7-3-2007

Specific binding of mammalian high mobility group protein at-hook 2 to the minor groove of at-rich DNAS : thermodynamic and specificity studies

Tengjiao Cui
Florida International University

DOI: 10.25148/etd.FI14061559

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

SPECIFIC BINDING OF THE MAMMALIAN HIGH MOBILITY GROUP PROTEIN AT-HOOK 2 TO THE MINOR GROOVE OF AT-RICH DNAS: THERMODYNAMIC AND SPECIFICITY STUDIES

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CHEMISTRY

by

Tengjiao Cui

2007

To: Interim Dean Mark Szuchman
College of Arts and Sciences

This dissertation, written by Tengjiao Cui, and entitled Specific Binding of the Mammalian High Mobility Group Protein AT-hook 2 to the Minor Groove of AT-Rich DNAs: Thermodynamic and Specificity Studies, having been approved in respect to style and intellectual content, is referred to you for judgment.

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

David Chatfield

Lidia Kos

John T. Landrum

Watson Lees

Fenfei Leng, Major Professor

Date of Defense: July 3, 2007

The dissertation of Tengjiao Cui is approved.

Interim Dean Mark Szuchman
College of Arts and Sciences

Dean George Walker
University Graduate School

Florida International University, 2007

ABSTRACT OF THE DISSERTATION

SPECIFIC BINDING OF THE MAMMALIAN HIGH MOBILITY GROUP PROTEIN AT-HOOK 2 TO THE MINOR GROOVE OF AT-RICH DNAS: THERMODYNAMIC AND SPECIFICITY STUDIES

by

Tengjiao Cui

Florida International University, 2007

Miami, Florida

Professor Fenfei Leng, Major Professor

The mammalian high mobility group protein AT-hook 2 (HMGA2) is a small transcriptional factor involved in cell development and oncogenesis. It contains three "AT-hook" DNA binding domains, which specifically recognize the minor groove of AT-rich DNA sequences. It also has an acidic C-terminal motif. Previous studies showed that HMGA2 mediates all its biological effects through interactions with AT-rich DNA sequences in the promoter regions. In this dissertation, I used a variety of biochemical and biophysical methods to examine the physical properties of HMGA2 and to further investigate HMGA2's interactions with AT-rich DNA sequences. The following are three avenues perused in this study: (1) due to the asymmetrical charge distribution of HMGA2, I have developed a rapid procedure to purify HMGA2 in the milligram range. Preparation of large amounts of HMGA2 makes biophysical studies possible; (2) Since HMGA2 binds to different AT-rich sequences in the promoter regions, I used a combination of isothermal titration calorimetry (ITC) and DNA UV melting experiment to characterize interactions of HMGA2 with poly(dA-dT)$_2$ and poly(dA)poly(dT). My

results demonstrated that (i) each HMGA2 molecule binds to 15 AT bp; (ii) HMGA2 binds to both AT DNAs with very high affinity. However, the binding reaction of HMGA2 to poly(dA-dT)$_2$ is enthalpy-driven and the binding reaction of HMGA2 with poly(dA)poly(dT) is entropy-driven; (iii) the binding reactions are strongly depended on salt concentrations; (3) Previous studies showed that HMGA2 may have sequence specificity. In this study, I used a PCR-based SELEX procedure to examine the DNA binding specificity of HMGA2. Two consensus sequences for HMGA2 have been identified: 5'-ATATTCGCGAWWATT-3' and 5'-ATATTGCGCAWWATT-3', where W represents A or T. These consensus sequences have a unique feature: the first five base pairs are AT-rich, the middle four to five base pairs are GC-rich, and the last five to six base pairs are AT-rich. All three segments are critical for high affinity binding. Replacing either one of the AT-rich sequences to a non-AT-rich sequence causes at least 100-fold decrease in the binding affinity. Intriguingly, if the GC-segment is substituted by an AT-rich segment, the binding affinity of HMGA2 is reduced approximately 5-fold. Identification of the consensus sequences for HMGA2 represents an important step towards finding its binding sites within the genome.

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

1. Literature Review

1.1 HMG proteins

Mammalian high mobility group (HMG) proteins are chromatin-associated, non-histone nucleoproteins, which are essential architectural factors affecting DNA-dependent processes, like DNA replication, repair and gene transcription, in the context of chromatin (Reeves, 2003; Bianchi & Agresti, 2005). HMG proteins are grouped into a superfamily due to their similarity in physical and chemical properties. Based upon their different structures and recognition of distinct DNA substrates, the HMG superfamily can be subdivided into three groups: HMGA, HMGB and HMGN (Bustin, 2001). With respect to their primary structure, all HMG proteins have a similar sequence pattern, which is characterized by basic DNA binding domains at the N-terminal and an acidic tail at the C-terminal. However, each HMG protein has its signature DNA-binding motif and carboxyl end (Goodwin, 1998).

The HMGA family contains three proteins: HMGA1a (previously HMGI), HMGA1b (previously HMGY), and HMGA2 (previously HMGI-C). They bind to AT-rich sequences in the DNA minor groove via three AT-hook DNA-binding domains (Reeves & Nissen, 1990). These proteins are accessory factors for the construction of enhancesomes in the transcription regulation of certain genes (Reeves & Beckerbauer, 2001). By altering of DNA conformation and through specific protein/protein interactions, they promote the subsequent recruitment of other proteins to the binding site. The C-terminal function of HMGA proteins is not completely clear, but has been suggested to be involved in the regulation of DNA binding (Xiao et al., 2000;

Schwanbeck et al., 2001) and in interactions with other nuclear factors. Dysfunction of HMGA proteins has been associated with many benign and malignant tumors (Sgarra et al., 2004).

The HMGB family consists of HMGB1 (previously HMG-1) and HMGB2 (previously HMG-2) proteins, which are the most abundant HMG proteins in the nucleus. By modulating chromatin structure, HMGB proteins regulate the activities of other nuclear proteins, which are involved in transcription, repair and replication (Thomas, 2001; Widlak et al., 2006). The HMG box is the functional DNA-binding motif of the HMGB proteins (Read et al., 1993; Weir et al., 1993); it contains approximately 80 amino acids and has a characteristic L-shaped fold that is formed by three α-helices. These proteins recognize the DNA by shape, rather than by sequence, and have a high affinity for bent or distorted DNA (Bruhn et al., 1992; Ferrari et al., 1992). A hydrophobic surface of the HMG box intercalates into the DNA minor groove, where it unwinds and bends DNA significantly. The ability of the HMG box domain to induce DNA conformational change at a specific site is an important aspect of its biological function. In different DNA-binding models, non-sequence-specific HMGB proteins can function as a flexible partner with other sequence-specific DNA-binding proteins in altering the DNA structure at the binding site (Onate et al., 1994; Romine et al., 1998). Alternatively, certain HMG box-containing proteins first can recognize target sequences using their sequence-specific binding motifs, and later change the DNA conformation at the binding site via its HMG box domain (Landsman & Bustin, 1993).

The HMGN family includes HMGN1 (previously HMG-14) and HMGN2 (previously HMG-17). Both specifically recognize the 146-bp nucleosome core through the nucleosomal binding motif (Paton et al., 1983; Bustin & Reeves, 1996). When binding to a nucleosome, two HMGN1 or HMGN2 cooperatively contact the nucleosomeal core DNA at the major groove flanking the nucleosomal dyad axis, whereas no cooperation between one HMGN1 and one HMGN2 has been discovered (Alfonso et al., 1994). The nucleosomal binding domain is a positively charged stretch of approximately 30 amino acids, enriched with arginine residues (Walker et al., 1979). The C-terminal, called the chromatin unfolding domain, contains a high content of lysine and proline, which is fundamentally required for changing the architecture of a higher-order chromatin structure (Trieschmann et al., 1995). HMGN proteins enhance transcription and replication through their ability to induce decompaction of the nucleosome array on the chromatin fiber; consequently, they are chromatin-modulating regulators, but not putative transcription factors. HMGN mutants with a truncated C-terminus do not activate transcription or de-compact the chromatin, something which appears to be caused by their deprived interaction with core histones and H1. It has been observed that binding of HMGN to chromatin inhibits the acetylation of H3 by PCAF (Herrera et al., 1999), and that HMGN1 competes with H1 binding to the nucleosome (Ding et al., 1997); this strongly suggests a crucial role of HMGN proteins in relief of histone-dependent chromatin condensation. By regulating histone activity on chromatin, HMGN proteins are involved in enhancing diverse DNA-dependent activities at the chromatin level.

Based upon the knowledge regarding the multiple functions of the HMG superfamily *in vivo*, several models have been proposed for their chromatin binding and interactions with other proteins: (1) By bending the DNA, as previously shown for HMGA and HMGB proteins, HMG proteins prevent or facilitate the binding of other factors to the target site, which usually is located in the vicinity (Whitley et al., 1994; Das et al., 2004); (2) HMGN proteins alter chromatin structure by regulating the activity of other chromatin modulating factors (Lim et al., 2006; Postnikov et al., 2006); (3) HMG proteins are part of the chromatin binding complex; for example, HMGA proteins observed in the transcriptional enhancesome (Du et al., 1993; Borrmann et al., 2003); (4) HMG proteins, like HMGA and HMGB, modulate chromatin structure at a specific site, with the assistance of site-specific binding factors (Bianchi & Agresti, 2005); and (5) HMG proteins compete with other chromatin regulators, such as histone, in chromatin binding (Catez et al., 2004). It is noteworthy that, in some cases, more than one mechanism may be involved at a specific time and under certain circumstances; hence, HMG proteins may play a dynamic role in DNA-dependent events.

1.2 HMGA proteins

HMGA proteins are high mobility group AT-hook proteins, of which there are three members: HMGA1a, HMGA1b and HMGA2. HMGA1a and HMGA1b collectively are named HMGA1 protein, since they are encoded by the same gene *Hmga1*, but are different mRNA splicing products (Johnson et al, 1988, 1989). HMGA2 is encoded by different gene, *Hmga2* (Zhou et al., 1996). In terms of amino acid sequence, HMGA1 and HMGA2 are 50% homologous. Both proteins have three

positively charged DNA binding domains, called AT-Hook DNA binding domains, at the N-terminal, and a negatively charged carboxyl terminus (Goodwin, 1998). The AT-hook DNA-binding domains of HMGA proteins are highly conserved, consisting of a 9 amino acid sequence composed of BBXRGRPBB (B=K or R; X=G or P) (Reeves & Nissen, 1990). The C-terminus of HMGA proteins is enriched with glutamic acids, so it is negatively charged. In addition, there is high content of serine and threonine at the C-terminal domain, which can be phosphorylated by nuclear kinases (Sgarra et al., 2004). HMGA1a, HMGA1b and HMGA2 contain 107, 96, and 108 amino acids respectively. HMGA1b has the same sequence as HMGA1a, except for an internal 11 amino acid deletion between the first and second DNA binding domain. Compared to HMGA1a, HMGA2 has a 12 amino acid shorter sequence between the first and second DNA binding domains, but contains an extra 9 amino acid linker region between the third DNA binding domain and the carboxyl C-terminus. *In vivo*, HMGA proteins are architectural transcription factors (Reeves & Beckerbauer, 2001). Although they do not have any transcriptional activity alone, HMGA proteins are important for multiple DNA-manipulation and protein/protein interactions associated with transcription. By specifically contacting transcription factors and/or changing DNA conformation at promoters free of nucleosomes or remodeling nucleosomes in chromatin, HMGA proteins either promote or prevent the binding of transcription factors and other chromatin-regulating factors to the promoter/enhancer region (Arlotta et al., 1997; Mantovani et al., 1998; Fedele et al., 2006). So, by under different mechanisms, HMGA proteins can either enhance or repress the transcription of target genes.

Expression of HMGA protein *in vivo* is stringently regulated, and only is detected in proliferating and undifferentiated cells (Reeves, 2001). Correlating HMGA expression with embryogenesis, and observing that HMGA2 knockout mice exhibit a pygmy phenotype (Zhou et al., 1995), both indicate the functional importance of HMGA protein in cell growth and development. Moreover, because (1) the induction of HMGA expression has been shown in rat and human transformed cells (Giancotti et al., 1987, 1989); (2) the rearrangement of chromosome 12q13-15, where the *Hmga2* gene is located, is associated with many benign human tumors (Hess, 1998); and (3) the over-expression of full-length or C-terminal-truncated HMGA protein has been demonstrated to be oncogenic in transgenic mice model (Fedele et al., 2002), it is apparent that the malfunction of HMGA proteins is connected to the genesis of a variety of tumors *in vivo*.

1.3 HMGA proteins in Transcription regulation

1.3.1 HMGA1 protein in human interferon β gene transcription

HMGA proteins are called architectural transcription factors (Grosschedl et al., 1994), because they do not have transcriptional activity *per se*, but are important for the construction of transcription machinery through protein/protein and protein/DNA interactions (Goodwin, 1998; Reeves, 2000). Specifically, the accessory role of HMGA1 in the transcription regulation of several genes has been described (Reeves & Beckerbauer, 2001). Among them, the transcription regulation of the human interferon β (INF β) gene has been most extensively studied.

The human interferon β gene is virus inducible. Viral induction of the gene is under the control of four positive regulation domains, which are located approximately 100 bp upstream of the transcription initiation site (Du et al., 1993). The four positive regulation domains, named PRD I-IV, span 50 bp of the INF β promoter and contain a specific binding site recognized by transcription factors NF-κB, ATF-2 and IRF (Visvanathan & Goodbourn, 1989; Du & Maniatis, 1992). NF-κB binds to a consensus sequence in PRD II, which also includes a 5 bp AT-sequence that overlaps with the NF-κB binding site. Specific binding of NF-κB to PRD II has been proven essential for viral induction of the human INFβ gene. *In vivo*, inhibition of the p50/p65 NF-κB heterodimer, by transfecting the cell with an antisense p65 construct, results in a 100-fold decrease in transcription activity (Thanos & Maniatis, 1992). Even though NF-κB is needed for transcription activation of the INFβ gene, it is not the only protein that interacts with the PRD II element *in vivo*. In fact, mutations in PRD II, which keep NF-κB binding intact, decreased the level of viral induction (Visvanathan & Goodbourn, 1989). By screening the nuclear factors against PRD II, HMGA1 later was discovered binding to PRD II as well. Although the NF-κB and HMGA1 binding sites identified by DNase I footprinting partially overlap, there is no competition between NF-κB and HMGA1 when both bind to the INF β promoter. On the contrary, HMGA1 can stimulate NF-κB binding to PRD II. Methylation interference analysis of PRD II fragment, reveals how HMGA1 contacts PRD II at the minor groove of the 5'-AAATT-3' sequence, whereas NF-κB binds to the major groove at the same time. By changing the DNA conformation at PRD II, HMGA1 not only enhances the binding of NF-κB, it also

facilitates cooperation between NF-κB and ATF-2 through its interaction with both of these transcription factors (Du et al., 1993). In addition to the binding site in PRD II, two other HMGA1 binding sites have been identified in the flank region of the ATF-2 binding site in PRD IV. Mutation of any of these HMGA binding sites, especially the one in PRD II, significantly decreases the INFβ promoter activity, which proves that HMGA1 binding is required for the transcription activation. Furthermore, if an antisense HMGA1 construct is introduced, the human INF β gene is no longer virus inducible (Thanos & Maniatis, 1992).

