentially exist as a homodimer or homo-oligomer.

Initial studies involved sedimentation analysis and gel filtration experiments. These determined the physical characteristics of HMGA2, such as molecular weight and the Stokes' radius (Horiike *et al.*, 1983). It was suggested that HMGA2 adopts a homodimer arrangement when compared to ribonuclease A (Baez and Leng, unpublished results). Ribonuclease A is a monomeric protein with a similar molecular weight to HMGA2 of 13,700 Da but a Stokes' radius of only 16.4Å compared to 30.2 Å for HMGA2 (Smyth *et al.*, 1963). Current research tested the cross-linking ability of HMGA2 with two chemical cross-linking agents, EDC and DSS. EDC is a zero-length cross-linker that cross-links between the carboxyl and amino groups of closely associated amino acids. HMGA2 has 13 carboxyl groups (11 glutamic acids, 1 aspartic acid, and one C-terminus carboxy group), 8 of which are located within the negative C-terminus. There are also 14 amino groups (13 lysine residues and one N-terminal NH_2 group), 6 of which are located within the "AT hook" regions and none are present in the C terminal (Manfioletti *et al.*, 1991). These amino acids can therefore provide linkage between the C-terminal domain and the "AT hook" domains. The experiments using EDC show a high percentage of cross-linking which provided evidence that amino acids located within two HMGA2 monomers exist in close contact (Fig. 6). It is also interesting to note that monomers of the protein remain after the cross-linking reaction (Fig. 6, Lane 2). These monomers migrated faster on an SDS polyacrylamide gel than the untreated monomers (no cross-linking). This may indicate that the protein bends, and links occur between

regions within the monomer. These associations are most likely to take place between the C-terminal and another region within the protein due to the charge distribution.

DSS is a homobifunctional cross-linking agent that has a spacer arm length of 11.4 Å. This cross-linking agent cross-links two primary amines using amine-reactive *N*-hydroxysuccinimide (NHS) esters (Partis *et al.*, 1983). DSS was able to cross-link HMGA2, but with a lower rate efficiency than EDC (Fig. 7). When the protein concentration was high, both cross-linking agents were able to produce additional homo-oligomers, which might represent trimers or tetramers, as judged from the position of the bands on the gels (Figs. 6 and 7).

It is of interest to note that the enhanceosome formation at the β-interferon promoter involves two molecules of HMGA1 (Yie *et al.*, 1997). This indicates that a homodimer of HMGA proteins is likely to be involved in the formation of the enhanceosome. Each of the two proteins aids in binding to a different region of the enhancer. One molecule binds to positive regulatory domain II (PRDII) using "AT hooks" one and two. This reverses an intrinsic bend in the DNA, facilitating the binding of activating transcription factor-2 (ATF-2)/c-Jun. The second molecule binds to PRDIV using "AT hooks" two and three. This assists with the binding of nuclear factor-κB (NF-κB) (Yie *et al.*, 1997, 1999). The HMGA family of proteins has been shown to function differently, as in the case of adipocyte cell growth where HMGA1 and HMGA2 have opposing effects (Melillo *et al.* 2001). However, all HMGA molecules possess three "AT hook" binding domains and an acidic C-terminal domain; and all bind to DNA via the minor groove of AT rich DNA (Zhou *et al.*1996, Solomon *et al.* 1986). This indicates that the mechanisms of binding are likely conserved between all HMGA proteins.

HMGA2 has been associated with many protein partners indicating that protein-protein interactions are able to occur with the HMGA2 protein (Sgarra *et al.*, 2005 and references within). A broad spectrum of protein partners exist, they include: transcription factors, mRNA processing proteins, chromatin-remodelling related factors and structural proteins (Sgarra *et al.*, 2005). Some of these proteins bind to HMGA in the absence of DNA demonstrating that the protein-protein interactions are not a result of close contacts when proteins are bound to DNA. The binding of nuclear factor-Y (CCAAT binding factor) (NF-Y (CBF)) to the CCAAT box of the α2(I) collagen promoter involves HMGA1. No high affinity-binding site for HMGA1 exists at this promoter region, but this protein has been shown to be essential for transcription. It was found that HMGA1 interacts with NF-Y (CBF) and stabilizes the binding of this factor to the CCAAT box (Currie, 1997). This further illustrates that protein-protein interactions can occur free in solution between two or more HMGA molecules and between HMGA and other proteins.

The C-Terminal is Required for Homodimer Formation

A truncated version of the HMGA2 protein, missing the C-terminal tail, has been implicated in many benign tumors (reviewed in Fedele *et al.*, 2001). The mechanisms by which this truncated HMGA2 protein operates remain unknown. Rearranged copies of the protein also exist, where the C-terminal has been removed and replaced with various fusion partners (reviewed in Fedele *et al.*, 2001). These fusion partners can contribute as little as a few amino acids to the molecule and therefore not impart function on the fusion protein (Kools and Van de Ven, 1996). This indicates that the loss of the C-terminal tail, rather than the fusion to another protein contributes to the benign tumor phenotype (Fedele *et al.*, 2001). The tumor formation is a result of a loss in cell cycle control. This

implicates the C-terminal in cell cycle control regulation and therefore essential to the correct functioning of the protein.

We proposed that the C-terminal tail is involved in the homodimer formation due to the charge distributions within the protein. Preliminary data using truncated HMGA2 demonstrated that with the cross-linker EDC, homodimers were unable to form (Baez and Leng unpublished results). This current research provides further evidence of the C-terminal tail involvement in homodimer formation. Using truncated HMGA2 and a C-terminal peptide labeled with tetramethylrhodamine-5-maleimide, size exclusion chromatography was performed. The results showed that the labeled C-terminal co-eluted with the truncated HMGA2 protein (Fig. 2). This was shown as a correlation was observed between the peaks in the graph corresponding to the concentration of the TMR and the concentration of the protein observed on an SDS-polyacrylamide gel. This suggests that the C-terminal is bound to the truncated protein, and points to an involvement of the C-terminal in protein-protein interactions. This result also explains how the protein monomer interacts after cross-linking. The C-terminal tail must interact with another region of the protein to produce a more compact molecule. This compact protein is therefore able to run faster in an SDS polyacrylamide gel, as observed previously (Baez and Leng, unpublished results).

More evidence for the C-terminal involvement in protein-protein interactions includes *in vivo* experiments concerning the loss of the C-terminal, which have varying results. In the case of transcription involving serum response factor (SRF) and HMGA1 at the SRF-responsive promoter, the loss of the C-terminal region had no effect on transcription (Chin *et al.*, 1998). In contrast, transcription is halted at the β-interferon

gene when the C-terminal tail of HMGA1 is lost. Therefore the C-terminal tail was deemed necessary for NF-κB coactivation of the *β-interferon* gene (Yie *et al.*, 1997). The binding affinity of other proteins for the HMGA family is altered after the loss of the C-terminal domain (Sgarra *et al.*, 2005). Proteins found to have an increased binding affinity for HMGA when the C-terminal domain is absent include hnRNP H and K (Sgarra *et al.*, 2005).

The loss of the C-terminal of HMGA2 has been found to have opposing effects for the regulation of insulin-like growth factor II mRNA binding protein (IMP2) compared to the wild type protein (Brants *et al.*, 2004). The wtHMGA2 protein was found to up-regulate transcription of *IMP2* in wild type mouse embryos compared to pygmy mutant embryos not expressing HMGA2 (Brants *et al.*, 2004). On the other hand, the truncated HMGA2 missing its C-terminal domain down-regulated transcription compared to that seen in pygmy mutant embryos with no HMGA2 expression (Brants *et al.*, 2004). These opposing effects suggest an important role for the C-terminal in transcriptional control.

Loss of the C-terminal has implications on the transcriptional regulation of genes such as *IMP2*. As the C-terminal has been proven necessary in homodimer formation, the loss of the homodimer may facilitate protein-protein interactions with other factors and affect their binding to promoter or enhancer elements. The opposite may also be true, in such cases; the loss of the C-terminal and therefore the loss of the homodimer may decrease protein-protein interactions and diminish transcription of some genes. The presence of the C-terminal and therefore the existence of HMGA2 as a homodimer would be necessary for the correct functioning of HMGA2 as an architectural transcription

factor. Also, to the correct functioning of the protein to enhance or repress transcription by altering protein-protein and protein-DNA contacts at the enhancer or promoter.

Model of Homodimer Formation

After the C-terminal was demonstrated necessary in the formation of the HMGA2 homodimer, it was logical to uncover which region it binds to in the protein. Due to charge interactions, the region of C-terminal binding to another HMGA2 protein was narrowed down to the three "AT hook" regions (Fig. 3).

Protein aggregation from an unstructured protein or peptide into structured amyloid fibrils can be accomplished through unique properties of the protein including charge interactions. Proteins that possess hydrophobic residues and therefore have a tendency to form β-sheets, as well as amino acids that possess paired charges encourage fibrillar formation (Tjernberg *et al.*, 2002). It has also been found that positively and negatively charged peptides bind together or copolymerize (Tjernberg *et al.*, 2002). Proteins which form a structured molecule via charge interactions include the interactions between polycations such as spermine and polyarginine with the unstructured protein α-synuclein (Goers *et al.*, 2003). This information supports the formation of a homodimer by the HMGA2 protein, via electrostatic charge interactions between the negatively charged C-terminal and a positively charged "AT hook".

It has been shown that the "AT hook" motifs appear essential in protein-protein interactions with other factors implicating them in the formation of the homodimer. Reeves found that many factors interact with the HMGA proteins (Reeves, 2001). The sites of contact are dispersed throughout the entire length of the HMGA protein. These contacts have a commonality; the points of association always involve at least one "AT

hook" region along with some flanking amino acids. Reeves also noted that these "AT hook" regions are known to be modified through phosphorylation, acetylation or methylation, providing molecular switches for association and dissociation of protein-protein interactions (Reeves, 2001). Proteins found to interact with HMGA via the "AT hook" regions include NF-Y (CBF), which requires at least one "AT hook" (Currie, 1997) and SRF which specifically requires the third "AT hook" region (Chin *et al.*, 1998).

Three single methionine mutants were created by our lab, where a methionine residue was placed between the "AT hook" domains, and between the third "AT hook" and the C terminal (Fig. 3). These mutants were cross-linked using EDC and the homodimers cleaved using cyanogens bromide. Cyanogens bromide exclusively cleaves methionine residues on the carboxyl side (Berg *et al.*, 2002). Due to the specificity of the cyanogens bromide digestion, pattern analysis from the resulting gel was achieved. Three basic models were devised, and results suggested that the C-terminal domain interacts with the second "AT hook" DNA binding domain (Fig. 4).

This finding confirms the importance of the second "AT hook" in protein-protein interactions. The second "AT hook" has been shown to interact not only with DNA but also with many other proteins (Reeves, 2001). Many protein partners of HMGA lose their ability to interact with DNA or drive transcription if the second "AT hook" is missing. These proteins include CCAAT / enhancer binding protein β (C / EBPβ) at the leptin promoter (Melillo *et al.*, 2001), and specificity protein 1(Sp1) at the human insulin receptor gene promoter (Foti *et al.*, 2003). Both of these factors interact with HMGA1. This indicates that the loss of the homodimer may play a role in the loss of transcriptional

control of these genes. The importance of the second "AT hook" region in HMGA2 has also been illustrated with the protein partner $p120^{E4F}$ at the cyclin A promoter (Patel *et al.*, 1994). Protein partners involving HMGA2 need to be further studied to demonstrate the importance of the second "AT hook" region.

Phosphorylation sites for Cdc2 flank the second "AT hook" DNA binding domain at Ser-43 and Ser-58, four other sites for casein kinase 2 (CK2) occur in the C-terminal at Thr-98, Ser-99, Ser-100 and Ser-103 (Schwanbeck *et al.*, 2000). Both of these regions are involved in homodimer formation. Thus, it is possible that phosphorylation at one or all of these sites mediates monomer-dimer association / dissociation. It has been shown that phosphorylation of cAMP response element-binding protein (CREB) dimers by calmodulin kinase II (CamKII) leads a decrease in recruitment of CREB-binding proteins inhibiting transcription of CRE-dependent reporter genes (Wu and McMurray, 2001). This could be biologically significant if a homodimer of HMGA2 is essential for transcription factor recruitment, and could indicate a possible cell-cycle control mechanism in HMGA2 mediated gene transcription.

Cysteine (C41) is Not Required for the Homodimer Formation

The presence of a cysteine residue at position 41 within HMGA2 led us to consider the possibility of a disulphide bond contribution to the establishment of the homodimer. In the absence of a reducing agent such as β-mercaptoethanol, a homodimer was visualized on an SDS polyacrylamide gel indicating that establishment of a disulphide bond is possible (Fig. 5). These bonds are formed when oxidation occurs between sulfhydryl groups particularly involving cysteine residues. Disulphide bonds covalently attach the two cysteine residues (Berg *et al.*, 2002). In eukaryotes, these bonds

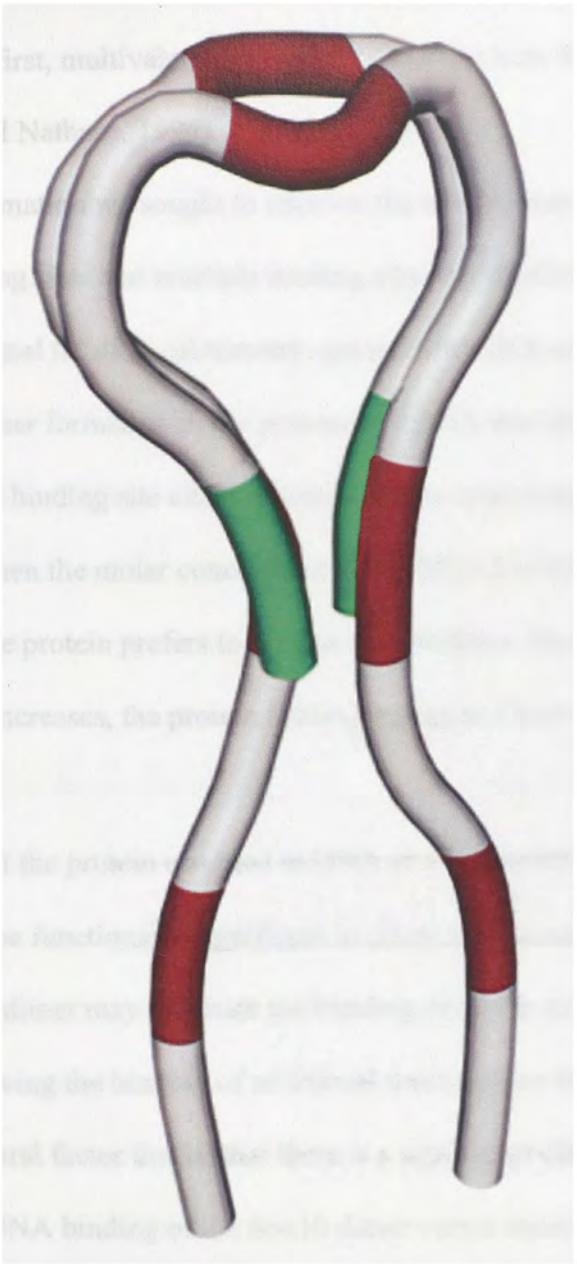
are only found within or between proteins in the lumen of the rough endoplasmic reticulum. The reducing environment in the rest of the cell makes bond formation of this type unfavorable (Berg *et al.*, 2002). As HMGA2 is a nuclear protein we investigated the establishment of the homodimer without the presence of cysteine. This determined whether the protein could form a homodimer within the cell. A mutant protein was created where the cysteine residue was removed and replaced with a glycine residue. This mutant was cross-linked using EDC and DSS. Our results demonstrated that a homodimer could occur without the need for the disulphide bond (Figs. 6 and 7). The protein is able to establish a disulphide bond involving the cysteine residues located between the first and second "AT hook" domains of the monomers. The homodimer model was adjusted to reflect that the interaction between the C-terminal domain and the second "AT hook" domain could occur as well as the disulphide bond (Fig. 12). This was achieved by bending the homodimer allowing the cysteine residues to come in close contact. The EDC cross-linker was also able to link amino acids surrounding the third "AT hook" domains but these domains would then repel one another due to charge-charge interactions.