More recently, another form of regulation of the human interferon β gene has been described. The formation of the enhanceosome complex at the promoter region creates a novel activation surface that recruits the CREB-binding protein, CBP/p300 (Merika et al., 1998). CBP is an enormous protein with multiple functional motifs involved in interactions with different transcription factors, cofactors and basal transcription machinery (Chan & La Thangue, 2001). Interestingly, CBP has intrinsic histone acetyltransferase (HAT) activity (Bannister & Kouzarides, 1996). This HAT activity turns out to be important for both optimum activation and turnoff of the human INFβ gene. Recruiting CBP and the CBP-related cofactor P/CAF to the enhanceosome, the HMGA1 protein is acetylated by CBP and P/CAF at K 65 and K 71, respectively. *In vitro*, acetylation of HMGA1 by CBP, but not P/CAF, decreases HMGA1 binding to PRD II. This result is consistent with the three-dimensional structure of HMGA1 bound to PRD II, in which K 65, but not K 71, makes backbone contact with the DNA (Munshi

et al., 1998). To investigate the effect of acetylation of HMGA1 *in vivo*, COS cells have been co-transfected with a CAT reporter vector along with either wild-type CBP and P/CAF or CBP and P/CAF bearing amino acid substitutions in their HAT domains. It was demonstrated that CAT activity increases two- to three-fold in the presence of wild type CBP or P/CAF; and that CBP, but not P/CAF, HAT activity is required for the dissociation of HMGA1 from PRD II, which in turn causes the post-induction turnoff of the INFβ gene.

Based upon previous research on the transcription of the human INF β gene, Thanos et al., in 2000, described the details of this canonical model (Agalioti et al., 2000). They found that the initiation of transcription of the INF β gene starts at a nucleosome-free promoter, followed by recruitment of different transcription factors to the enhanceosome in a well-defined order. Furthermore, chromatin remodeling, which removes a nucleosome (nucleosome II) that masks the TATA box, has been proven important for the recruitment of TFIID to the basal transcription complex. According to their study, GCN-5, not CBP, acetylates the HMGA protein at lysine 71 in the early stages of the enhancesome pack, which solves the puzzle of why CBP-mediated acetylation of HMGA protein seems to be required for both optimal transcription initiation and posttranscriptional turnoff (Munshi et al., 1998).

1.3.2 HMGA1 protein in E-selectin gene transcription

In addition to mediating binding of NF-κB to the PRD II element in the human

interferon β gene promoter, a similar relationship between HMGA1 and NF-κB regulating transcription of the E-selectin gene also has been observed (Whitley et al., 1994). E-selectin is a member of the selectin family of transmembrane glycoproteins, which is important for leukocyte adhesion to vascular endothelium during an inflammatory response (Pober & Cotran, 1990). Expression of the E-selectin gene is stringently regulated; it only is expressed in endothelial cells following exposure to either interleukin-1β (Il-1) or tumor necrosis factor alpha (Ghersa et al., 1992). A careful analysis of the *cis* regulatory elements in the E-selectin gene has revealed a striking similarity in the organization of E-selectin and the human interferon β promoter. Four positive regulatory domains (designated PDs) were identified in a 150 bp region that is upstream of the transcription start site. Three of these domains (PD I, PD III and PD IV) contain NF-κB binding sites, while PD II contains a putative binding site for ATF2 transcription factor. By NF-κB and ATF-2 alone, the promoter cannot be maximally activated, so other transcription cofactors, which also may be involved in the transcription complex, were screened. It became evident that the HMGA1 protein was the cofactor that also specifically binds to the E-selectin promoter. In the transcription regulation region, four HMGA1 binding sites were identified using DNase I footprinting; three of them were mapped in PD I, PD III and PD IV; the other HMGA1 binding site was found in PD II, located at the 3' flanking region of the ATF-2 binding site. All these HMGA1 binding sites featured an AT-rich sequence with a length of 9 bp in PD II and PD III and 5 bp in PD IV and PD I. Specific binding of HMGA1 to these determined binding sites was confirmed further by gel mobility shift assay, using a DNA fragment

containing each PD as a probe. The specific binding of HMGA1 to the positive regulatory domains indicates that this protein might facilitate the binding of NF-κB and ATF-2, as observed in the human interferon β promoter (Du et al., 1993).

As expected, by comparing the binding of NF-κB with PD III and PD I in the presence and absence of HMGA1 protein, HMGA1 was found to be able to significantly enhance NF-κB binding. This observation has triggered interest in knowing if Il-1-induced transcription activation is HMGA1-regulated. In order to answer this question, the E-selectin gene was put under the control of either wild type or mutated promoters, and the expression level of the E-selectin gene was monitored *in vivo*. The mutated promoter either bears disrupted NF-κB binding sites without affecting HMGA1 binding, or, conversely, the A/T core in the HMGA1 binding sites is replaced by G/C, which completely eliminates HMGA1 binding but leaves NF-κB binding unaffected. Results from this *in vivo* experiment showed that mutation in NF-κB binding almost completely abolish promoter activity; meanwhile, the promoter containing mutated HMGA1 binding sites results in at least a 10-fold decrease in Il-1-induced transcription activity. It, thus, has been proven that both NF-κB and HMGA1 are essential for full activity of the E-selectin promoter.

Moreover, *in vitro*, HMGA1 also strongly supports ATF2 binding to PD II on the E-selectin promoter. The stimulatory effect of HMGA1 on binding of ATF2 with PD II may be DNA-independent, since the PD II-ATF2 complexes that are formed in the

presence or absence of HMGA1 exhibit the same mobility during electrophoresis (Whitley et al., 1994). This is different from what has been observed in HMGA1 interactions with NF-κB and PD III, in which a ternary complex of NF-κB, PD III and HMGA1 is detected. The presence of HMGA1 protein on the NF-κB-PD III complex suggests that HMGA1 consistently may be required to maintain the DNA conformation at the NF-κB binding site, or that HMGA1 may be needed for the interaction of other transcription cofactors with NF-κB in the context of transcription machinery.

1.3.3 HMGA2 protein in human interferon β gene transcription

Since the early 1990s, accumulating results have demonstrated that HMGA1 functions in gene transcription as an architectural transcription factor. However, little is known about HMGA2, the other member of the HMGA family, in terms of transcription regulation. Mantovani et al. (1998) first reported that NF-κB mediated activation of the human interferon β promoter can be enhanced by HMGA2. This result implicated all members of the HMGA family as potential participants in gene transcription. As a homologue to HMGA1, it was shown that HMGA2 can bind to PRD II as well as HMGA1, and that it can increase the transcription potential of NF-κB, as HMGA1 does. Disruption of AT-rich core of PRD II eliminates binding of HMGA2, which also abolishes HMGA2-dependent transcription enhancement *in vivo*. Like HMGA1, HMGA2 also increases the binding affinity of NF-κB for PRD II *in vitro*. The effect of HMGA2 binding to PRD II and potentially interacting with NF-κB was studied further in the differentiated rat epithelial cell line, PC C13. This cell line does not express

endogenous HMGA1 protein, a condition that is perfect for studying the transcription activation of INF β gene by HMGA2 alone. By comparing luciferase activity in the presence of either HMGA1 or HMGA2, Goodwin et al. concluded that HMGA2 has the same capacity as HMGA1 as an activator of INF β gene transcription.

Even though it was shown that HMGA2 functions similarly to HMGA1 in stimulating the binding of NF-κB to PRD II and enhancing the transcription activity of the human interferon β promoter, this is not true in all instances. First of all, although the HMGA1 and HMGA2 share highly conserved DNA binding domains, the sequence outside of the DNA binding domains are distinct (Johnson et al., 1988; Manfioletti et al., 1991). This difference may affect both their DNA binding specificity and interactions with other proteins *in vivo* (Huth et al., 1997). Secondly, in a specific transcription complex, different protein/protein interactions are involved; HMGA1 and HMGA2 may recruit different transcription factors, based upon specific protein/protein interactions. Thirdly, the promoter structure also may determine which protein will be recruited. Regulatory elements, which contain multiple AT sites, interact with HMGA protein in different ways (Maher & Nathans, 1996; Piekielko et al., 2001). In this context, the different lengths of the spacer region between the AT-hooks of the HMGA protein might be important. A specific arrangement of the AT-rich sequence on the promoter might recruit only one of these HMGA proteins preferentially.

1.3.4 HMGA2 protein interferes with the HD binding

An interesting example of HMGA2 in gene transcription regulation has been demonstrated by its interference with homeodomains (HDs) (Arlotta et al., 1997). Homeodomains are 61-amino acid-long structures that specifically bind to DNA sequences with a 5'-TAAT-3' core (Gehring et al., 1994). This DNA binding motif has been found in a few of the transcription factors that control cell fate decisions (Kaufman et al., 1990). A simple mechanism that can be involved in the control of HD-containing transcription factors binding to their response element in the gene promoter could be realized via regulation of competitors' activity. It has been demonstrated that HMGA2 can interact with some HD-binding sequences and exclude HD from binding to the same DNA. Using Berenil, a minor-groove binding drug, in the competitive binding of HD to target sequences, it has been demonstrated that HMGA2 protein competes with HD, not only by binding to the same sequence, but also through DNA conformational changes. *In vivo*, with the expression of HMGA proteins, the transcription activity of the HCR element, which contains several 5'-TAAT-3' motifs, was 50% inhibited.

1.3.5 HMGA2 protein in cyclin A gene transcription

Cyclin A is a key regulatory protein which, in mammalian cells, is involved in both the S phase and the G2/M transition of the cell cycle through its association with distinct cdks. (Rosenblatt et al., 1992). Transcription of the cyclin A gene can be repressed when $p120^{E4F}$ binds to the CRE regulation element on the promoter (Fajas et al., 2001). Recently, it was shown that HMGA2 protein can counteract $p120^{E4F}$-mediated inhibition (Tessari et al., 2003). *In vitro*, HMGA2 specifically binds to $p120^{E4F}$, which

requires the two N-terminal zinc-fingers of $p120^{E4F}$, and the second AT-hook with flanking spacer region between the second and third AT hooks. Similar binding also has been observed during interactions between HMGA2 and NF-κB. As suggested by the electrophoretic mobility shift assay (EMSA), HMGA2 competes with $p120^{E4F}$ in terms of binding to the promoter of cyclin A. Its relevance to the regulation of the cyclin A gene *in vivo* has been demonstrated by chromatin immunoprecipitation experiments using either purified antibody for HMGA2 or $p120^{E4F}$. Remarkably, HMGA2 was found to be associated with the cyclin A gene during the 6 to 24 hours after stimulation of G_0 arrested cells, but was not detectable during the G_0 or G_1 phase. HMGA2 binding with the cyclin A gene peaked 18 hours after stimulation, which corresponded to the S phase, during which transcription was active. In contrast, no binding of $p120^{E4F}$ to the cyclin A gene was detected at 18 hours, and only very weak binding at other times during G_0 to 24 hours after stimulation. These data reveal the phase-dependent and mutually exclusive association of HMGA2 and $p120^{E4F}$ with the cyclin A promoter, indicating that induction of the cyclin A gene is mediated by HMGA2 counteracting p120 repression. Previously, it has been implied that HMGA2 might be involved in cell proliferation and cell differentiation (Goodwin, 1998); actually, $Hmga2^{-/-}$ knock-out mice exhibited a pygmy phenotype (Zhou et al., 1995). The finding that HMGA2 can regulate the expression of cyclin A *in vivo* provides one possible mechanism for this.

1.3.6 HMGA2-mediated E2F1 activity in pituitary adenomas

It previously has been reported that over-expression of HMGA2 protein in transgenic mice causes them to develop pituitary adenomas (Fedele et al., 2002). This

indicates that abnormal expression of HMGA2 is related to pituitary tumorigenesis. Interestingly, if one mates transgenic mice that over-express HMGA2 with E2F1 knockout mice, the adenomatous phenotype is almost totally rescued in the double mutant mice (Field et al., 1996). Such results imply that HMGA2-dependent pituitary tumorigenesis is mediated through a pathway that requires E2F1. The E2F family of transcription factors is crucial for the expression of several genes required to enter the S phase of the cell cycle (Vigo et al., 1999; Muller et al., 2001). RB (retinoblastoma) protein, also known as a tumor repressor, controls cell cycle progression by suppressing the activity of E2F (Chellappan et al., 1991; Seville et al., 2005). In non-proliferating cells, RB is associated with E2F1 and prevents E2F1 contact with the basal transcription complex by masking the activation domain of E2F1 (Xiao et al., 2003). Moreover, RB recruits class I histone deacetylase proteins (HDAC1) to the E2F1 binding site (Magnaghi-Jaulin et al., 1998). HDAC1 can repress transcription by removing acetyl groups from the histone, which causes condensation of the chromatin, thereby blocking access of transcription factors to the promoter. To better understand the mechanism behind HMGA2's activation of E2F1-mediated transcription, a vector containing an E2F1 response element was fused to the luciferase report gene. Constructs expressing RB and HMGA2 were transformed, either individually or simultaneously, to examine for effects on transcription. Results were that transcription in the cells with an RB plasmid was suppressed, while transcription in cells containing an HMGA2 plasmid was increased. When both plasmids were co-transfected, RB-mediated repression was reversed by increasing the dose of HMGA2. Interestingly, if an HMGA2 plasmid was transfected alone into a RB-knockout cell, transcription activity was not affected. This

indicates that HMGA2 does not interact directly with E2F1, but activates transcriptional activity of E2F1 by releasing it from RB. In addition to releasing E2F1 from RB, binding of HMGA2 to RB simultaneously displaces HDAC1. Under this circumstance, acetylation of the E2F1 binding site opens up chromatin and makes it accessible to transcription factors.