This finding verified that the HMGA2 homodimer may exist within a cell and led us to further investigate the homodimer interactions with DNA.

HMGA2 Binding To DNA

HMGA2 has been shown to function as an architectural factor that binds to DNA to facilitate the recruitment of transcription factors to the promoter (reviewed in Goodwin, 1998). Hence, it is crucial to understand what conformation the protein achieves upon binding to DNA. Maher and Nathans have shown that HMGA binding to a

Figure 12: Model of HMGA2 homodimer formation.



Model of HMGA2 homodimer formation with the C-terminal interacting with the second "AT hook" DNA binding domain. The red areas represent the "AT hook" domains and the green areas represent the C-terminal domain.

single AT rich tract, such as TATA boxes, occurs via univalent low affinity binding (Maher and Nathans, 1996). By adding one or two appropriately spaced AT tracks in close proximity to the first, multivalent, high affinity binding sites for the protein are constructed (Maher and Nathans, 1996).

With this information we sought to uncover the stoichiometries involved in DNA binding to single binding sites and multiple binding sites for HMGA2. Using a combination of isothermal titration calorimetry, gel mobility shift assays and chemical cross-linking, homodimer formation in the presence of DNA was observed (Figs. 8, 9, 10 and 11). Using a single binding site either a homodimer or a homotetramer binds to two DNA binding sites. When the molar concentration of HMGA2 is low compared to the DNA concentration, the protein prefers to bind as a homodimer. However, when the protein concentration increases, the protein favors binding as a homotetramer or higher homo-oligomer.

The finding that the protein can bind to DNA as a homodimer has important implications and may be functionally significant to allow for transcription factor binding. The binding of a homodimer may facilitate the bending of DNA, altering its conformation and allowing the binding of additional transcription factors. It has been shown in the architectural factor Sox10 that there is a significant difference in DNA bending between the DNA binding of the Sox10 dimer versus monomer (Peirano and Wegner, 2000). Sox10 has a high mobility group binding domain and interacts with the consensus (A/T)(A/T)CAA(A/T)G, binding to this sequence introduced a bend in the DNA (Peirano and Wegner, 2000). Binding of a monomer introduced a bend angle of ~75-80°, while binding of a dimer produced a greater bend angle of ~101° (Peirano and

Wegner, 2000). It was also observed that the dissociation of Sox10 from the DNA was much slower for the dimer than for the monomer indicating a stronger affinity of the DNA for the dimer (Peirano and Wegner, 2000). As Sox10 is an architectural transcription factor and contains a high mobility group binding domain, the binding structures of this protein can be assumed similar to HMGA2. As Sox10 can form a dimer and this protein conformation has an affect on DNA bending, HMGA2 may have a similar effect. This difference in DNA conformation has also been associated with the binding of other HMG proteins due to alterations in protein stoichiometry. The binding of HMGA1 to α-satellite DNA at high protein to DNA ratios introduced bends in the double helix. Binding of the protein at low molar ratios introduces a smaller bend in the DNA (Slama-Schwok *et al.*, 2000). Also in closed circular plasmids the binding of HMGA proteins at high molar ratios introduced negative supercoils into the DNA, these negative supercoils were thought to be a result of both bending and unwinding of the plasmid (Nissen and Reeves, 1995). At lower molar ratios, positive supercoils were introduced; these supercoils were a result of bending of the DNA, where no unwinding was observed (Nissen and Reeves, 1995). Binding of HMGA2 at high molar ratios may have a similar effect as that seen with both α-satellite DNA and closed circular DNA. This is due to the DNA binding "AT hook" domains of the proteins in these studies being the same as those within HMGA2.

The related protein HMGA1 is involved in the formation of the enhanceosome at the human interferon-β gene promoter (Yie *et al.*, 1999). The formation of this structure involves two molecules of HMGA1. Both HMGA1 molecules are required in order to produce conformational changes in the DNA. The HMGA1 proteins are involved in

additional protein-protein interactions with NF-κB and ATF-2/c-Jun, which stabilize the enhanceosome (Yie *et al.*, 1999). Whether the two HMGA1 molecules form a homodimer remains unknown. Studies related to enhanceosomes involving HMGA2 are required to discern if two molecules of this protein and hence a possible homodimer are involved in the enhanceosome formation and what protein and DNA interactions occur.

Forensic Relevance

Proteins are essential in the field of Forensic Science. The tools used to determine the characteristics of these proteins and the genes they are transcribed from also have wide impacts in forensics. Some of the techniques used in this research have already transferred over into the forensic setting. The polymerase chain reaction is now widely used and accepted in many forensic DNA labs. Other techniques used here have potential forensic uses that could be utilized in the future.

DNA profiling has become an integral part of any forensic laboratory. The polymerase chain reaction (PCR); is a widely used technique that can aid in the identification of individuals or determine paternity. The procedure involves amplifying a region between two known primers to create multiple copies of the DNA region of interest. The section that is amplified for these purposes contains "short tandem repeats" (STRs). STRs are repetitions of a short sequence where the number of repeats varies between individuals. The repeat numbers are observed at several loci and provide an analysis of an individual with a high level of discrimination (Rudin and Inman, 2002).

Another identifying region in DNA is the single nucleotide polymorphism (SNP). This is characterized by a single base change in a region of DNA that is polymorphic (Rudin and Inman, 2002). It has become increasingly important for forensic techniques to

become portable and robust enough to be used at the crime scene itself. As a result, the SNP regions of DNA can now be tested in microassays or biosensors. Biosensors involve immobilizing oligonucleotides on solid supports. When an evidence sample is introduced, binding to form double stranded DNA will occur if the sequence is found to be complimentary. A chemical cross-linker such as EDC can be utilized. The cross-linker is able to link ssDNA to a self-assembled, thioglycolic acid (TGA) monolayer modified gold electrode. A color indicator is then used which reacts differently with single stranded compared to double strand DNA. The intensity of the signal is often an indication of concentration, however, the mere presence of a signal is often the only indication required to show the presence of the sequence tested (Ye and Ju, 2003).

Identification of biotoxins is an important aspect of Forensic Biology. Due to the threat of bioterrorism a rapid method is required to identify an infectious disease. Size-exclusion chromatography has been successfully used along with trypsin digestion, desalination and reversed-phase high performance liquid chromatography, for the identification of proteins and infectious disease agents. Size exclusion chromatography is used to filter the protein mixture to a smaller number of proteins within a size range. Calibration of the size exclusion column can be achieved by first using control proteins of various known molecular weights and plotting the molecular weight against elusion time. This produces a standard curve which unknowns can be measured against. When the system is fully automated, it can collect the appropriate fraction of protein based on pre-calculated elution times. Hence, a known range of molecular weights can be isolated. In this way size exclusion chromatography can be used to make the trypsin digest analysis

simpler and less time consuming. This technique has been successfully used for cholera toxin and staphylococcal enterotoxin (Carol *et al.*, 2005).