1.3.7 Summary

To date, HMGA proteins have been reported to function in the transcription regulation of approximately 30 genes (Reeves & Beckerbauer, 2001). In most cases, HMGA proteins exert a positive influence on gene transcription. However, negative regulation also has been described (Bustin & Reeves, 1996). Because of their divergent interactions with transcription factors and DNA, different mechanisms have been proposed to explain HMGA-mediated transcription regulation. In some instances, for example, in the case of the human interferon β promoter, HMGA proteins bind to AT-rich sequences in the enhancer, and recruit transcription factors to their specific response elements. This function of HMGA proteins can be mediated through DNA conformational changes at the transcription factor-binding site, which increase the binding affinity between transcription factors and their target sequence. In other instances, HMGA proteins facilitate the binding of transcription factor without contacting DNA, as observed in the transcription activation of c-*fos* and smooth muscle-specific SM22α genes (Chin et al., 1998). A possible explanation for this is that HMGA can associate with target transcription factors, thereby inducing a conformational change required for binding. This kind of interaction recently was implicated in a study of the

ETS transcription factor, PU.1, binding to the enhancer of the immunoglobulin heavy chain gene (Lewis et al., 2001).

In eukaryotic cells, chromatin structure also plays an important role in gene transcription (Wolffe et al., 1998; Aalfs & Kingston, 2000). The observation that HMGA proteins directly interact with H1 (Bonnefoy et al., 1999) and nucleosomes (Reeves & Nissen, 1993) implies that HMGA proteins might execute their transcriptional function by modulating chromatin structure. Experimental results supporting this proposal include the finding that HMGA protein can displace H1 from scaffold attachment regions (SARs) (Zhao et al., 1993) and from HMGA-induced chromatin remodeling in the transcription activation of the human interleukin-2 receptor α chain gene (Reeves et al., 2000). In a quiescent state, the promoter/enhancer regions are protected either by nucleosomes or inhibitory chromatin proteins, such as histone H1. By replacing the chromatin repressor and/or via nucleosome remodeling, HMGA proteins play an important role in opening up the promoter/enhancer region, which makes it much more accessible to transcription factors and cofactors.

1.4 Hmga2 gene and its Transcription regulation

1.4.1 c-DNA cloning of the Hmga2 gene

In 1991, the c-DNA for mouse HMGA2 was cloned for the first time, using Lewis lung carcinoma cells, in which HMGA2 is over-expressed (Manfioletti et al., 1991). Earlier, the c-DNA cloning of the HMGA1a and HMGA1b revealed that the two proteins are products of the same gene, but are alternative mRNA splicing products

(Johnson et al, 1988, 1989). Although HMGA2 is 50% homologous with HMGA1 in terms of its primary structure, it is encoded by a separate gene located on a different chromosome (Zhou et al., 1996). Figure 1 shows the amino acid sequence derived from the HMGA2 c-DNA sequence, in which the open reading frame encodes a 108 amino acid protein with a molecular weight around 12 kDa.

One feature of HMGA2 protein is that it contains high levels of the amino acids lysine, arginine and glutamic acid. The distribution of these charged amino acids on the protein is very asymmetrical. The positively charged residues are almost exclusively concentrated within three 9 amino acid segments called AT-hook DNA binding domains; while the carboxyl acids largely are located at the C-terminal (Reeves & Beckerbauer, 2001). Each DNA binding domain has a sequence of KR(G/P)RGRP(R/K)(K/G), which not only is highly conserved between all members of the HMGA family, but also is homologous among different species (Ashar et al., 1996). In fact, human HMGA2 only differs from its mouse counterpart in four amino acids (none of the four in the DNA binding domains), with an extra glutamic acid residue at position 95 (Patel et al., 1994). NMR structures reveal that the amino and carboxyl groups on the backbone of the AT-hook peptide form hydrogen bonds with the DNA base in the minor groove; meanwhile, the positively charged amino groups on the arginine stick out of the minor groove and form charge-charge interactions with the phosphate groups on the DNA helix (Geierstanger et al., 1994). The C-terminal of the HMGA2 contains 40% glutamic acid and about 33% serine and threonine; consequently, it is highly negatively charged and contains possible phosphorylation target sites for cdc2 kinase (Reeves et al., 1991). It has

been shown that phosphorylation decreases the DNA binding of HMGA proteins, and the phosphorylation is cell cycle-dependent. Function of the HMGA proteins during different stages of cell proliferation and in different cells may be regulated via phosphorylation at the C-terminal.

With respect to its primary structure, HMGA2 is 100% identical to HMGA1 protein in the first and second DNA binding domains; 66% identical in the third DNA binding domain, and 60% identical at the acidic C-terminal. Aside from these conserved regions, little similarity has been observed in the remaining parts of the two proteins (Goodwin, 1998). Figure 1 shows the sequences of HMGA2 aligned with HMGA1a and HMGA1b. Compared to HMGA1a, HMGA1b is characterized by an internal 11-amino acid deletion between the first and second AT-hook DNA binding domains (Johnson et al., 1988; Reeves & Nissen, 1995); hence, in terms of the arrangement of the DNA binding domains, HMGA1b is more closely related to HMGA2 (Ashar et al., 1996). Different from the HMGA1 proteins, HMGA2 has a featured linker between the third DNA binding domain and the C-terminal, so that the DNA binding domains of HMGA2 are further separated from the C-terminal. This specific arrangement of the DNA binding domains in different members of the HMGA family recently has been shown to be related to the proteins' DNA binding properties (Schwanbeck et al., 2000). DNA with multiple AT sequences may have a preference for a specific member of the HMGA family. The most distinct region between HMGA1 and HMGA2 is the 25 amino acids before the first AT-hook DNA binding domain. The spacer region between every single

DNA binding domain potentially is involved in DNA binding and in interactions with other proteins, like different transcription factors (Huth et al., 1997).

```
                10        20        30        40        50        60
                                  *********                 *********
HMGA2     MSARGEGAGQPSTSAQGQPAAPVPQKRGRGRPRKQQQ           EPTCEPSPKRPRGRPKG
HMGA1b    MSESSSKSSQPLASKQEKDGT   EKRGRGRPRKQPP          KEPSEVPTPKRPRGRPKG
HMGA1a    MSESSSKSSQPLASKQEKDGT   EKRGRGRPRKQPPVSPGTALVGSQKEPSEVPTPKRPRGRPKG

              70        80        90       100
                        *********          ----------
HMGA2     SKNKSPSKAAQKKAETIGEKRPRGRPRKWPQQVVQKKPAQETEETSSQESAEED
HMGA1b    SKNKGAAKT  RKTTTTPGRKPRGRPKK     LEK    EEEEGISQESSEEEQ
HMGA1a    SKNKGAAKT  RKTTTTPGRKPRGRPKK     LEK    EEEEGISQESSEEEQ
```

Figure 1. The amino acid sequence of HMGA2 aligned with HMGA1a and HMGA1b. (*: DNA-binding domain; -: C-terminal)

1.4.2 Genomic structure of the Hmga2 gene

The mouse *Hmga2* gene is approximately 150 kb in length, and is composed of five exons and four introns (Manfioletti et al., 1995; Zhou et al., 1996). Compared with the c-DNA sequence, Exon I contains the 5'-UTR and amino acids 1-37, which include the ATG translation start codon and the first DNA-binding domain; Exon II encodes amino acids 38-66, which contain the second DNA-binding domain; Exon III encodes amino acids 67-83, which contain the third DNA-binding domain; Exon IV encodes amino acids 84-94, which contain the linkage region between the third DNA-binding domain and the C-terminal; and Exon V contains amino acids 95-108 and the 3'-UTR (Figure 2). In genomic structure, the *Hmga2* gene is characterized by a very long Intron 3 (>60 kb) between the coding sequences for the third DNA-binding domain and the linker

region. Breakpoints in this intron result in a C-terminal truncated protein, which has been detected frequently in many benign tumors (Hess, 1998).

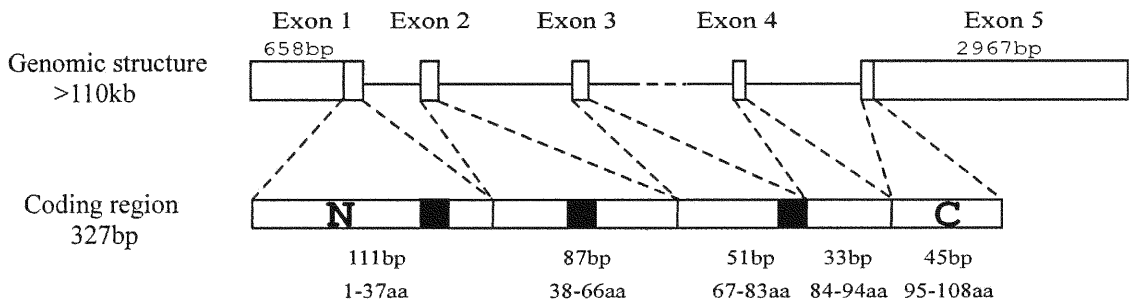


Figure 2. The genomic structure of mouse HMGA2 (DNA-binding domains are represented by solid boxes)

The c-DNA cloning of the mouse Hmga2 gene reveals an unusually long 3-UTR (Zhou et al., 1996). The DNA sequence for the 2.8 kb 3'-UTR has been analyzed. Two polyadenylation signals (AATAAA) have been detected 75 bp apart. The first polyadenylation signal is located 18 bp downstream from the poly(A) tail, and is associated with a TGTGTTCA sequence, 35 bp downstream. Since this combination of a polyadenylation signal and a YGTGTTYY (Y=C or T) element has been described as an efficient formation for the 3' terminus of mRNA, the first polyadenylation signal, instead of the second one, may be used as the main transcription termination signal *in vivo* (McLauchlan et al., 1985). Compared to the 3'-UTR, the 5'-UTR of the mouse Hmga2 gene is more complicated. In fact, three different transcripts, with lengths between 3.8 and 4.2 kb, have been detected by Northern blot analysis, which indicates the existence of multiple transcription start sites.

Characterization of the human *Hmga2* gene reveals a structure that is very homologous with the mouse *Hmga2* gene, in which the N-terminal, AT-hook DNA binding domains, spacer and C-terminal are encoded by five exons (Chau et al., 1995). However, the 3'-UTR and 5'-UTR are distinct. In human 3'-UTR, three possible polyadenylation signals associated with a TGTGTTTT sequence have been identified. The second of these is characterized by a 23 bp space between the two sequences, which meets the requirement of <50 bp between the signal sequence AATAAA and the consensus termination sequence YGTGTTYY (Y=C or T) for the efficient termination of transcription. In an attempt to locate the transcription start site; multiple starting sites have been identified (Chau et al., 1999), which is consistent with the observation of more than one mRNA transcript *in vivo*. A $(CT)_{28}$ repeat interrupted by a G, and a $(CA)_{17}$ repeat interrupted by four Ts can be located between the transcription initiation site and the ATG translation start codon in the 5'-UTR (Chau et al., 1995). The functions of these repeat segments are not yet completely clear, but they are conserved in both the human and mouse *Hmga2* gene.

1.4.3 Transcription regulation of the Hmga2 gene

Analyzing the DNA sequence upstream of the transcription start site, neither a TATA box nor a CAT box was found. However, a ppyr/ppur tract featuring several TCC repeats and multiple GC boxes that potentially bind to the transcription factor Sp1 were present at the 5' flank region of the transcription start site (Zhou et al., 1996; Rustighi et al., 1999). DNase I footprinting already has demonstrated that Sp1 binds specifically to the ppyr/ppur tract and activates the HMGA2 promoter (Rustighi et al., 1999). Since Sp1

has been shown to recruit and stabilize the TFIID complex *in vivo* (Pugh & Tjian, 1991), it can substitute TBP function in transcription from a promoter lacking a TATA box. One feature of the TATA-less promoter is that it usually contains more than one transcription start site. Accordingly, three *Hmga2* transcripts of 4.7, 4.2 and 3.9 kb have been detected by Northern blot analysis (Ayoubi et al., 1999). Sequence analysis has revealed that the two longer transcripts contain longer 5'-UTR and are transcriptional products from alternate transcription initiation sites (Chau et al., 1999). Moreover, it has been determined that the ratio of these transcripts is altered in different cell lines. Even within the same cells, the dominant mRNA species changes in different stages of cell growth. The reason for the coexistence of different transcripts remains unknown. In addition to the Sp1 site, some other putative transcription factor binding sites like ATF, Ets and E2F also have been identified in the 5' region of the transcription start site; this may explain the induction of the *Hmga2* gene as a delayed early gene in G_1, after serum or growth factor stimulation (Ayoubi et al., 1999).

Another mechanism that can be used in transcription regulation involves the posttranscriptional process of mRNA. Such a mechanism has been implicated as potentially involved in expressing the HMGA2 gene, because it has been observed that an active HMGA2 promoter does not express HMGA2 at a level detectable by Northern blot, and that the level of HMGA protein does not always correlate with the amount of mRNA (Klotzbucher et al., 1999). In investigations addressing the possible function of 3'-UTR in posttranscriptional process, three characterized regions have been defined within the 3'-UTR of the HMGA2 transcript (Borrmann et al, 2001). The 3'-UTR

contains 10 repeated AUUUA sequences, in addition to a 58 bp A/U rich sequence 13 bp upstream. The role of the AUUUA motif in the destabilization and rapid degradation of mRNA has been described. An insertion of an A/U-rich element 20-30 bp in front of the AUUUA motif can further increase the destabilization of mRNA (Xu et al., 1997). This is consistent with the observation that of full-length 3'-UTR leads to a 12.7-fold decrease in HMGA2 expression, relative to a positive control. This result also can be used to explain why transgenic mice bearing a truncated HMGA2 protein exhibit a neoplastic phenotype (Battista et al., 1999), and why a truncated or short ectopic- sequence fused HMGA2 protein frequently is identified in benign tumors (Hess, 1998). There is another 39 bp G/C-rich sequence conserved in both HMGA1 and HMGA2 that is found at the beginning of the 3'-UTR, though the function of this region remains unknown.

1.5 Association of HMGA proteins with Benign and Malignant tumors

1.5.1 HMGA proteins in cell transformation

Over-expression of HMGA proteins often is associated with a malignant phenotype. When rat thyroid epithelial cells were transfected by a Sarcoma virus or Kirsten murine sarcoma virus that carries a retroviral transforming oncogene, v-*mos* or v-*ras*, respectively, a significant increase was noted in the expression of HMGA proteins (Giancotti et al., 1987, 1989). Conversely, HMGA proteins only are present at a much lower level in untransformed cells, and are almost undetectable in differentiated cells. The role of HMGA proteins in neoplastic transformation has been further confirmed using an antisense HMGA construct (Berlingieri et al., 1995). Berlingieri et al. showed that rat thyroid cells infected with oncogentic retrovirus fail to exhibit a malignant

phenotype, if HMGA2 protein synthesis is inhibited. The cell with an antisense HMGA2 construct can neither grow in soft agar, nor induce any tumors after injection into athymic mice. Interestingly, the expression of HMGA1 protein also was inhibited at the mRNA level when the antisense HMGA2 construct was introduced. Since the expression of HMGA1 protein also correlates with the malignant phenotype of human or rat thyroid cell neoplasms (Chiappetta et al., 1995), and because transgenic mice bearing a disrupted HMGA2 gene develops thyroid carcinomas in the same way that wild-type mice do after a radioactive iodine feed (Scala et al., 2001), it is possible that HMGA1, instead of HMGA2, plays a casual role in malignant cell transformation. In order to address this question, approximately 10 years after their antisense HMGA2 work, Berlingieri et al. (2002) studied HMGA1 function during thyroid cell transformation, using the same strategy. To their surprise, they discovered that transfection of a HMGA1 antisense construct into a normal rat thyroid cell line, followed by infection with Kirsten murine sarcoma virus, failed to result in a malignant neoplasias. Transfected cells carrying the antisense HMGA1 construct also were unable to grow on soft agar or to stimulate tumorgenesis in athymic mice. By RT-PCR, these investigators found that HMGA2 expression was not affected by the presence of an HMGA1 antisense construct, which is different from the inhibition of HMGA1 in the presence of antisense HMGA2 construct observed previously (Berlingieri et al., 1995). They explained their results by postulating a control of HMGA2 on HMGA1 synthesis, but not vice versa. This assumption also can be used to explain why induction of AP-1 transcriptional activity is repressed when either an HMGA1 (Berlingieri et al., 2002) or HMGA2 (Vallone et al., 1997) antisense construct is transfected into rat thyroid cells. Berlingieri et al. concluded that HMGA1

plays a crucial role, and HMGA2 an accessory role in thyroid cell transformation. One noteworthy point is that, though HMGA1 is crucial for cell transformation, over-expression of HMGA1 in normal thyroid cells causes cell death; Cells exhibit earlier entry into the S phase, but delays in G_2/M transition lead to apoptosis (Fedele et al., 2001).

In many benign tumors, a rearrangement in chromosome 12q13-15 causes a break within the third intron of the HMGA2 gene, which results in a C-terminal-truncated HMGA2 or a chimeric HMGA2 containing N-terminal DNA binding domains and ectopic sequences from other genes (Hess, 1998). To determine whether or not a given rearranged HMGA2 gene is potentially oncogenic *in vivo*, Fedele et al. transfected murine NIH3T3 embryonic fibroblasts with a construct expressing either a C-terminal-truncated HMGA2 or a fused HMGA2 containing the three AT-hook DNA binding domains and three LIM domains from the LPP gene (Fedele et al., 1998). They discovered that these two arranged forms of HMGA2 are equal in their ability to transform NIH3T3 cells, and that stable transfectants exhibit typical markers of neoplastic transformation. They can grow on soft agar and can induce cancer when inoculated into athymic mice. Based upon the small difference in transforming properties they observed between the two constructs, Fedele et al. concluded that the truncated HMGA2 containing only three DNA binding domains is sufficient to transform the NIH3T3 cell line. The chimeric sequence obtained from other genes, at least from the LPP gene, had no significant effect on transformation. Their data are consistent with the fact that various fusion sequences have been characterized in different rearrangements;

and in some cases, only a few amino acids are fused to the HMGA2 DNA-binding domains, suggesting that truncation of the HMGA2 gene, rather than its fusion with other genes, is responsible for cell transformation. Despite the cell transformation capacities of these rearranged forms of HMGA2, their transformation ability is relatively weak compared to positive controls transfected by *Ret/MEN2A* or the *ras* oncogene. The foci number is 30-fold lower and appears roughly 1 to 2 weeks later than what is observed in positive controls. However, it seems consistent with the observation that only benign and non-aggressive tumors are associated with arranged forms of HMGA2.

1.5.2 HMGA proteins in the transformation of thyroid cells in vivo

As suggested by studies on the virus-induced transformation of rat thyroid cells, the induction of both HMGA1 and HMGA2 genes is required for thyroid cell oncogenesis (Giancotti et al., 1987, 1989). In 2001, Scala et al. studied this phenomenon *in vivo*, using transgenic dwarf mice, in which the HMGA2 locus had been disrupted (Scala et al., 2001). In their experiments, the pygmy (pg-/-) mice (Green, 1989) and wild type mice were fed radioactive iodine. The mice were sacrificed within 40~60 weeks of treatment, and all thyroid nodules that developed in these animals were analyzed. In terms of carcinoma frequency and histological features, the investigators found that the pygmy (pg-/-) mutants were the same as the wild type mice, which suggests that HMGA2 expression is not required for radiation-induced thyroid cell transformation in vivo. The same phenomenon was observed in another experiment in which E7 oncogene-induced carcinomas were examined. Ledent et al. (1995) used TgE7 transgenic mice carrying a viral oncogenetic E7 gene, under the control of a bovine thyroglobulin

promoter. By crossing the pygmy (pg-/-) and TgE7 mice, they generated TgE7/pg-/- offspring. Unlike their TgE7 parent, the TgE7/pg-/- mice did not develop any well-defined thyroid cancer within 12 months. However, typical thyroid carcinoma did develop in TgE7 mice. In conclusion, the pygmy (pg-/-) and wild type (pg+/+) mice exhibited the same level of susceptibility to radiation or E7-mediated malignant thyroid tumorigenesis, which indicates that HMGA2 is not necessary for *in vivo* thyroid cell transformation. Moreover, over-expression of both HMGA1 and HMGA2 has been characterized in the transformed thyroid cell lines. The expression of both HMGA proteins in these transgenic mice has been analyzed by immunohistochemistry and RT-PCR. Scala et al. identified no HMGA2 expression in homozygous pygmy mice, but HMGA2 was detected in all thyroid tumors from the wild type and heterozygous mice. As far as the expression of HMGA1 is concerned, the induction of HMGA1 gene was observed in all samples from wild type, homozygous (pg-/-), heterozygous (pg+/-), TgE7/pg-/- and TgE7 mice. These results, considered together, strongly argue that the induction of the HMGA1 gene *in vivo* is sufficient for malignant thyroid tumor development.

1.5.3 HMGA2 protein in human benign tumors

Chromosome rearrangement targeting 12q14-15, where the HMGA2 gene is located, has been reported in a wide range of benign human tumors, like lipomas (Asher et al., 1995; Petit et al., 1996; Merscher et al., 1997; Nilsson et al, 2005), uterine leiomyomas (Kazmierczak et al., 1995; Hennig et al., 1997; Bhugra et al., 1998), pleomorphic adenomas (Geurts et al., 1997, 1998), aggressive angiomyxomas (Micci et

al, 2006), endometrial polyps (Bol et al., 1996), pulmonary chondroid hamartomas (PCH) (Wanschura et al., 1996; Rogalla et al., 1998), and osteosarcomas (Kools & Van de Ven, 1996). Most, but not all these tumors are of mensenchymal origin. In all these different solid tumor types, the HMGA2 gene recombines with other chromosomal sites through translocations, insertions, and inversions, in which either transcriptional regulation of the HMGA2 gene has been altered, or a chimeric HMGA2 transcript has been synthesized. Different mechanisms have been proposed for these HMGA2-mediated tumors. However, further work is needed to definitively determine causative mechanisms. In all these rearrangements, a breakpoint in the extremely large Intron 3 of the HMGA2 gene appears to be the predominant target. Rearrangements within Intron 3 result in a truncated transcript lacking Exons 4 and 5, or in a fused transcript containing Exons 1 to 3 of the HMGA2 gene linked to other sequences from other known genes or unknown sources. In this situation, some preferred partner gene for specific tumorigenesis has been identified. For example, the LPP (lipoma preferred partner) gene (Petit et al., 1996) and LHFP (lipoma HMGA2 fusion partner) gene (Petit et al., 1999) have been found in lipomas; the FHIT (fragile histidine triad) gene (Geurts et al., 1997) and NFIB (Nuclear factor I/B) gene (Geurts et al., 1998) have been found associated with HMGA2 in pleomorphic adenomas. Diverse collections of partner genes are from different chromosomes and possess a variety of functions. In some cases, different tumors share a partner gene, such as HMGA2-LPP in both lipomas and pulmonary chondroid hamartomas (Rogalla et al., 1998). Conversely, multiple partner genes have been described within a single tumor type, as well (Petit et al., 1999). Even though the rearrangement details behind the phenotype may be distinct in each case, the partner

gene seems not to be a determinant of histological type of tumor. Another interesting point is that, in most instances, the ectopic sequence fused to the DNA binding domain of HMGA2 only encodes a few amino acids before an in-frame stop codon (Kazmierczak et al., 1995; Kools et al., 1996), leading to the postulation that truncation of the HMGA2 protein, rather than the fused sequence, leads to tumorigenesis. Recently, this assertion was supported further by the results of a study on 3'-UTR of the HMGA2 gene, in which the functions of several AUUUA destabilization elements in the regulation of mRNA have been characterized (Borrmann et al, 2001). Borrmann et al also showed that a C-terminal missing HMGA2 has 12-fold more transcription activity than wild type. With this background, a breakpoint in either the third Intro or the 3'-UTR of the HMGA2 gene can result in loss of a negative regulation element in the 3'-UTR. The posttranscriptional process for either a truncated or fused or 3'-UTR lacking transcript will be disrupted, due to the 12q13-15 arrangements. To date, no positive regulation of mRNA resulting from third intron breakage has been reported. However, theoretically, this could happen, since some RNA binding proteins may stabilize the mRNA or facilitate the translational process. Compared to rearrangements in the third intro or the 3'-UTR, disruption of the HMGA2 gene in the 5'-UTR has been reported infrequently. However, such a rearrangement has been described in a few cases of PCH (pulmonary chondroid hamartomas) (Wanschura et al., 1996). Disruption of the HMGA2 gene in this region can alter the transcriptional regulation of *HMGA2 in vivo*. Placing *HMGA2* under the control of a heterogeneous promoter/enhancer may alter the transcription factors required for transcription activation, and its response to specific stimuli.

1.5.4 HMGA proteins in malignant tumors

Over-expression of HMGA proteins has been found to correlate with the development of a highly malignant phenotype of rat thyroid cells transformed by viral oncogenes (Giancotti et al., 1985, 1987). This correlation later was extended to virus-induced thyroid carcinomas *in vivo* and to human thyroid cells (Giancotti et al., 1989; Chiappetta et al., 1995). Conversely, the expression of the HMGA gene in normal, differentiated cells was negligible. Recently, abnormal expression of HMGA proteins has been detected in several human malignant tumors, including beast cancer (Flohr et al., 2003), prostate cancer (Bussemakers et al., 1991) and colon cancer (Fearon & Vogelstein, 1990). For instance, prostate cancer currently is the second most common cancer effecting males in the United States males. In men older than 50 years, one out of 11 will be diagnosed with prostate cancer, and one-third of those affected will die from this disease (Boring et al., 1991). A significant increase in HMGA expression characterizes this malignant tumor. In 71 clinic samples analyzed by Tamimi et al. (1993), HMGA1 expression was detected in metastatic prostatic tumors, but not in non-metastatic specimens. In accordance with this, the HMGA1 mRNA level correlated with the degree of malignancy; and was not detectable in normal tissue. A similar relationship also exists between HMGA expression and colorectal cancer. In an analysis of samples from human colorectal carcinomas with a malignant phenotype, the expression of HMGA1 was detected at both the mRNA and protein levels in colorectal carcinoma tissue, but not in normal intestinal mucosa. In addition, no HMGA1 expression was detected in colonic polyps; whereas half of the colon adenomas expressed HMGA1. Among the adenomas analyzed, only adenomas positive for HMGA1 exhibited evidence

of cellular dysplasia; and they revealed considerably less intense immunohistochemical staining, relative to malignant colonic tissue. This observation is very similar to what has been observed in benign thyroid diseases, like goiters or adenomas, which suggests that HMGA1 can be used as a marker, both to distinguish between benign and malignant tumors, and to predict the existence of metastases.

The most common cancer in women is breast cancer, which has received considerable research attention, primarily striving towards earlier diagnosis and treatment. After analyzing 212 breast tissue specimens, Chiappetta et al. (2004), concluded that a positive correlation exists between the level of HMGA1 expression and the histological grade of breast carcinomas. Because of this, they suggested that HMGA1 could be used as a novel indicator, both to diagnose and predict prognosis in human breast cancer. HMGA1 expression in breast carcinomas appears to be related to its negative regulation of the *BRCA1* gene (Baldassarre et al., 2003), which has been identified as a risk and probable causative factor for increased susceptibility to familial breast and ovarian cancer (Miki et al., 1994). Mutation in *BRCA1* is associated with approximately 50% of the cases of familial breast cancer, and reduced expression of *BRCA1* is positively correlated with increased invasiveness of human breast cancer (Thompson et al., 1995). Therefore, induction of HMGA1 expression in breast cells might down-regulate *BRCA1* expression, which logically would account for the aggressive phenotype.

More recently, HMGA expression in lung cancer has been described by Sarhadi et al. (2006). They analyzed the expression of both HMGA1 and HMGA2 proteins in tissue from 152 lung carcinoma patients. Using RT-PCR and immunohistochemistry methods, they identified a five and six-fold increase in HMGA1 and HMGA2 expression, respectively, in cancerous versus normal lung tissue; in addition, mRNA levels correlated with the nuclear expression of the two proteins. They also assessed the relationship between cell proliferation and apoptosis and the expression of HMGA proteins. They found that HMGA1 expression levels do not correlate significantly with either cell proliferation or apoptosis; on the other hand, expression of HMGA2 is significantly positively correlated with the degree of cell proliferation. This is consistent with the known role of HMGA2 in the transcriptional activation of the cyclin A gene (Tessari et al, 2003), a gene which is important in cell cycle control, with the established association between *Hmga2* mutations and the mice pygmy phenotype (Zhou et al., 1995). Because over-expression of both HMGA1 and HMGA2 protein has been detected in up to 90% of lung carcinomas, and because increased expression levels of HMGA proteins correlate with poor survival in lung cancer patients, these proteins may be useful markers for use in diagnosing and predicting prognosis and course among lung cancer patients.