In the discipline of Forensic Chemistry, the detection and identification of drugs is essential. The technique of ion-exchange chromatography can be used to extract illicit drugs and their metabolites from a matrix ready for gas chromatography/mass spectrometry analysis (GC/MS). Cation-exchange chromatography has been used for the extraction of many illicit drugs including cocaine, amphetamine, benzoylecgonine (cocaine metabolite) and methamphetamine. It has been utilized in samples taken both from urine and serum (Stout *et al.*, 2002, Weinmann *et al.*, 2000). Ion-exchange chromatography functions to separate out the drugs from the matrix. This is utilized in order to make the analysis by GC/MS easier to interpret by sifting out the contaminants.

Finally the model protein HMGA2 used in this study may also have a forensic application. Gestational age has come under much debate with many different measurements being used to determine the age of the fetus. Some measurements use the beginning of the last menstruation, others use time of fertilization and still others use time of implantation. Time discrepancies between these measurements are in the weeks and do not provide an accurate account of fetal age (Santee and Henshaw, 1992). This is extremely important when considering legality of abortion. In the USA, five states, FL, GA, IA, SC and VA prohibit abortions in the 3rd trimester (week 27 onward), nine other states, CA, DE, MA, NV, NY, NC, PA, SD, and VT prohibit abortions after a set number of weeks (usually 20 or 24). A more accurate account of gestational age could be developed by using a protein marker. HMGA2 expression is detected only within rapidly dividing cells of the fetus (Patel *et al.*, 1994, Rogalla *et al.*, 1996). The presence or

absence, level of expression or location of expression of a protein can indicate the degree of development of a fetus. The expression pattern of HMGA2 has been studied in both humans and mice, as discussed previously (Patel *et al.* 1994, Zhou *et al.*, 1996). A more detailed pattern and quantitative analysis would have to be undertaken in order to confirm that the protein could be utilized in this way. HMGA2 expression may therefore have a role in determining gestational age for the legal implications of abortion.

It can take many years for scientific tools to be utilized in the field of Forensic Science. The techniques used must undergo stringent controls and peer review before being considered as techniques that will stand up to the rigors of a court room (Saferstein, 2003). These stringent controls include a great knowledge of the technique and many peer reviewed articles concerning the technique and its advantages and downfalls (Saferstein, 2003). The same is true for proteins that may be used as markers for Forensic purposes. For example the protein marker cardiac Troponin I was discovered in 1963 (reviewed in Filatov *et al.*, 1999). This marker was not realized as a potential forensic tool until 2003 when it was tested for the estimation of postmortem interval (Sabucedo and Furton, 2003). This protein marker is still not used in court for the time since death determination. Many more peer reviewed articles will be required before cardiac Troponin I would be considered within a court of law. For this reason HMGA2 may also have a forensic relevance. However, more research concerning the protein, its structure and expression are required before its impact may be understood.

CHAPTER VI

CONCLUSIONS

This research was performed to further understand the mammalian HMGA2 biochemical properties. Previous studies in this lab indicated that HMGA2 exists in solution as a homodimer. Results in this thesis strongly support this hypothesis. This research included determining if the protein exists as a homodimer when free in solution, as well as when bound to DNA. The structural arrangement of this homodimer was also addressed. The potential forensic impact of the protein and the techniques utilized was also discussed.

Five conclusions concerning homodimer formation of HMGA2 were reached

1. HMGA2 exists as a homodimer when free in solution
2. The C-terminal is required for the establishment of the homodimer
3. The dimer forms between the C-terminal of one HMGA2 protein and the second "AT hook" of another HMGA2 protein due to charge-charge interactions
4. Homodimer formation is not dependent on the existence of a disulphide bond
5. HMGA2 binds to AT-rich DNA as a homodimer

This research indicates that the unstructured protein HMGA2 can form a structured homodimer. The biological implications of HMGA2 forming a homodimer are widespread. Aberrant expressions of a truncated HMGA2 protein, as well as fusion products involving the truncation of HMGA2 have been implicated in cell cycle

deregulation. The C-terminal involvement in dimer formation indicates that the existence of the homodimer may be important in normal cell functions.

The binding of a dimer to DNA has been shown to have different effects on the DNA conformation compared to those seen during monomer DNA binding. The dimer has been shown to effect the bending in α-satellite DNA and changes in superhelicity of closed circular plasmids (Slama-Schwok *et al.*, 2000, Nissen and Reeves, 1995). Binding of a homodimer at a promoter or to DNA wrapped around a nucleosome, may introduce greater conformational changes and allow for binding or dissociation of transcription factors (Strauss and Varshavsky, 1984). Binding as a monomer to these sites may not produce the desired bending, preventing the recruitment of transcription factors. HMGA2 has been shown to modulate the expression of cyclin A (cell cycle control – S phase entry and G_2/M transition), Bfl-1/A1 (apoptosis inhibitor) and ERCC1 (DNA repair) (Tessari *et al.*, 2003, Edelstein *et al.*, 2003, Borrmann *et al.*, 2003). It remains to be seen if HMGA2 binds to these promoters as a dimer and whether this impacts transcription factor recruitment.

The HMGA2 proteins itself, as well as the techniques utilized in this research study, have realized or potential uses in the forensic field. The techniques used in this research cover a wide range of forensic applications that involve Forensic Biology with DNA and protein, and also Forensic Chemistry in the detection of drugs.

REFERENCES

- B. D. Abercrombie, G. G. Kneale, C. Crane-Robinson, E. M. Bradbury, G. H. Goodwin, J. M. Walker and E. W. Johns, Studies on the conformational properties of the high-mobility-group chromosomal protein HMG17 and its interaction with DNA. Eur. J. Biochem. 84 (1978) 173-7

- J. E. Adair, Y. Kwon, G. A. Dement, M. J. Smerdon and R. Reeves, Inhibition of nucleotide excision repair by high mobility group protein HMGA1. J. Biol. Chem. 280 (2005) 32184-92

- L. M. al-Alousi, R. A. Anderson, D. M. Worster and D. V. Land, Factors influencing the precision of estimating the postmortem interval using the triple-exponential formulae (TEF). Part II. A study of the effect of body temperature at the moment of death on the postmortem brain, liver and rectal cooling in 117 forensic cases. Forensic Sci. Int. 125 (2002) 231-6

- F. H. Allain, Y. M. Yen, J. E. Masse, P. Shultze, T. Dieckmann, R. C. Johnson and J. Feigon, Solution structure of the HMG protein NHP6A and its interaction with DNA reveals the structural determinants for non-sequence-specific binding. EMBO J. 18 (1999) 2563-79

- A. Anand and K. Chada, In vivo modulation of Hmgic reduces obesity. Nat. Genet. 24 (2000) 377-80

- S. J. Araujo and R. D. Wood, Protein complexes in nucleotide excision repair. Mutat. Re. 435 (1999) 23-33

- L. Aravind and D. Landsman, AT-hook motifs identified in a wide variety of DNA-binding proteins. Nucleic Acids Res. 26 (1998) 4413-21

- P. Arlotta, A. K. Tai, G. Manfioletti, C. Clifford, G. Jay, and S. J. Ono, Transgenic mice expressing a truncated form of the high mobility group I-C protein develop adiposity and an abnormally high prevalence of lipomas. J. Biol. Chem. 275 (2000) 14394-400

- H. R. Ashar, L. Cherath, K. M. Przybysz and K. Chada, Genomic characterization of human HMGIC, a member of the accessory transcription factor family found at translocation breakpoints in lipomas. Genomics 31(1996) 207-14

- C. T. Ashley, C. G. Pendleton, W. W. Jennings, A. Saxena and C. V. Glover, Isolation and sequencing of cDNA clones encoding Drosophila chromosomal protein D1. A repeating motif in proteins which recognize at DNA. J. Biol. Chem. 264 (1989) 8394-401

- G. Baldassarre, S. Battista, B. Belletti, S. Thakur, F. Pentimalli, F. Trapasso, M. Fedele, G. Pierantoni, C. M. Croce and A. Fusco, Negative regulation of BRCA1 gene expression by HMGA1 proteins accounts for the reduced BRCA1 protein levels in sporadic breast carcinoma. Mol. Cell Biol. 23 (2003) 2225-38