Even though both HMGA1 and HMGA2 expression have been established for several human malignant tumors, it is noteworthy that these two proteins may play different roles in tumorigenesis. In fact, accumulating evidence suggests that HMGA1 proteins largely function in cell differentiation; meanwhile, HMGA2 has more of a role

in cell proliferation. First of all, it has been demonstrated that HMGA1, rather than HMGA2, is required for malignant thyroid cell transformation *in vivo*. In fact, malignant transformation is suppressed by the presence of antisense HMGA1 (Berlingieri et al., 2002); second, in human lung cancer, HMGA2, but not HMGA1, correlates with cell proliferation (Sarhadi et al., 2006). A similar observation has been made in the Dunning rat prostate cancer model, in which HMGA1 levels specifically correlate with metastatic capacity, but not with tumor-doubling time (Bussemakers et al., 1991); Third, Tamimi et al. (1993) have noted that, in prostate cancer, HMGA1 expression is confined only to the malignant prostatic structure, and is not detected in adjacent tissues. These results imply distinct, but closely related, functions of the different HMGA proteins in oncogenesis.

Transfection of the rat thyroid cell with antisense HMGA2has been shown to suppress transformation that otherwise is inducible by the viral oncogene *v-mos* or *v-ras* (Berlingieri et al., 1995). Later, HMGA1 antisense was shown to inhibit thyroid transformation *in vivo* (Berlingieri et al., 2002). Prompted by the correlation between HMGA1 level and the level of malignancy in all malignant tumors, Scala et al. (2000) generated a recombinant adenovirus carrying a HMGA1 sequence in an antisense orientation. By infecting two human thyroid anaplastic carcinoma cells (ARO and FB-1), they demonstrated that the antisense construct inhibits the expression of HMGA1, and leads to programmed cell death. Conversely, normal thyroid cells were unaffected. They also demonstrated, *in vivo*, the suppression of xenograft tumor growth in athymic mice, which can be transformed by anaplastic thyroid carcinoma cells in the absence of the

HMGA1 antisense plasmid. This interesting finding shed lights on the selective inhibition of HMGA proteins *in vivo*, which suggests a novel approach to the treatment of HMGA-related human cancers.

1.6. Transgenic Mouse model used in the study of HMGA proteins

1.6.1 Transgenic mice with a pygmy phenotype

While investigating the genomic aberrations associated with mouse spontaneous pygmy phenotype, Zhou et al. (1995) identified a disruption of the Hmga2 gene to be the causative factor. Two pygmy mutants, designated $pg^{TgN40ACha}$ and *pg*, were used to localize the mutation site on the mouse genome. The mutant $pg^{TgN40ACha}$ contains a transgenic insertion at the locus $pg^{TgN40ACha}$; meanwhile, mutant *pg* is spontaneous. By means of a bidirectional chromosomal walk along the normal mouse genome, using a DNA sequence derived from a transgenic insertion in $pg^{TgN40ACha}$, they identified a 56 kb common deletion in the $pg^{TgN40ACha}$ and *pg* mutants. The sequence of the 56 kb deletion was compared with DNA sequence databases from GenBank and EMBL, which revealed 100% homology with the Hmga2 gene. Southern blots, using the full sequence of mouse Hmga2 gene, further confirmed the presence of a disruption at the Hmga2 locus in both mutants, the Hmga2 gene being completed deleted in $pg^{TgN40ACha}$, and the 5' sequence and Exons 1 and 2 being lost in the *pg* mutant. Mutation of the Hmga2 gene also was identified in another pygmy mutant, In(10)17Rk, which was characterized by an inversion of chromosome 10 at the third intron of *Hmga2*. Although distinct mutations have been characterized in different pygmy mice strains, they share a unique growth problem, which is HMGA2-related, rather than related to growth hormone aberration. To

exclude the possibility that the pygmy phenotype results from other genes, they generated a transgenic mouse, whose Exons 1 and exon 2 were replaced by a neomycin-resistance gene. Mating this null-mutant transgenic mouse with a heterozygous mouse carrying one mutated allele produced homozygous mutations at both *Hmga2* loci. As expected, the homozygous (*Hmga2*$^{-/-}$) mice exhibited typical pygmy mice features, such as low birth weight, much-less developed fatty tissue, and 40% the body weight of the wild type mice. Interestingly, the heterozygous mice expressed HMGA2 at approximately 50% the wild type level, suggesting that the wild type allele did not increase its expression to compensate for the loss of the mutant allele. In fact, mice with one mutated *Hmga2* allele exhibited 80% the weight of the wild type mice. Because of this, the *Hmga2* mutation should be considered recessive. Meanwhile, these investigators also found that the expression level of HMGA1 is not affected by the *Hmga2* mutation, which implicats different functions for these two related proteins during cell development. Since it has been was observed that HMGA proteins only are expressed during embryogenesis, and are not detectable in adult cells, their pivotal role in cell proliferation and differentiation has been assumed.

1.6.2 Transgenic mice carrying a C-terminal truncated HMGA2 gene

Rearrangement at chromosome 12q13-15, where the HMGA2 gene is located, has been described in many benign tumors. In most of these rearrangements, a breakpoint in the large third Intron of the HMGA2 gene results in a chimeric transcript containing the DNA binding domain of HMGA2 and an ectopic sequence from another gene. In many cases, the fusion sequence only contains a few amino acids without any

obvious biological function, which indicates that truncation of the HMGA2, rather than its fusion with other gene, is responsible for the cell transformation (Kazmierczak et al., 1995; Hess, 1998). Furthermore, Fedele et al. (1998) have demonstrated that a C-terminal truncated HMGA2 can transform NIH3T3 fibroblasts as well as the HMGA2-LPP fused protein, in which the ectopic sequence does not increase or decrease the transform ability of the truncated HMGA2. Due to its ability to malignantly transform the NIH3T3 cell line, and because it has been observed in many benign tumors, truncated Hmga2 could be oncogenic *in vivo*. To test this hypothesis, Battista et al. (1999) studied the oncogenic potential of truncated Hmga2 using transgenic mice. Transgenic mice bearing a C-terminal truncated Hmga2 gene were generated by means of embryonic stem cells mediated strategy (Soriano et al., 1991; Ledermann, 2000). The cDNA of truncated HMGA2 (*HMGA2/T*), deprived of the C-terminal, was linked to a cytomegalovirus promoter (CMV) and electroporated into ES cells AB2.2. Clones with G418-resistence were selected. Two transfected cell clones expressing the highest level of HMGA2/T mRNA were microinjected into C57BL6/J mouse blastocysts and transferred to foster females. Chimaeric offspring subsequently were mated with wild type C57BL6/J mice to select *HMGA2/T* mutants. Battista et al. found that all of the *HMGA2/T* mice developed a giant and obese phenotype, which exhibited much faster weight gain than wild type mice. By magnetic resonance imaging (MRI) analysis of the fat deposition, the transgenic mice were found to have extensive expansion of the retroperitoneal and subcutaneous white adipose tissue, as well as abundant fat stores in the perirenal and epididymal areas. In addition, other abnormalities were identified in these transgenic mice, such as dilated bladders, mild hydronephorosis, and urinary infections. These

results directly demonstrate the abnormal activities of adipocytes induced by HMGA2 truncation *in vivo*.

At almost the same time, Arlotta et al. (2000) reported their results regarding C-terminal truncated HMGA2 in transgenic mice. Like Battista et al., Arlotta et al. observed a giant phenotype associated with HMGA2 truncation. The transgenic mice were an average of 15 to 28% heavier than non-transgenic mice during the early stages of growth. Gross examination and histologic analysis revealed that the weight gain of the transgenic mice was due to increased adipogenesis. No obvious differences were observed between the transgenic and nontransgenic mice with respect to muscle and skeletal development. While the majority of the obese mice exhibited normal morphology of fat tissue, in 7 of 33, the adipose tissue was abnormal, indicating that the transgene predisposed to lipomagenesis. Considering that lipomas seldom happen spontaneously in mice (the authors reported no such case in 1000 wild type mice analyzed), this result strongly supports a casual role of truncated HMGA2 in lipomagenesis, which is consistent with the transforming ability of truncated HMGA2 that was observed in malignant transformation of NIH3T3 fibroblast cells (Fedele et al., 1998). Even though Western blots confirmed HMGA2/T expression in all tissues except for heart, testes and brain in the transgenic mice, only the adipose tissue was affected. No other tissue abnormalities were observed, suggesting that the tumorigenic effects of HMGA2/T are stringently tissue-specific. In an effort to locate candidate gene whose transcription could be affected by disrupted HMGA2, Arlotta et al. compared the differential expression in the adipocytes from transgenic and non-transgenic mice. They

isolated the differentially expressed DNA and compared the DNA sequence with the NCBI DNA database. The blast results revealed an interesting candidate that encodes the c-Cbl associate protein, CAP (Ribon et al., 1998). CAP is expressed exclusively in adipocytes, and has been identified as an adaptor protein between the insulin receptor and the c-Cbl product in the insulin receptor singling pathway (Ribon et al., 1998). Since CAP is required for the tyrosine phosphorylation of c-Cbl products by insulin receptors, it is essential to the adipocyte response to insulin. It has been demonstrated that transgenic mice have increased expression levels of CAP compared to wild type.mice. Under this circumstance, abnormities that develop in HMGA2/T transgenic mice might be associated with the enhanced CAP-mediated response of adipoctyes to insulin in the insulin-dependent metabolic pathway.

Interestingly, HMGA1 has a negative effect on fat cell development. Melillo et al. (2001) studied the role of HMGA1 in adipocyte growth using the 3T3-L1 cell line as a model system. 3T3-L1 cells are pre-adipocytes, whose growth are arrested at G_1, but which can differentiate into adipocytes upon stimulation by specific agents, like growth hormone or fetal bovine serum (FBS) (Ayoubi et al., 1999). Differentiation of the pre-adipocyte requires a group of transcription factors called CCAAT/enhancer-binding proteins (C/EBPs), which are expressed at specific stages of adipogenesis (Morrison & Farmer, 2000; Farmer, 2005). Expression of these transcription factors and their regulation of target genes happen in a cascade fashion, which eventually promotes the induction of several adipocyte-specific genes, including the gene for the fatty acid-binding protein 422/aP2 and that for leptin (Christy et al., 1989, 1991). Using Northern

and Western blots to analyze the HMA2 levels of 3T3 cells at different stages of differentiation after FBS stimulation, Melillo et al. observed increased HMGA1 levels relative to non-stimulated cells. HMGA1 expression reached a peak between 6 hours and 1 day, and decreased after 4 days. The correlation between *Hmga1* expression and adipocyte differentiation was studied further using antisense methodology. They found that the cell response to hormone stimulation was completely eliminated in the presence of an antisense *Hmga1* construct. Moreover, the induction of two important adipocyte-specific genes aP2 (Distel et al., 1992) and leptin (Hwang et al., 1996), was suppressed. At the same time, they observed a much shorter doubling time in cells transfected by antisense *Hmga1*, indicating an increased cell growth rate and a shorter G_1 arrest than that required for the differentiation of 3T3-L1 cells. These results strongly suggest the need for HMGA1 during adipocyte differentiation. In a study of HMGA1 effects on the transcriptional activation of adipocyte-specific genes like the *obese* gene encoding for leptin (Pelleymounter et al., 1995), they demonstrated that HMGA1 physically interacts with C/EBP factors *in vivo* and *in vitro*. The transcriptional activity of the *obese* promoter is dramatically down regulated, either because of absence of HMGA1 or mutations that impair the interaction of HMGA1 with C/EBP factors. Because transgenic mice (Lep^{ob}/Lep^{ob}) carrying double-mutant *obese* alleles exhibit a giant phenotype, HMGA1 might negatively regulate adipocyte differentiation by enhancing the transcriptional activity of the *obese* gene. Combining the results described by Melillo et al. and other related work using transgenic mice, the proliferation and differentiation of adipocytes seems to be a balance between HMGA2 and HMGA1 activities.

1.6.3 HMGA2 mutation reduces obesity in transgenic mice

Abnormal transcripts resulting from disruptions in the Hmga2 gene have been observed in many benign mesenchymal tumors, including lipomas (Fedele et al., 2001). In addition, *Hmga2*$^{-/-}$ mice exhibit a pygmy phenotype, as well as considerably less developed fatty tissues. These observations have promoted interest in better understanding HMGA2's role in adipogenesis and obesity. In the late 1990's, Anand and Chada investigated the role of *Hmga2* in obesity, using transgenic mice (Anand & Chada, 2000). They first attempted to induce obesity by means of a high-fat diet. In this experiment, wild type mice were fed either standard or high-fat food. RNA isolated from individual fat depots was used to determine the expression of *Hmga2*. They found that RNA isolated from the wild type mice fed standard food did not express HMGA2. However, *Hmga2* expression was detected in RNA isolated from individual fat depots taken from wild type mice after one week on a high-fat diet. Similarly, *Hmga2* expression also was observed in the individual fat depots of two genetically obese mouse models, *Lep*^{ob}/*Lep*^{ob} and *Lepr*^{db}/*Lepr*^{db} (Green, 1989). As a next step, these investigators fed the *Hmga2*$^{-/-}$ and *Hmga2* $^{+/-}$ mutant mice either a standard or high-fat diet. By 26 weeks, they witnessed obesity in the wild type mice, but no significant difference in the final weights of *Hmga2*$^{-/-}$ or *Hmga2*$^{+/-}$ mice, irrespective of diet. Besides, they found that *Hmga2* is semi-dominant, since *Hmga2*$^{+/-}$ heterozygous mutants had final weights that were close to those of wild type mice, when a standard diet was administrated. Interestingly, this semi-dominant property of the *Hmga2* gene was not observed under high-fat diet conditions. In fact, the *Hmga2*$^{+/-}$ mice, like their homozygous sisters, *Hmga2*$^{-/-}$, failed to develop obesity, despite a high-fat diet induction. This result

demonstrated that mutation in one or both *Hmga2* alleles can confer resistance to obesity.