- G. C. Banks, B. Mohr and R. Reeves, The HMG-I(Y) A.T-hook peptide motif confers DNA-binding specificity to a structured chimeric protein. J. Biol. Chem. 274 (1999) 16536-44

- S. Battista, M. Fedele, J. Martinez Hoyes, F. Pentimalli, G. M. Pierantoni, R. Visone, I. De Martino, C. M. Croce and A. Fusco, High-mobility-group A1 (HMGA1) proteins down-regulate the expression of the recombination activating gene 2 (RAG2). Biochem. J. 389 (2005) 91-7

- S. Battista, V. Fidanza, M. Fedele, A. J. Klein-Szanto, E. Outwater, H. Brunner, M. Santoro, C. M. Croce and A. Fusco, The expression of a truncated HMGI-C gene induces gigantism associated with lipomatosis. Cancer Res. 59 (1999) 4793-7

- J. M. Berg, J. L. Tymoczko and L. Stryer, Biochemistry 5th Ed. W. H. Freeman and Company, New York. (2002)

- M. E. Bianchi, M. Beltrame, G. Paonessa, Specific recognition of cruciform DNA by nuclear protein HMG1. Science 243 (1989) 1056-9

- Y. Birger, K. L. West, Y. V. Postnikov, J. H. Lim, T. Furusawa, J. P. Wagner, C. S. Laufer, K. H. Kraemer and M. Bustin Chromosomal protein HMGN1 enhances the rate of DNA repair in chromatin. EMBO J. 22 (2003) 1665-75

- L. Borrmann, R. Schwanbeck, T. Heyduk, B. Seebeck, P. Rogalla, J. Bullerdiek and J. R. Wisniewski, High mobility group A2 protein and its derivatives bind a specific region of the promoter of DNA repair gene ERCC1 and modulate its activity. Nucleic Acids Res. 31 (2003) 6841-51

- L. Borrmann, S. Wilkening and J. Bullerdiek, The expression of HMGA genes is regulated by their 3'UTR. Oncogene 20 (2001) 4537-41

- P. J. Bradley, 'Metastasizing pleomorphic salivary ademona' should now be considered a low-grade malignancy with a lethal potential. Curr. Opin. Otolaryngol. Head Neck Surg. 13 (2005) 123-6

- J. R. Brants, T. A. Ayoubi, K. Chada, K. Marchal, W. J. Van de Ven and M. M. Petit, Differential regulation of the insulin-like growth factor II mRNA-binding

protein genes by architectural transcription factor HMGA2, FEBS Lett. 569 (2004) 277-83

- R. Breathnach, C. Benoist, K. O'Hare, F. Gannon and P. Chambon, Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. Proc. Natl. Acad. Sci. U S A. 75 (1978) 4853-7

- K. Broberg, M. Zhang, B. Strombeck, M. Isaksson, M. Nilsson, F. Mertens, N. Mandahl, I. Panagopoulos, Fusion of RDC1 with HMGA2 in lipomas as the result of chromosome aberrations involving 2q35-37 and 12q13-15. Int. J. Oncol. 21 (2002) 321-6

- M. Bustin, Revised nomenclature for high mobility group (HMG) chromosomal proteins. Trends Biochem. Sci. 26 (2001) 152-3

- M. Bustin and R. Reeves, High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function. Prog. Nucleic Acid Res. Mol. Biol. 54 (1996) 35-100

- J. Carol, M. C. Gorseling, C. F. de Jong, H. Lingeman, C. E. Kientz, B. L. van Baar and H. Irth, Determination of denaturated proteins and biotoxins by on-line size-exclusion chromatography-digestion-liquid chromatography-electrospray mass spectrometry. Anal. Biochem. 346 (2005) 150-7

- C. Castagnoli, M. Stella, G. Magliacani, S. Ferrone and P. M. Richiardi. Similar ectopic expression of ICAM-1 and HLA class II molecules in hypertrophic scars following thermal injury. Burns 20 (1994) 430-3

- K. Y. Chau, P. Arlotta, U. A. Patel, C. Crane-Robinson, G. Manfioletti and S. J. Ono, A novel downstream positive regulatory element mediating transcription of the human high mobility group (HMG) I-C gene. FEBS Lett. 457 (1999) 429-36

- K. Y. Chau, U. A. Patel, K. L. Lee, H. Y. Lam and C. Crane-Robinson, The gene for the human architectural transcription factor HMGI-C consists of five exons each coding for a distinct functional element. Nucleic Acids Res. 23 (1995) 4262-6

- M. T. Chin, A. Pellacani, H. Wang, S. S. Lin, M. K. Jain, M. A. Perrella and M. E. Lee, Enhancement of serum-response factor-dependent transcription and DNA binding by the architectural transcription factor HMG-I(Y), J. Biol. Chem. 273 (1998) 9755-60

- P. Claus, E. Schulze and J. R. Wisniewski, Insect proteins homologous to mammalian high mobility group proteins I/Y (HMG I/Y). Characterization and binding to linear and four-way junction DNA. J. Biol. Chem. 269 (1994) 33042-8

- T. Cui. S. Wei, K. Brew and F. Leng, Energetics of binding the mammalian high mobility group protein HMGA2 to poly(dA-dT)2 and poly(dA)-poly(dT). J. Mol. Biol. 352 (2005) 629-45

- R. A. Currie, Functional interaction between the DNA binding subunit trimerization domain of NF-Y and the high mobility group protein HMG-I(Y), J. Biol. Chem. 272 (1997) 30880-8

- A. L. Dragan, J. R. Liggins, C. Crane-Robinson and P. L. Privalov, The energetics of specific binding of AT-hooks from HMGA1 to target DNA. J. Mol. Biol. 327 (2003) 393-411

- J. Dressler, L. Bachmann, M. Kasper, J. G. Hauck and E. Muller, Time dependence of the expression of ICAM-1 (CD 54) in human skin wounds. Int. J. Legal Med. 110 (1997) 299-304

- J. Dressler, L. Bachmann, P. Strejc, R. Koch and E. Muller, Expression of adhesion molecules in skin wounds: diagnostic value in legal medicine. Forensic Sci. Int. 113 (2000) 173-6

- L. C. Edelstein, L. Lagos, M. Simmons, H. Tirumalai and C. Gelinas, NF-kappa B-dependent assembly of an enhanceosome-like complex on the promoter region of apoptosis inhibitor Bfl-1/A1. Mol. Cell. Biol. 23 (2003) 2749-61

- K. B. Ellwood, Y. M. Yen, R. C. Johnson and M. Carey, Mechanism for specificity by HMG-1 in enhanceosome assembly. Mol. Cell. Biol. 20 (2000) 4359-70

- L. Fajas, C. Paul, A. Vie, S. Estrach, R. Medema, J. M. Blanchard, C. Sardet and M. L. Vignais, Cyclin A is a mediator of p120E4F-dependent cell cycle arrest in G1. Mol. Cell Biol. 21 (2001) 2956-66

- M. Fedele, S. Battista, G. Manfioletti, C. M. Croce, V. Giancotti, and A. Fusco, Role of the high mobility group A proteins in human lipomas, Carcinogenesis. 22 (2001) 1583-91

- M. Ferguson, P. A. Henry and P. A. Currie, Histone deacetylase inhibition is associated with transcriptional repression of the Hmga2 gene. Nucleic Acids Res. 31 (2003) 3123-33

- V. L. Filatov, A. G. Katrukha, T. V. Bulargina and N. B. Gusev, Troponin: structure, properties, and mechanism of functioning. Biochemistry (Mosc) 64 (1999) 969-85

- P. Finelli, G. M. Pierantoni, D. Giardino, M. Losa, O. Rodeschini, M. Fedele, E. Valtorta, P. Mortini, C. M. Croce, L. Larizza and A. Fusco, The High Mobility Group A2 gene is amplified and overexpressed in human prolactinomas. Cancer Res. 62 (2002) 2398-405