Investigators then used Lep^{ob}/Lep^{ob} transgenic mice to study obesity. Lep^{ob}/Lep^{ob} mice are deficient in the neuroendocrine hormone leptin, so that they genetically develop more fatty tissue than wild type (Lep^{+}/Lep^{+}) mice (Zhang et al., 1994). When Lep^{ob}/Lep^{ob} was placed under the *Hmga2* mutant background, they observed $Hmga2^{+/-}$ Lep^{ob}/Lep^{ob} mice weighed 72% as much as wild type (Lep^{ob}/Lep^{ob}) mice; meanwhile, the double-homozygous mutant, $Hmga2^{-/-}$ Lep^{ob}/Lep^{ob} exhibited 27% the weight of wild type mice. It is known that increased weight of adipose tissue in Lep^{ob}/Lep^{ob} mice is due to hypertrophy and hyper-proliferation of fat cells (Johnson & Hirsch, 1972). However, the reduced weight of the *Hmga2* mutant mice was not due to blocking the conversion of pre-adipocytes to differentiated adipocytes. In fact, the adipocytes in $Hmga2^{-/-}$ Lep^{ob}/Lep^{ob} mice looked similar to those in Lep^{ob}/Lep^{ob} mice, and no difference in either the expression or regulation of adipogenesis-related genes has been identified. Instead, there is a decrease in the number of cells in the respective adipose depots in $Hmga2^{-/-}$ Lep^{ob}/Lep^{ob} versus Lep^{ob}/Lep^{ob} wild type mice.

Based on these results, Anand and Chada proposed that proliferation of undifferentiated pre-adipocytes requires HMGA2 expression. In wild type fatty tissue, HMGA2 is not detectable, because the pre-adipocyte population is too small in the adipose tissue. Under obesity-induction conditions, the number of adipocytes increases, due to expansion of the pre-adipocytes population. However, in the absence of *Hmga2*,

expansion of pre-adipocytes is severely inhibited, which eventually leads to retarded growth of fatty tissue.

1.6.4 Over-expression of HMGA2 induces pituitary adenomas in vivo

In order to establish the influence of HMGA2 over-expression on tumorigenesis, transgenic mice ($HMGA2^{TG}$) bearing a *Hmga2* gene under the control of a CMV promoter were generated (Fedele et al., 2002). Abundant expression of HMGA2 was observed in all transgenic mice tissues, but not in non-transgenic controls. At the age of 12 months, most of the transgenic mice developed a deformed skull and several symptoms of head tilt. MRI demonstrated a large tumor at the skull base pushing the brain upward against the cranium. Compared to male mice, female mice exhibited a higher susceptibility to this kind of lesion. By 6 months of age, 85% of the females had developed pituitary adenomas, while only 40% of males exhibited the same phenotype at 18 months. Fedele et al. (2006) revealed that the effect of HMGA2 in pituitary adenomas depends upon its interaction with pRB, which thereafter affects E2F1-meditaed gene transcription (Fedeal et al., 2006). It has been described that HMGA2 can transactivate the E2F-binding promoter, by releasing E2F1 from a pRB repressor. Meanwhile, HMGA2 displaces HDAC1 from the pRB/E2F complex, allowing pCAF/p300-mediate acetylation at the E2F1 binding site. This acetylation results in less condensed chromatin at the promoter/enhancer region, which makes it available for transcription factor binding. Fedele et al. demonstrated that the binding of HMGA2 to pRB is required for transformation of Rat-2 cells *in vivo*. Rat-2 cells transfected with HMGA2 mutant deprived of pRB binding ability did not develop any features of

malignant transformation. Moreover, it has been reported that over-expression of HMGA2 promots acetylation of histones and E2F1 protein on E2F-target promoters *in vivo*. These results, considered together, argue that HMGA2 might induce pituitary adenomas by enhancing E2F1 activity. To test this, Fedele et al. generated transgenic mice $HMGA2^{TG}$ $E2f1^{-/-}$ and $HMGA2^{TG}$ $E2f1^{+/-}$ to study HMGA2 over-expression in pituitary adenomas. By the age of 15 months, almost all (13/14) the $HMGA2^{TG}$ $E2f1^{+/-}$ mutants had developed pituitary tumors, as did the $HMGA2^{TG}$ $E2f1^{+/+}$ mice. In contrast, only 25% (4/16) of the $HMGA2^{TG}$ $E2f1^{-/-}$ mice developed pituitary adenomas. This result demonstrates how the loss of E2F1 function significantly decreases the susceptibility of $HMGA2^{TG}$ mice to pituitary adenomas, which further suggests that E2F1 is required for pituitary neoplasia in $HMGA2^{TG}$ mice.

1.7 Biochemical and Biophysical studies of HMGA proteins

1.7.1 CD and NMR studies

HMGA proteins initially were characterized by their low molecular weight (around 10-12 kDa) and their solubility in 5% perchloric acid (Lund et al., 1983). All of them contain about 100 amino acids, so they move fast during electrophoresis. In terms of amino acid composition, HMGA proteins are enriched with both acidic and basic polar residues and prolines, but contain only a few residues with bulky hydrophobic side chains. This unusual amino acid component prevents the polypeptide backbone of HMGA proteins from folding into any defined secondary structure. In fact, the circular dichroism (CD) spectra of HMGA1a in solution have revealed a "random coil" conformation, which is characterized by a 70% unstructured polypeptide chain and only

a 15-20% region with a secondary structural component, like an α-helix or β-sheet (Lehn et al., 1988).

As stated earlier, in each HMGA protein, there are three highly conserved regions, called AT-hook DNA binding domains, which are named after their predicted structure (Reeves & Nissen, 1990). The DNA binding properties of these conserved regions have been studied extensively by different groups (Reeves & Nissen, 1990; Geierstanger et al., 1994; Maher et al., 1996). Among them, Reeves and Nissen originally identified the DNA binding domain in the HMGA1 protein, and subsequently studied the DNA binding properties of the identified region using a synthesized 11 amino acid peptide, TPKRPRGRPKK, corresponding to one of the conserved sequences (Reeves & Nissen, 1990). Using the Hoechst 33258 competition assay, they discovered that this peptide specifically binds to the minor groove of AT-rich DNA, in a manner that is indistinguishable from how the full-length HMGA1 protein binds. However, the binding affinity was much lower than that of the intact protein. The DNA binding specificity of the peptide was elucidated further by DNA footprinting, by which a similar protection pattern in the AT region of the DNA was observed between the peptide and several minor groove binding drugs, such as netropsin and Hoechst 33258. Complementary to this earlier finding, Solomon et al. showed that the preference of HMGA1 to bind to an A/T, instead of G/C, sequence in the minor groove was due to the spatial hindrance caused by the 2-amino group on guanine (Solomon et al., 1986). Replacing this 2-amino group with a hydrogen atom changed the guanine to inosine. They found that HMGA1 can bind to poly(dI-dC)$_2$ as well as poly(dA-dT)$_2$, which

confirms the 2-amino group on guanine is the major obstacle that prevents the binding of HMGA1 to a G/C sequence.

Although, the HMGA protein exhibits a random coil conformation in solution, it adopts a characteristic shape when it binds to DNA. Both molecular modeling and NMR studies have revealed a planar but crescent-shaped conformation of the peptide upon DNA binding (Geierstanger et al., 1994; Evans et al., 1995). The PRGRP core lies deep at the bottom of the DNA minor groove, fitting snugly in the minor groove. The carboxyl and amino groups on the peptide backbone form hydrogen bonds with adjacent AT base pairs, while the trans-prolines on either side of the RGR core direct the flanking arginine or lysine residues out of the minor groove. This specific conformation provides extra contact with the DNA minor groove, in which the side chain of the residues form van der Waals interactions with the walls of the minor groove, and the amino or guanidinium groups form charge-charge interactions with the phosphate groups on the DNA helix. Electrostatic interactions have been shown to be important for the binding of HMGA proteins to DNA substrate. Ethylation of the phosphate group at an ATATTT binding site significantly decreases its binding with HMGA1 (Solomon et al., 1986). Consistently, in studies on HMGA2 binding to poly(dA-dT)$_2$ and poly(dA)poly(dT), the polyelectrolyte effect has been determined to be the major contributor to the binding free energy (Cui et al., 2005). Moreover, compared to certain minor groove binding drugs, HMGA proteins bind to DNA in a more salt-dependent manner, which can be explained if the number of charged groups is proportional to binding affinity. Since these drugs are much less charged than HMGA proteins, their DNA binding affinities are less affected

by salt concentration. Interestingly, if the arginine residues in the PRGRP are replaced with lysine, the peptide completely loses its DNA-binding ability, which indicats a crucial role of the RGR core in binding specificity (Geierstanger et al., 1994). Although arginine and lysine both are positively charged, the remarkable difference in DNA binding affinity between PRGRP and PKGKP underlines the significant difference between these two residues within the context of the binding complex.

The conserved sequence KRPRGRPKK has been thought to be the dominant DNA-binding motif; however, other regions around the DNA-binding domains also contribute to binding (Frank et al., 1998). It has been shown that phosphorylation at serine and threonine residues at the N- and/or C-terminal of the DNA binding domains significantly decreased the DNA binding affinity of HMGA proteins (Piekielko et al., 2001). Furthermore, in a recent NMR study of HMGA1 binding to the interferon β promoter, Huth et al. attributed the significant difference in binding affinity between the second and third DNA binding domains of HMGA1 to their different modular components (Huth et al., 1997). According to their results, there are three modular components that contribute to binding: (1) a central RGR core that sits deep at the bottom of the minor groove; (2) a pair of lysine and arginine residues at either end of the core, which mediate electrostatic and hydrophobic contacts with the DNA backbone; and (3) in the case of the second, but not the third, DNA-binding domain, six amino acids towards the C-terminal of the core that interact with the sugar-phosphate backbone of the double helix. These investigators demonstrated that this last component is the distinguishing feature between the second and third DNA binding motif, and that it

generates additional contact with the DNA surface, which accounts for the higher binding affinity of the second DNA-binding motif.

1.7.2 Posttranslational modification of HMGA proteins

Primarily using a short peptide, EPSEVPTPK, that represents the second DNA-binding domain of HMGA1, and later extending to the full-length HMGA1 protein, Thr 53 appears to be phosphorylated by mammalian cdc2 kinase, both *in vivo* and *in vitro*, in a cell cycle-dependent manner (Nissen et al., 1991). By ^{32}P labeling synchronized NIH3T3 cells at different times during the cell cycle, Nissen et al. discovered that Thr 53 was most extensively phosphorylated in mitotic cells, whereas unphosphorylated in cells of G_0-G_2 phase. Exponentially growing cells exhibited moderate phosphorylation at Thr 53, due to the random distribution of cells in different phases of the cell cycle. The cdc2 kinase recognizes substrate that contains the consensus amino acid sequence (Z)-Ser/Thr-Pro-(X)-Z, where Z usually is lysine or arginine, X frequently is a polar residue, and the parentheses indicates residues that appear at some, but not all sites (Holmes & Solomon, 1996). In contrast to murine HMGA1, which contains only one such consensus sequence that can be phosphorylated by cdc2 kinase at Thr 53 (Reeves et al., 1991), human HMGA1 has two such sites, at Thr 53 and Thr 78 (Reeves & Nissen, 1995). Both of these sites have been shown to be a substrate for cdc2 kinase *in vitro*. Phosphorylation at Thr 53 and Thr 78 can change the DNA binding property of HMGA1 protein. It also has been demonstrated that phosphorylation decreases the DNA binding affinity of HMGA1 (Reeves et al., 1991; Schwanbeck et al., 2000); more interestingly, DNA binding affinity of the phosphorylated HMGA1 has been shown to be more sensitive to the ionic strength

than that of the unphosphorylated HMGA1. At a low salt concentration (50 mM), the difference in binding affinity between phosphorylated and unphosphorylated HMGA1 is only 1.12 fold, but it jumps to approximately 20-fold at a physiological salt concentration (200 mM) (Nissen et al., 1991). This significant difference suggests a possible similarity with respect to changes in the DNA-binding affinity of HMGA1 proteins *in vivo* as they are phosphorylated by cdc2 kinase in a cell cycle-dependent manner.

Assessing the effects of phosphorylation of HMGA1 protein on its binding to the interferon β promoter, poly(dA-dT)$_2$ and 4 H DNA, Piekielko et al. noted decreased DNA binding affinity of HMGA1 protein with all these DNA substrates, when they are phosphorylated at the C-terminal by casein kinase 2 (CK2) (Piekielko et al., 2001). Further phosphorylation by cdc2 kinase additionally attenuates the binding. Using hydroxyl radical protein footprinting to define the changes resulting from cdc2 phosphorylation, they uncovered a general conformational change of HMGA1 protein, both at the DNA-binding domains and within the spanning regions between them. Similarly, a remarkable change in the contact between HMGA2 and interferon β was observed after phosphorylation at Ser 43 and Ser 58, two amino acids which are located at the N- and C-terminal, respectively, of the second AT-hook binding to PRD II (Schwanbeck et al., 2000). DNA footprinting revealed almost abolished protection of the PRD II, after cdc2 phosphorylation, and hydroxyl-radical-based protein footprinting demonstrated a significant change in the complex conformation in the region between the second and third DNA binding domains. These studies demonstrate that the properties and specific functions of HMGA proteins likely are regulated by

phosphorylation. Whereas CK2 phosphorylation plays a constitutive role in protein conformation (Szewczuk et al., 1999), cdc2 kinase modulates conformation in a cell cycle-dependent manner. The DNA-binding affinity of cdc2 phosphorylated HMGA1 and HMGA2 is several times lower than that of the unphosphorylated protein; hence conformational changes resulting from cdc2 phosphorylation may contribute to the disassembly of promoter complexes prior to cells entering the M phase.

In addition to cdc2-mediated phosphorylation, at least two serine residues, Ser 102 and Ser 103, at the C-terminus of HMGA1 have been identified as phosphorylation sites for CK2 (Palvimo & Linnala-Kankkunen, 1989). Even though phosphylation at the C-terminus of HMGA1 by CK2 has been proven less influential upon DNA binding affinity than cdc2-phosphorylation, it has been suggested that CK2-medidated phosphorylation adds extra negative charges to the C-terminal domain, which may affect the interactions between this domain and other basic nuclear proteins, like histones. Additionally, reports on interleukin 4-induced phosphorylation of HMGA1 protein, both in B lymphocytes (Wang et al., 1995) and hematopoietic cells (Wang et al., 1997) provide further evidence for the phosphorylation-dependent regulation of HMGA functions in the transcription of genes related to cell proliferation and differentiation, upon extra- or intra-cellular stimulation.

Besides cdc2 and CK2 modifications, Banks et al. (2000) and Xiao et al. (2000) have described another example of phosphorylation of HMGA1 protein in neoplastic and brain cells, respectively. In both cases, the tumor-promoting phorbol ester (TPA)-

induced activation of protein kinase C (PKC) leads to the phosphorylation of HMGA1 *in vivo*. Xiao et al. reported two major PKC phosphorylation sites: Ser 64, which is adjacent to the C-terminus of the second DNA-binding domain; and Ser 44, located within the spanning region between the first and second DNA-binding domains. They demonstrated that the phosphorylation of HMGA1 by cdc2 and PKC together leads to more than 100-fold decrease in the DNA binding affinity. Although Banks et al. described a similar decrease in DNA binding affinity induced by the phosphorylation of HMGA1 by PKC; by means of MALDI mass spectra, they identified extra target sites for PKC at amino acids 5, 9, 21, and 75-77. They found that both HMGA1a and HMGA1b are more extensively phosphorylated in the malignant human breast epithelial cell line MCF-7/PKC-α than in non-transformed MCF-7 cells. In addition, HMGA1b is more extensively modified by other modifications, like methylation and acetylation, than HMGA1a in both MCF-7/PKC-α and MCF-7 cells. However, the relationship between phosphylation and cell malignancy is not yet clear. It is rational to relate the phosphorylation of HMGA1 to the regulation of target genes involved in cell development.

As for modifications other than phosphorylation, acetylation of Lys 65 at the C-terminus of the second DNA binding domain of HMGA1 has been proven essential for interferon β turnoff after transcription (Munshi et al., 1998). More recently, methylation of HMGA1a at R25 of the RGR core in the DNA binding domain has been shown to be strictly related to the execution of programmed cell death (Sgarra et al., 2003). Among

all the HMG proteins, HMGA1a is the only species that has been found methylated in examined human tumor or transformed cell lines. Interestingly, the methylation of HMGA1a protein remarkably increases during apoptosis (Sgarra et al., 2003). Because chromatin condensation is required for the later stages of cell apoptosis, an increase in HMGA1a methylation could be related to heterochromatin and chromatin remodeling in apoptotic cells.

2. Research objectives

The mammalian high mobility group protein AT-hook 2 (HMGA2) is a non-histone nuclear protein related to cell development and differentiation during embryogenesis. Abnormal expression of this protein is associated with a variety of benign and malignant tumors. HMGA2 contains three positively charged AT-hook DNA-binding domains and a negatively charged C-terminus. This unique asymmetrical charge distribution suggests that HMGA2 may be purified using a combination of cation- and anion-exchange chromatography.

Previous studies showed that HMGA2 regulates transcription through binding to AT-rich DNA sequences in promoter regions. Intriguingly, for different promoters, it binds to different AT-rich DNA sequences. For example, HMGA1a binds to a 27 A-track, i.e. 5'-AAAAAAAAAAAAAAAAAAAAAAAAAAA-3' within the insulin receptor gene promoter E3 region to modulate transcriptional activities. For other promoter regions such as human IFN β promoter, human HIV-1 proviral promoter and human interleukin-15 promoter region, HMGA2 binds to the alternate AT base pairs. So far, the molecular basis of HMGA2 recognizing different AT-rich DNA sequences remains obscure. The energetics that drives the formation of HMGA2-DNA complexes is still unknown. A detail thermodynamic analysis should provide valuable information regarding the mechanism by which HMGA2 recognizes different AT-rich DNA sequences.

Since their discovery, HMGA proteins were shown to preferentially bind to AT-rich DNA sequences in the promoter regions. The early qualitative footprinting studies showed that they could bind to any run of five to six AT base pairs with similar DNA binding affinity. These studies also suggested that HMGA proteins only recognize the minor groove configuration of AT base pairs rather than specific sequences. However, more recent studies showed that HMGA proteins should have sequence specificity. For example, the high affinity binding of HMGA proteins requires two to three appropriately spaced AT-rich sequences as a single multivalent binding site. The DNA binding affinity of this multivalent binding is much higher than that of the single valent binding. More significantly, each HMGA protein always simultaneously binds to two to three runs of AT base pairs in the regulatory transcription regions, such as the human interferon β enhancer, the promoter regions of interleukin-2 gene and interleukin-2 receptor α-chain gene. These results suggest that HMGA proteins bind to specific DNA sequences as transcriptional factors. More evidence for sequence-specific DNA binding of HMGA proteins comes from the NMR structural studies. The structure of an HMGA1a-DNA complex showed that the AT-hook DNA binding domain binds to the 5'-AAATT-3' sequence in a fixed orientation with the N-terminal arginine of the core RGR sequence located near the 3'-end of the sequence, and the C-terminal arginine located near the 5'-end of the sequence. This binding orientation resulted from hydrophobic interactions of the arginine side chains of the RGR core with the adenine bases of the DNA sequence, and hydrogen bonding between the AT-hook DNA binding domains and the bases of the binding site. These results suggest that HMGA proteins do not randomly bind to any AT-rich DNA sequences. Until now, a systemic investigation of HMGA proteins' sequence

selectivity has not been carried out. An *in vitro* selection of HMGA2's high affinity DNA-binding sites should provide essential information towards identification of the binding sites within genome.

In this dissertation, I have used a combination of biochemical and biophysical experiments to explore the thermodynamics of HMGA2 binding to two AT polymers: poly(dA-dT)$_2$ and poly(dA)poly(dT), and its sequence selectivity. Our hypotheses are: (1) HMGA2 should bind to cation-exchange resins, such as SP-Sepharose, and anion-exchange resins, such as Q-Sepharose. A combination of cation- and anion-exchange chromatography can be used to develop a rapid procedure for purifying HMGA2; (2) HMGA2 binds to AT-rich DNA with high affinity; the binding site size of HMGA2 on poly(dA-dT)$_2$ and poly(dA)poly(dT) should be significantly greater than 5 base pairs; (3) HMGA2 binds to specific AT-rich sequences.

The following are my research objectives:

Objective 1: To develop a rapid procedure for preparation of HMGA2 protein in a large scale.

Objective 2: To characterize the interaction of HMGA2 with AT-rich DNA using a combination of biochemical and biophysical methods.

Objective 3: To investigate DNA sequence specificity of HMGA2 *in vitro*.

3. Large Scale Preparation of the Mammalian High Mobility Group Protein A2 for Biophysical Studies

3.1 Abstract

Due to asymmetrical charge distribution of the mammalian high mobility group protein A2 (HMGA2), which makes HMGA2 bind to both cation- and anion-exchange columns, we developed a rapid procedure for purifying HMGA2 in the milligram range. This purification procedure greatly facilitated biophysical studies, which require large amounts of the protein.

3.2 Introduction

The mammalian high mobility group protein A2 (HMGA2) is a nuclear oncoprotein (Goodwin, 1998; Reeves, 2001). Abnormal expression of HMGA2 and its mutants is directly linked to tumorigenesis of a variety of benign tumors of mesenchymal origin such as lipomas, uterine leiomyomas, and fibroadenomas (Schoenmakers et al., 1995; Klotzbucher et al., 1999; Fedele et al., 2001). These solid tumors are very common in humans. For example, approximately 39% of all hysterectomies performed in the United States annually are uterine leiomyomas (1,361,786 cases from 1994 to 1999 (Keshavarz et al., 2002)) and up to 77% of women of reproductive age may have these benign smooth-muscle tumors (Cramer & Patel, 1990). In many cases, rearrangements of chromosomal bands 12q13-15 where the HMGA2 gene is located cause disruptions of HMGA2's normal functions (Tallini & Dal, 1999; Aman, 1999), which correlate with the formation of the tumors. Over-expression and deregulation of HMGA proteins including HMGA2 also causes several types of malignant tumors such as lung cancer (Sarhadi et

al., 2006), hepatocellular carcinoma (Chang et al., 2005), prostate cancer (Tamimi et al., 1996), oral cancer (Miyazawa et al., 2004), and leukemia (Kottickal et al., 1998). The expression level is correlated with the degrees of malignancy and metastatic potential of the transformed cells (Tallini et al., 1999). A recent study showed that HMGA proteins were expressed in about 90% of lung carcinomas (152 cases), and the expression level is inversely associated with survival rate and prognosis (Sarhadi et al., 2006). These results suggest that HMGA proteins could be used as a biomarker for diagnosing the neoplastic transformation and the metastatic potential of many cancers (Giancotti et al., 1991; Rogalla et al., 1998).

HMGA2 is only expressed in proliferating, undifferentiated mesenchymal cells and is undetectable in normal fully differentiated adult cells (Zhou et al., 1995; Gattas et al., 1999). Disruption of its normal expression patterns causes deregulations of cell growth and differentiation. Results from Chada's laboratory showed that *Hmga2* knock-out mice developed the pygmy phenotype (Zhou et al., 1995). These mutant mice were severely deficient in fat cells and other mesenchymal tissues (almost 20-fold decrease). Furthermore, it was discovered that disruption of the *Hmga2* gene caused a dramatic reduction in obesity of leptin-deficient mice (Lep^{ob}/Lep^{ob}) in a gene-dosage dependent manner: $Hmga2^{+/+}$ Lep^{ob}/Lep^{ob} mice weighed over three times more than $Hmga2^{-/-}$ Lep^{ob}/Lep^{ob} animals, and the weight of $Hmga2^{+/-}$ Lep^{ob}/Lep^{ob} mice was in-between (Anand & Chada, 2000). These results suggest that HMGA2 plays an important role in fat cell proliferation and is a potential target for the treatment of obesity (Anand & Chada, 2000).

Due to its role in tumorigenesis and obesity described above, HMGA2 has been widely studied by a variety of biochemical and biophysical methods (Solomon et al., 1986; Huth et al., 1997; Cui et al., 2005). Unfortunately, it was difficult to purify large quantities of recombinant HMGA2 for biophysical studies such as nuclear magnetic resonance (NMR) and isothermal titration calorimetry (ITC) studies. One reason is that over-expression of HMGA2 is toxic to the host cells (Reeves, 2004), which results in low level of expression of the protein. Another reason is the lack of a rapid purification procedure for large-scale preparation of HMGA2. As reported previously (Bustin & Reeves, 1996), one intriguing feature of HMGA2 is the asymmetrical charge distribution of its primary structure (Figure 3). The positive charges are concentrated in the three "AT hook" DNA binding domains, which tightly binds to the minor groove of AT-rich DNA sequences (Maher & Nathans, 1996; Cui et al., 2005), The negative charges are mainly located in the C-terminus. This unique feature allowed for development of a rapid purification procedure to prepare HMGA2 in large quantities for various biophysical studies. Using a combination of a strong cation-exchange column, such as SP-Sepharose FF column, and an anion-exchange column, such as Q-Sepharose FF column, we routinely purify 50 to 100 mg of HMGA2 from 10 liters of E. coli cell culture over-expressing recombinant HMGA2 (>98% pure).

3.3 Materials and Methods

3.3.1 Materials

Q-Sepharose FF, SP-Sepharose FF, and poly(dA-dT)$_2$ were purchased from Amersham Biosciences (Piscataway, NJ). Isopropyl-beta-D-thiogalactopyranoside

(IPTG), DTT, lysozyme, Na_2HPO_4; NaH_2PO_4; NaCl; glycerol, PMSF, SDS, ethidium, guanidine hydrochloride, urea, ampicillin, kanamycin, acrylamide, N, N'-methylenebisacrylamide were bought from Sigma-Aldrich corporation (St. Louis, MO). All restriction enzymes were obtained from New England Biolabs (Beverly, MA). pfu DNA polymerase was commercially obtained from Stratagene Corporate (La Jolla, CA). pET30a(+) was purchased from Novagen (Madison, WI). The synthetic oligonucleotides were purchased from MWG-Biotech, Inc. (High Point, NC).

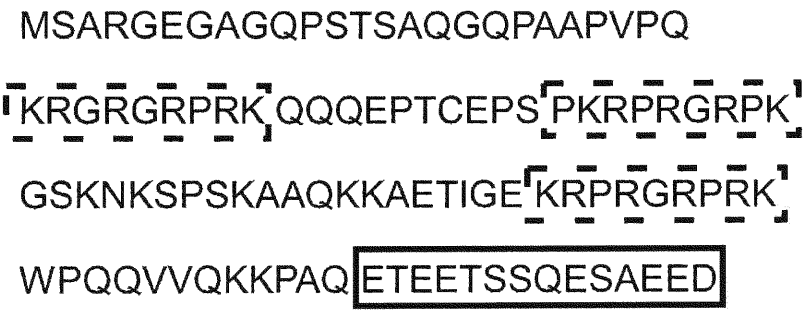


Figure 3. The amino acid sequence of the mouse HMGA2. The positively charged "AT hook" DNA binding domains are boxed with dot lines and the negatively charged C-terminus are boxed with solid lines.

3.3.2 Plasmids

A plasmid, pET21c(+)-*Hmga2* that contains a full length of murine Hmga2 cDNA, which directs expression of HMGA2 with a His-tag in its C-terminus, was kindly provided by Dr. Alfredo Fusco at Universita di Catanzaro, Italy. A point mutation (R47S), which is located in the second AT hook DNA-binding domain, was found in the DNA sequence of *Hmga2* gene in the plasmid. Since this point mutation and the His-tag

may affect the protein's conformation and its DNA binding properties, a PCR-based site-directed mutagenesis (Allemandou et al., 2003) was used to create a plasmid, pMGM1, which expresses the wild type protein HMGA2 without mutation and the His-tag.