- M. Folin and E. Contiero. Electrophoretic analysis of non-human primates hair keratin. Forensic Sci. Int. 83 (1996) 191-9

- D. Foti, E. Chiefari, M. Fedele, R. Iuliano, L. Brunetti, F. Paonessa, G. Manfioletti, F. Barbetti, A. Brunetti, C. M. Croce, A. Fusco and A. Brunetti, Lack of the architectural factor HMGA1 causes insulin resistance and diabetes in humans and mice. Nat. Med. 11 (2005) 765-73

- D. Foti, R. Iuliano, E. Chiefari, and A. Brunetti, A nucleoprotein complex containing Sp1, C/EBPbeta and HMGI-Y controls human insulin receptor gene transcription, Mol. Cell. Biol. 23 (2003) 2720-32

- G. Freyd, S. K. Kim and H. R. Horvitz, Novel cysteine-rich motif and homeodomain in the product of the Caenorhabditis elegans cell lineage gene lin-11. Nature 344 (1990) 876-9

- B. H. Geierstanger, B. F. Volkman, W. Kremer and D. E. Wemmer, Short peptide fragments derived from HMG-I/Y proteins bind specifically to the minor groove of DNA. Biochemistry 33 (1994) 5347-55

- V. Giancotti, A. Bandiera, E. Buratti, A. Fusco, R. Marzari, B. Coles and G. H. Goodwin, Comparison of multiple forms of the high mobility group I proteins in rodent and human cells. Identification of the human high mobility group I-C protein. Eur. J. Biochem. 198 (1991) 211-6

- V. Giancotti, M. T. Berlingieri, P. P. DiFiore, A. Fusco, G. Vecchio and C. Crane-Robinson, Changes in nuclear proteins on transformation of rat epithelial thyroid cells by a murine sarcoma retrovirus. Cancer Res. 45 (1985) 6051-7

- S. C. Gill and P. H. von Hippel, Calculation of protein extinction coefficients from amino acid sequence data, Anal. Biochem. 182 (1989) 319-26

- F. Girard, B. Bello. U. K. Laemmli and W. J. Gehring, In vivo analysis of scaffold-associated regions in Drosophila: a synthetic high-affinity SAR binding protein suppresses position effect variegation. EMBO J. 17 (1998) 2079-85

- J. Goers, A. B. Manning-Bog, A. L. McCormack, I. S. Millett, S. Doniach, D. A. Di Monte, V. N. Uversky and A. L. Fink, Biochemistry 42 (2003) 8465-71

- G. Goodwin, The high mobility group protein, HMGI-C, Int. J. Biochem. Cell. Biol. 30 (1998) 761-6

- W. Grellner, T. Georg and J. Wilske, Quantitative analysis of proinflammatory cytokines (IL-1beta, IL-6, TNF-alpha) in human skin wounds. Forensic Sci. Int. 113 (2000) 251-64

- J. Grievink, R. T. Houterman and K. De Groot, Protein molecular weight computation from sedimentation velocity data. Anal. Biochem. 57 (1974) 137-56

- C. H. Hardman, R. W. Broadhurst, A. R. Raine, K. D. Grasser, J. O. Thomas and E. D. Laue, Structure of the A-domain of HMG1 and its interaction with DNA as studied by heteronuclear three-and four-dimensional NMR spectroscopy. Biochemistry 34 (1995) 16596-607

- T. Hayashi, Y. Ishida, A. Kimura, T. Takayasu, W. Eisenmenger and T. Kondo. Forensic application of VEGF expression to skin wound age determination. Int. J. Legal Med. 118 (2004) 320-5

- C. Henssge, L. Althaus, J. Bolt, A. Freislederer, H. T. Haffner, C. A. Henssge, B. Hoppe and V. Schneider, Experiences with a compound method for estimating the time since death. I. Rectal temperature nomogram for time since death. Int. J. Legal Med. 113 (2000) 303-19

- D. A. Hill, M. L. Pedulla and R. Reeves, Directional binding of HMG-I(Y) on four-way junction DNA and the molecular basis for competitive binding with HMG-1 and histone H1. Nucleic Acids. Res. 27 (1999) 2135-2144

- D. A. Hill and R. Reeves, Competition between HMG-I(Y), HMG-1 and histone H1 on four-way junction DNA. Nucleic Acids Res. 25 (1997) 3523-31

- U. Hirning-Folz, M. Wilda, V. Rippe, J. Bullerdiek and H. Hameister, The expression pattern of the Hmgic gene during development. Genes Chromosomes Cancer 23 (1998) 350-7

- K. Horiike, H. Tojo, T. Yamano and M. Nozaki, Interpretation of the stokes radius of macromolecules determined by gel filtration chromatography. J. Biochem (Tokyo) 93 (1983) 99-103

- J.R. Huth, C. A. Bewley, M. S. Nissen, J. N. Evans, R. Reeves, A.M. Gronenborn and G. M. Clore, The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif, Nat. Struct. Biol. 4 (1997) 657-65

- K. Ikematsu, R. Tsuda, T. Kondo, H. Kondo, K. Ozawa, S. Ogawa and I. Nakasono, The expression of '150-kDa oxygen regulated protein (ORP-150)' in human brain and its relationship with duration time until death. Leg. Med. (Tokyo) 6 (2004) 97-101

- K. R. Johnson, D. A. Lehn, T. S. Elton, P. J. Barr and R. Reeves, Complete murine cDNA sequence, genome structure, and tissue expression of the high mobility group protein HMG-I(Y). J. Biol. Chem. 263 (1988) 18338-42

- L. Kamenev, M. Leclercq and C. Francois-Gerard. An enzyme immunoassay for prostate-specific p30 antigen detection in the postcoital vaginal tract. J. Forensic Sci. Soc. 29 (1989) 233-41

- S. Kang, N. Kassam, M. L. Gauthier and D. H. O'Day, Post-mortem changes in calmodulin binding proteins in muscle and lung. Forensic Sci. Int. 131 (2003) 140-7

- Y. Kohwi and T. Kohwi-Shigematsu, Structural polymorphism of homopurine-homopyrimidine sequences at neutral pH. J. Mol. Biol. 231 (1993) 1090-101

- T. Kondo, J. Tanaka, Y. Ishida, R. Mori, T. Takayasu and T. Ohshima, Ubiquitin expression in skin wounds and its application to forensic wound age determination. Int. J. Legal Med. 116 (2002) 267-72

- T. Kondo, T. Ohshima, R. Mori, D. W. Guan, K. Ohshima and W. Eisenmenger. Immunohistochemical detection of chemokines in human skin wounds and its application to wound age determination. Int. J. Legal Med. 116 (2002) 87-91

- P. F. Kools, and W. J. Van de Ven, Amplification of a rearranged form of the high-mobility group protein gene HMGIC in OsA-CI osteosarcoma cells, Cancer Genet. Cytogenet. 91 (1996) 1-7

- E. Lees, B. Faha, V. Dulic, S. I. Reed and E. Harlow, Cyclin E/cdk2 and cyclin A/cdk2 kinases associate with p107 and E2F in a temporally distinct manner. Genes Dev. 6 (1992) 1874-85

- D. A. Lehn, T. S. Elton, K. R. Johnson and R. Reeves, A conformational study of the sequence specific binding of HMG-I(Y) with the bovine interleukin-2 cDNA. Biochem. Int. 16 (1988) 963-71

- B. Lewin. Genes VIII. Prentice Hall (2003)

- D. Li, H. H. Lin, M. McMahon, H. Ma, and D. K. Ann, Oncogenic raf-1 induces the expression of non-histone chromosomal architectural protein HMGI-C via a

p44/p42 mitogen-activated protein kinase-dependent pathway in salivary epithelial cells. J. Biol. Chem. 272 (1997) 25062-70

- A. H. Ligon, S. D. Moore, M. A. Parisi, M. E. Mealiffe, D. J. Harris, H. L. Ferguson, B. J. Quade, and C. C. Morton, Constitutional rearrangement of the architectural factor HMGA2: a novel human phenotype including overgrowth and lipomas. Am. J. Hum. Genet. 76 (2005) 340-8

- H. Mach, C. R. Middaugh and R. V. Lewis, Statistical determination of the average values of the extinction coefficients of tryptophan and tyrosine in native proteins. Anal. Biochem. 200 (1992) 74-80

- J. F. Maher and D. Nathans, Multivalent DNA-binding properties of the HMG-1 proteins. Proc. Natl. Acad. Sci. USA 93 (1996) 6716-20

- G. Manfioletti, A. Rustighi, F. Mantovani, G. H. Goodwin and V. Giancotti, Isolation and characterization of the gene coding for murine high-mobility-group protein HMGI-C. Gene 167 (1995) 249-53

- G. Manfioletti, V. Giancotti, A. Bandiera, E. Buratti, P. Sautiere, P. Cary, C. Crane-Robinson, B. Coles and G. H. Goodwin, cDNA cloning of the HMGI-C phosphoprotein, a nuclear protein associated with neoplastic and undifferentiated phenotypes. Nucleic Acids Res. 19 (1991) 6793-7

- V. Masciullo, G. Baldassarre, F. Pentimalli, M. T. Berlingieri, A. Boccia, G. Chiappetta, J. Palazzo, G. Manfioletti, V. Giancotti, G. Viglietto, G. Scambia and A. Fusco, HMGA1 protein over-expression is a frequent feature of epithelial ovarian carcinomas. Carcinogenesis 24 (2003) 1191-8

- S. Matsuzawa, H. Kimura, Y. Itoh, H. Wang and T. Nakagawa. A rapid dot-blot method for species identification of bloodstains. J. Forensic Sci. 38 (1993) 448-54

- R. M. Melillo, G. M. Pierantoni, S. Scala, S. Battista, M. Fedele, A. Stella, M. C. De Biasio, G. Chiappetta, V. Fidanza, G. Condorelli, M. Santoro, C. M. Croce, G, Viglietto, and A. Fusco, Critical role of the HMGI(Y) proteins in adipocytic cell growth and differentiation, Mol. Cell. Biol. 21 (2001) 2485-95

- I. Miller, A. C. Gutleb, A. Kranz and M. Gemeiner. Forensics on wild animals: differentiation between otter and pheasant blood using electrophoretic methods. Electrophoresis 16 (1995) 865-8

- J. Minshull, R. Golsteyn, C. S. Hill and T. Hunt, The A- and B-type cyclin associated cdc2 kinases in Xenopus turn on and off at different times in the cell cycle. EMBO J. 9(1990) 2868-75

- S. Miyaishi, T. Kitao, Y. Yamamoto, H. Ishizu, T. Matsumoto, Y. Mizutani, A Heinemann and K. Puschel. Identification of menstrual blood by the simultaneous determination of FDP-D dimer and myoglobin contents. Nippon Hoigaku Zasshi 50 (1996) 400-3

- K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn and H. Erlich. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb. Symp. Quant. Biol. 51 (1986) 263-73

- M. S. Nissen and R. Reeves, Changes in superhelicity are introduced into closed circular DNA by binding of high mobility group protein I/Y. J. Biol. Chem. 270 (1995) 4355-60

- M. D. Partis, D. G. Griffiths, G. C. Roberts and R. B. Beechey, Cross-linking of protein by the omega-maleimido alkanoyl*N*-hydroxysuccinimido esters. J. Prot. Chem. 2 (1983) 263-77

- U. A. Patel, A. Bandiera, G. Manfioletti, V. Giancotti, K. Y. Chau and C. Crane Robinson. Expression and cDNA cloning of human HMGI-C phosphoprotein. Biochem. Biophys. Res. Commun. 201 (1994) 63-70

- R. I. Peirano and M. Wegner, The glial transcription factor Sox10 binds to DNA both as monomer and dimer with different functional consequences, Nucleic Acids Res. 28 (2000) 3047-55

- M. M. Petit, E. F. Schoenmakers, C. Huysmans, J. M. Geurts, N. Mandahl and W. J. Van de Ven, LHFP, a novel translocation partner gene of HMGIC in a lipoma, is a member of a new family of LHFP-like genes. Genomics 57 (1999) 438-41

- M. M. Petit, R. Mols, E. F. Schoenmakers, N. Mandahl and W. J. Van de Ven, LPP, the preferred fusion partner gene of HMGIC in lipomas, is a novel member of the LIM protein gene family. Genomics 36 (1996) 118-29

- D. V. Rao and V. K. Kashyap. Dot blot immunoassay for detection of human semen. J. Immunoassay 13 (1992) 537-44

- R. Reeves and A. P. Wolffe, Substrate structure influences binding of the non-histone protein HMG-I(Y) to free nucleosomal DNA. Biochemsitry 35 (1996) 5063-74

- R. Reeves and M. S. Nissen, The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. J. Biol. Chem 265 (1990) 8573-82

- R. Reeves, Molecular biology of HMGA proteins: hubs of nuclear function, Gene. 277 (2001) 63-81

- P. Rogalla, K. Drechsler, G. Frey, Y. Hennig, B. Helmke, U. Bonk and J. Bullerdiek, HMGI-C expression patterns in human tissues. Implications for the genesis of frequent mesenchymal tumors. Am. J. Pathol. 149 (1996) 775-9

- E. Roijer, A. Nordkvist, A. K. Strom, W. Ryd, M. Behrendt, J. Bullerdiek, J. Mark and G. Stenman, Translocation, deletion/amplification, and expression of HMGIC and MDM2 in a carcinoma ex pleomorphic adenoma. Am. J, Pathol. 160 (2002) 433-40

- N. Rudin and K. Inman, An Introduction to Forensic DNA Analysis 2nd Ed. CRC Press (2002)

- A. Rustighi, F. Mantovani, A. Fusco, V. Giancotti and G. Manfioletti, Sp1 and CTF/NF-1 transcription factors are involved in the basal expression of the Hmgi-c proximal promoter. Biochem. Biophys. Res. Commun. 265 (1999) 439-47

- A. Rustighi, M. A. Tessari, F. Vascotto, R. Sgarra, V. Giancotti and G. Manfioletti, A polypyrimidine/polypurine tract within the Hmga2 minimal promoter: a common feature of many growth-related genes. Biochemistry 41 (2002) 1229-40

- A. J. Sabucedo and K. G. Furton, Estimation of postmortem interval using the protein marker cardiac Troponin I. Forensic Sci. Int. 134 (2003) 11-6

- R. E. Saferstein, Criminalistics: An Introduction to Forensic Science 8th Ed. Prentice Hall (2003)

- S. K. Saha, G. L. Darmstadt, N. Yamanaka, D. S. Billal, T. Nasreen, M. Islam and D. H. Hamner, Rapid diagnosis of pneumococcal meningitis: implications for treatment and measuring disease burden. Pediatr. Infect. Dis. J. 24 (2005) 1093-8

- G. Sandeen, W. I. Wood and G. Felsenfeld, The interaction of high mobility proteins HMG14 and 17 with nucleosomes. Nucleic Acids Res. 8 (1980) 3757-78

- B. Santee and S. K. Henshaw, The abortion debate: measuring gestational age. Fam. Plann. Perspect. 24 (1992) 172-3

- I. Sato, M. Sagi, A. Ishiwari, H. Nishijima, E. Ito and T. Mukai. Use of the "SMITEST" PSA card to identify the presence of prostate-specific antigen in semen and male urine. Forensic Sci. Int. 127 (2002) 71-4

- I. Sato, M. Yoshiike, T. Yamasaki, K. Yoshida, S. Takano, T. Mukai and T. Iwamoto. A dot-blot-immunoassay for semen identification using a polyclonal antibody against semenogelin, a powerful seminal marker. Forensic Sci. Int. 122 (2001) 27-34

- I. Sato, T. Hamabe, K. Yamazaki, and Y. Watanabe. A new method for the determination of the ABO blood type of semen by immunoblotting using anti-ABH antibodies following immunoprecipitation. J. Immuno. Methods 188 (1995) 229-37

- H. Schagger and G. von Jagow, Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal. Biochem. 166 (1987) 368-79

- J. F. Schiltz, A. Rustighi, M. A. Tessari, J. Liu, P. Braghetta, R. Sgarra, M. Stebel, G. M. Bressan, F. Altruda, V. Giancotti, K. Chada and G. Manfioletti, Hmga2 promoter analysis in transgenic mice. Biochem. Biophys. Res. Commun. 309 (2003) 718-23

- R. Schwanbeck, G. Manfioletti and J. R. Wisniewski, Architecture of high mobility group protein I-C.DNA complex and its perturbation upon phosphorylation by Cdc2 kinase, J. Biol. Chem 275 (2000) 1793-801

- Y. Seo, E. Kakizaki and K. Takahama. A sandwich enzyme immunoassay for brain S-100 protein and its forensic application. Forensic Sci. Int. 87 (1997) 145-54

- L. G. Sheflin, N. W. Fucile and S. W. Spaulding, The specific interactions of HMG1 and 2 with negatively supercoiled DNA are modulated by their acidic C-terminal domains and involve cysteine residues in their HMG 1/2 boxes. Biochemistry 32 (1993) 3238-48

- R. Sgarra, M. A. Tessari, J. Di Bernardo, A. Rustighi, P. Zago, S. Liberatori, A. Armini, L. Bini, V. Giancotti and G. Manfioletti, Discovering high mobility group A molecular partners in tumour cells. Proteomics. 5 (2005) 1494-506

- A. Slama-Schwok, K. Zakrzewska, G. Leger, Y. Leroux, M. Takahashi, E. Kas, and P. Debey, Structural changes induced by binding of the high-mobility group I protein to a mouse satellite DNA sequence. Biophys. J. 78 (2000) 2543-59

- D. G. Smyth, W. H. Stein and S. Moore, The sequence of amino acid residues in bovine pancreatic ribonuclease: revisions and confirmations. J. Biol. Chem. 238 (1963) 227-34

- C. Soanes and S. Hawker, Compact Oxford English Dictionary of Current English 3rd Ed. Oxford University Press (2005)

- M. J. Solomon, F. Strauss and A. Varshavsky, A mammalian high mobility group protein recognizes any stretch of six A.T base pairs in duplex DNA. Proc. Natl. Acad. Sci. USA 83 (1986) 1276-80

- N. Sreerama and R. W. Woody, Estimation of protein secondary structure from circular dichrosim spectra: comparison of CONTIN, SELCON and CDSSTR methods with an expanded reference set. Anal. Biochem. 287 (2000) 252-60

- G. Stenman, Fusion oncogenes and tumor type specificity--insights from salivary gland tumors. Semin. Cancer Biol. 15 (2005) 224-35

- P. R. Stout, J. M. Gehlhausen, C. K. Horn and K. L. Klette, Evaluation of a solid-phase extraction method for benzoylecgonine urine analysis in a high-throughput forensic urine drug-testing laboratory. J. Anal. Toxicol. 26 (2002) 401-5

- F. Strauss and A. Varshavsky, A protein binds to a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome. Cell 37 (1984) 889-901

- R. Strick and U. K. Laemmli, SARs are cis DNA elements of chromosome dynamics: synthesis of a SAR repressor protein. Cell 83 (1995) 1137-48

- K. Takahama. Forensic application of organ-specific antigens. Forensic Sci. Int. 80 (1996) 63-9

- T. Takata, S. Miyaishi, T. Kitao and H. Ishizu. Identification of human brain from a tissue fragment by detection of neurofilament proteins. Forensic Sci. Int. 144 (2004) 1-6

- M. A. Tessari, M. Gostissa, S. Altamura, R. Sgarra, A. Rustighi, C. Salvagno, G. Caretti, C. Imbriano, R. Mantovani, G. Del Sal, V. Giancotti and G. Manfioletti, Transcriptional activation of the cyclin A gene by the architectural transcription factor HMGA2. Mol. Cell. Biol. 23 (2003) 9104-16

- D. Thanos and T. Maniatis. The high mobility group protein HMGI(Y) is required for NF-kappa B-dependent virus induction of the human IFN-beta gene. Cell 71 (1992) 777-89

- L. Tjernberg, W. Hosia, N. Bark, J. Thyberg, J. Johansson, Charge attraction and beta propensity are necessary for amyloid fibril formation from tetrapeptides, J. Biol. Chem. 277 (2002) 43243-6

- J. M. Walker, J. R. Hastings and E. W. Johns, The primary structure of a non-histone chromosomal protein. Eur. J. Biochem. 76 (1977) 461-8

- M. Wegner, G. Zastrow, A. Klavinius, S. Schwender, F. Muller, H. Luksza, J. Hoppe, J. Wienberg and F. Grummt, Cis-acting sequences from mouse rDNA promote plasmid DNA amplification and persistence in mouse cells: implication of HMG-I in their function. Nucleic Acids Res. 17 (1989) 9909-32

- W. Weinmann, M. Renz, S. Vogt and S. Pollak, Automated solid-phase extraction and two-step derivatisation for simultaneous analysis of basic illicit drugs in serum by GC/MS. Int. J. Legal Med. 113 (2000) 229-35

- H. M. Weir, P. J. Kraulis, C. S. Hill, A. R. Raine, E. D. Laue and J. O. Thomas, Structure of the HMG box motif in the B-domain of HMG1

- K. L. West, HMGN proteins play roles in DNA repair and gene expression in mammalian cells. Biochem. Soc. Trans. 32 (2004) 918-9

- K. L. West, Y. Ito, Y. Birger, Y. Postnikov, H. Shirakawa and M. Bustin. HMGN3a and HMGN3b, two protein isoforms with a tissue-specific expression pattern, expand the cellular repertoire of nucleosome-binding proteins. J. Biol. Chem. 276 (2001) 25959-69

- K. Wesylik, J. C. Kim and R. G. Amedec, Pleomorphic adenoma. J. La State Med. Soc. 153 (2001) 12-6

- Wikipedia, the Free Encyclopedia, www.wikipedia.com

- A. P. Wolffe, Architectural transcription factors. Science 264 (1994) 1100-1

- X. Wu and C. T McMurray, Calmodulin kinase II attenuation of gene transcription by preventing cAMP response element-binding protein (CREB) dimerization and binding of the CREB-binding protein, J. Biol. Chem 276 (2001) 1735-41

- Y. Ye and H. Ju, DNA electrochemical behaviours, recognition and sensing by combining with PCR technique. Sensors 3 (2003) 128-45

- J. Yie, M. Merika, N. Munshi, G. Chen, and D. Thanos, The role of HMGI(Y) in the assembly and function of the IFN-beta enhanceosome, EMBO J. 18 (1999) 3074-89

- J. Yie, S. Liang, M. Merika, and D. Thanos, Intra- and intermolecular cooperative binding of high-mobility group protein I(Y) to the beta-interferon promoter, Mol. Cell. Biol. 17 (1997) 3649-62

- X. Zhou, K. F. Benson, H. R. Ashar and K. Chada, Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. Nature 376 (1995) 771-4

- X. Zhou, K. F. Benson, K. Przybysz, J. Liu, Y. Hou, L. Cherath and K. Chada, Genomic structure and expression of the murine Hmgi-c gene. Nucleic Acids Res. 24 (1996) 4071-7

- S. Zwilling, H. Konig and T. Wirth, High mobility group protein 2 functionally interacts with the POU domains of octamer transcription factors. EMBO J. 14 (1995) 1198-208