3.3.3 HMGA2 purification

E. coli strain BLR(DE3) (Novagen, Madison, WI), which contained the plasmid pMGM1, was grown in Terrific Broth containing 50 μg/ml of the antibiotic kanamycin. The cell growth was monitored by measuring the OD_{595} every hour using an Amersham Ultrospec 2000 UV-VIS spectrophotometer. When the OD_{595} reached 0.6-0.7, HMGA2 expression was induced by adding 1 mM of IPTG into the cell culture. After an additional three to four hour incubation, the cells were harvested by centrifugation at 4,000 rpm for 25 minutes at 4 °C. The supernatant was discarded, and the cell pellet was re-suspended in an ice-cold lysis buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 0.5 mM PMSF, 0.1 mM DTT, 1 mg/ml lysozyme) at 3.5 ml per gram cells. The cells in the lysis buffer were incubated on ice for one hour and then the sample was frozen in liquid nitrogen and stored in a -80 °C freezer. In the following day, the frozen cells were thawed and the salt concentration was adjusted to 1M by adding solid NaCl. The solution was sonicated on ice 8 times at 300W each time with a 10 second interval between each sonication. This was followed by centrifugation of the solution at 16,000 rpm for 20 minutes at 4 °C. After centrifugation, the supernatant was dialyzed against buffer A (50 mM sodium phosphate, pH 8.0, 0.5 mM PMSF, 0.1 mM DTT, 10% glycerol) plus 200 mM NaCl at 4 °C overnight. After dialysis, the solution was loaded onto a 40 ml SP-Sepharose cation-exchange column pre-equilibrated with buffer A plus

200 mM NaCl. The loaded column was washed with 120 ml of buffer A plus 300 mM NaCl. Then the protein was eluted with a 300 ml NaCl gradient of 0.3 to 0.8 M NaCl in buffer A. HMGA2 was eluted at approximately 0.55 M NaCl. Peak fractions containing electrophoretically identified HMGA2 were pooled and dialyzed against buffer A plus 20 mM NaCl. The dialyzed protein was loaded onto a 40 ml Q-Sepharose anion-exchange column pre-equilibrated with buffer A plus 20 mM NaCl. Then the protein was eluted with a 300 ml NaCl gradient 20 to 300 mM NaCl in buffer A. The HMGA2 protein was eluted at approximately 190 mM NaCl.

3.3.4 Measurement of UV absorption and fluorescence emission spectra

Ultraviolet absorption spectrum of HMGA2 proteins in BPE buffer (6 mM Na_2HPO_4, 2 mM NaH_2PO_4, and 1 mM Na_2EDTA, pH 7.0) plus 4 mM NaCl was recorded in an Amersham Ultrospec 2000 UV-VIS spectrophotometer at 24 °C. Fluorescence measurements were made using a Jobin Yvon's FluoroMax-3 spectrofluorometer with λ_{ex} at 280 nm and both slits of 5 nm.

3.3.5 Differential scanning calorimetry (DSC)

DSC experiments were carried out using a VP-DSC calorimeter (Microcal Inc.). Samples were extensively dialyzed against BPE buffer containing 50 mM NaCl. DSC scans were conducted between 0 and 110 °C at a rate of 60 degrees per hour. Baselines, obtained by filling both calorimeter cells with the corresponding buffer, were subtracted from the sample experimental thermograms.

3.3.6 Isothermal titration calorimetry (ITC)

ITC experiments were carried out using a VP-ITC titration calorimeter (Microcal Inc.). Samples were extensively dialyzed against BPE buffer containing 100 mM NaCl. Typically, the titration was set up so that 15 µl of a 50 µM HMGA2 sample was injected every 200 seconds, up to a total of 18 injections, into a DNA sample (1.7 ml of 100 µM (bp)) 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.

3.3.7 Nuclear magnetic resonance (NMR) spectroscopy

Proton NMR spectra of HMGA2 were recorded at 25 °C on a Bruker Avance 600 FT NMR spectrometer operating at a proton frequency of 600.13 MHz. The residual solvent signal was suppressed with presaturation during relaxation delay. Chemical shift values were referenced to the residual HDO signal at 4.81 ppm.

Phase-sensitive NOESY spectra of HMGA2 were acquired at 25 °C with mixing times ranging from 100 ms to 800 ms. Typical NOESY spectra were collected with 256

experiments in F1 dimension using the hypercomplex method of States et al. In general, 80 scans were accumulated for each F1 experiment, which was acquired with 4096 complex points in F2 dimension over a spectral width of 9 to 12 kHz. The residual solvent signal in all NOESY experiments was suppressed using a 2s presaturation with a weak decoupler power. All 2D data were processed on a Dell Dimension 8400 PC with Pentium 4 processor using Felix 2004 (Accelrys, Inc.).

3.4 Results and Discussion

3.4.1 Purification of HMGA2

The mammalian HMGA2 is a small nuclear protein with several unique biochemical and biophysical properties: it has unusual high contents of basic, acidic, glycine, and proline residues; the charge distribution is asymmetrical with the positive charges mainly concentrated in the "AT hook" DNA binding domains and the negative charges located in the C-terminus (Figure 3); and the protein does not have a defined secondary structure (Huth et al., 1997). Therefore, we reasoned that HMGA2 should bind to both cation- and anion-exchange columns, such as SP-Sepharose FF and Q-Sepharose FF columns. Indeed, our results showed that HMGA2 binds to both types of ion-exchange columns. The basic purification steps include cell lysis, sonication, SP-Sepharose and Q-Sepharose ion-exchange chromatography. The purification can be completed in less than four days, which usually produces 50 to 100 milligrams of pure HMGA2 (the purity > 98% as judged by SDS-PAGE gels) from ten liters of cell culture.

(i) Cell growth and lysis. In this study, we used *E. coli* strain BLR(DE3) and a pET vector (pET30a(+)) for expression of HMGA2 in which the mouse *Hmga2* gene was subcloned into *NdeI* and *XhoI* sites to yield pMGM1. Over-expression of HMGA2 can be induced by IPTG. After addition of IPTG, the cell growth was significantly slowed down, which may result from over-expression of HMGA protein that is toxic to the host cells (Reeves, 2004). After additional three to four hour incubation, we harvested the cells by centrifugation at 4 °C. We also adopted a mild lysis procedure for cell lysis. Lysozyme (1 mg/ml) was used to digest the cell wall. After freezing and thawing, the cell's membrane was broken, resulting in a cell lysate with very high viscosity. The viscosity can be greatly reduced by sonication or incubation at 4 °C in the presence of DNase I. Cell debris was removed by centrifugation at 11,000 rpm at 4 °C. This cell lysis procedure is different from the previous one in which 5% perchloric acid or trichloroacetic acid and triton X-100 were used (Reeves, 2004). Although HMGA proteins do not have a defined structure (Huth et al., 1997), this harsh condition may cause the hydrolysis of certain groups of HMGA proteins, especially those isolated from mammalian cells and tissues.

(ii). SP-Sepharose FF column. After dialysis, the cell extract was loaded onto a SP-Sepharose FF column, a strong cation-exchange column equilibrated with the dialysis buffer. We usually washed the loaded column using 2 to 3 column volumes of buffer A plus 300 mM NaCl. No detectable HMGA2 was eluted at this stage. HMGA2 was eluted from the SP-Sepharose FF column using buffer A with a 0.3-0.8 M NaCl gradient. Figure 4a shows the elution profile in which HMGA2 was eluted in buffer A from 400 to

600 mM NaCl (the fractions before 400 mM NaCl are other proteins plus nucleic acid contaminants). There is an additional peak in this graph, which also contained HMGA2.

(iii). Q-Sepharose FF column. The pooled fractions from the SP-Sepharose FF column was dialyzed against buffer A plus 20 mM NaCl and loaded onto a Q-Sepharose FF column equilibrated with buffer A plus 20 mM NaCl. HMGA2 was eluted using buffer A with a 20 to 300 mM NaCl gradient. Figure 4b shows the elution profile in which HMGA2 was eluted in buffer A with approximately 200 mM NaCl. The negatively charged C-terminus is required for HMGA2 binding to the Q-Sepharose column. Therefore, this step is extremely efficient to remove the C-terminal truncated form of HMGA2, which may result from proteolytic degradation of HMGA2 inside the bacterial cells (Reeves, 2004). At this stage, HMGA2 is essentially pure as judged from SDS-PAGE gels (lane 6 of Figure 5). Figure 5 is a 15% SDS-PAGE gel of protein samples at various stages of HMGA2 purification.

3.4.2 Physical properties of HMGA2

Since HMGA2 contains a tryptophan residue, it has a maximum absorbance at 280 nm and also a maximum fluorescence around 360 nm when excited at 280 nm. The extinction coefficients of HMGA2 were determined to be 5,810 $cm^{-1}M^{-1}$ at 280 nm by Von Hippel's method (Gill & von Hippel, 1989). Ionic strength has no apparent effect on the extinction coefficients.

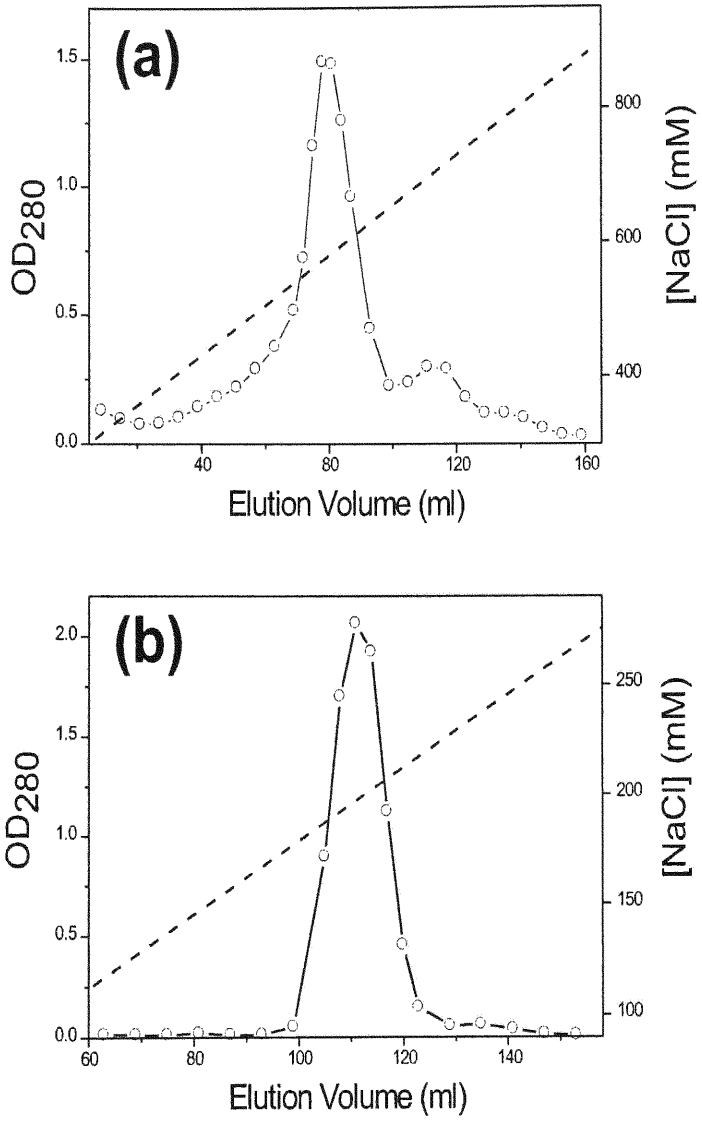


Figure 4. Ion-exchange chromatography of HMGA2. (a) Cation-exchange chromatography of HMGA2 on SP-Sepharose FF column. (b) Anion-exchange chromatography of HMGA2 on Q-Sepharose FF column.

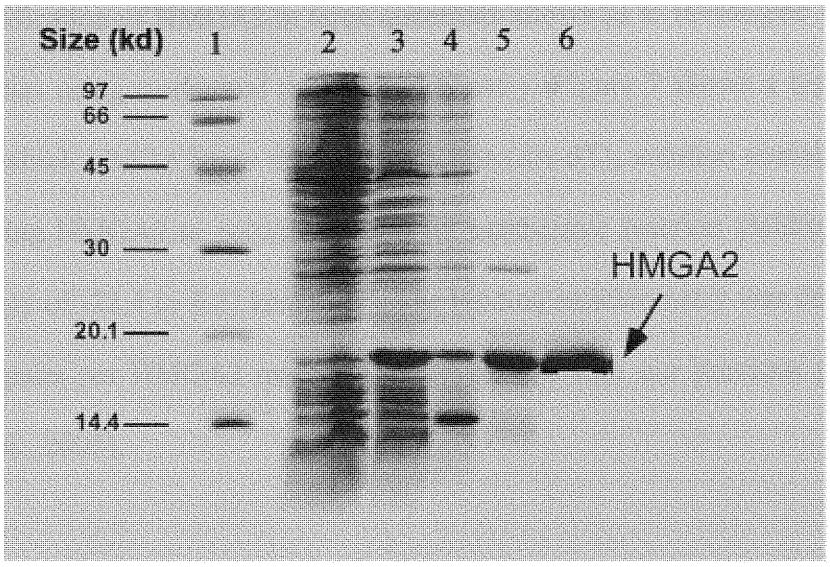


Figure 5. SDS-PAGE (15%) of the fractions from different purification stages. Lane 1. LMW-SDS Marker (Amersham Biosciences). Lane 2. Cell extracts prepared from BLR(DE3)/pMGM1 in the absence of IPTG.. Lane 3. Cell extracts prepared from BLR(DE3)/pMGM1 after 3 hour induction with 1 mM IPTG.. Lane 4. Cell lysates prior to loading onto SP-Sepharose FF column. Lane 5. HMGA2 prior to loading onto Q-Sepharose FF column (pooled peak fractions). Lane 6. Pure HMGA2 after eluted from Q-Sepharose FF column.

As demonstrated previously (Huth et al., 1997), HMGA2 is an intrinsically unstructured protein. Here, we used differential scanning calorimetry (Figure 6a) and 2-D NMR measurement (Figure 6b) to further confirm that HMGA2 is unstructured. No obvious transition in the thermogram and lack of spectral dispersion in NMR spectrum indicate an "unstructured" nature for the protein.

The purification of HMGA2 in large quantities allows us to investigate its DNA-binding properties using isothermal titration calorimetry. Figure 7 shows a typical ITC experiment in which HMGA2 was titrated into a poly(dA-dT)$_2$ solution containing 100 mM Na^+. Binding HMGA2 to poly(dA-dT)$_2$ is an exothermic reaction yielding a large negative enthalpy of -22.4 (± 0.5) kcal mol^{-1}. The binding stoichiometry was determined to be 15 base pairs per HMGA2 molecule, in agreement with our previous published results (Cui et al., 2005).

3.5 Conclusion

Based on a unique property of HMGA2, the asymmetrical charge distribution along the primary structure, we have developed a rapid purification procedure to prepare HMGA2 on a large scale for biophysical studies. Using a combination of cation- and anion-exchange chromatography, we can purify 50 to 100 mg of HMGA2 in four days from 10 liters of cell culture, which is free of nucleases and nucleic acids. The purity of the protein was more than 98% as judged by SDS-PAGE gels. HMGA2 purified from this rapid procedure was successfully used in various biophysical studies such as NMR measurements, DSC studies, and ITC analyses.

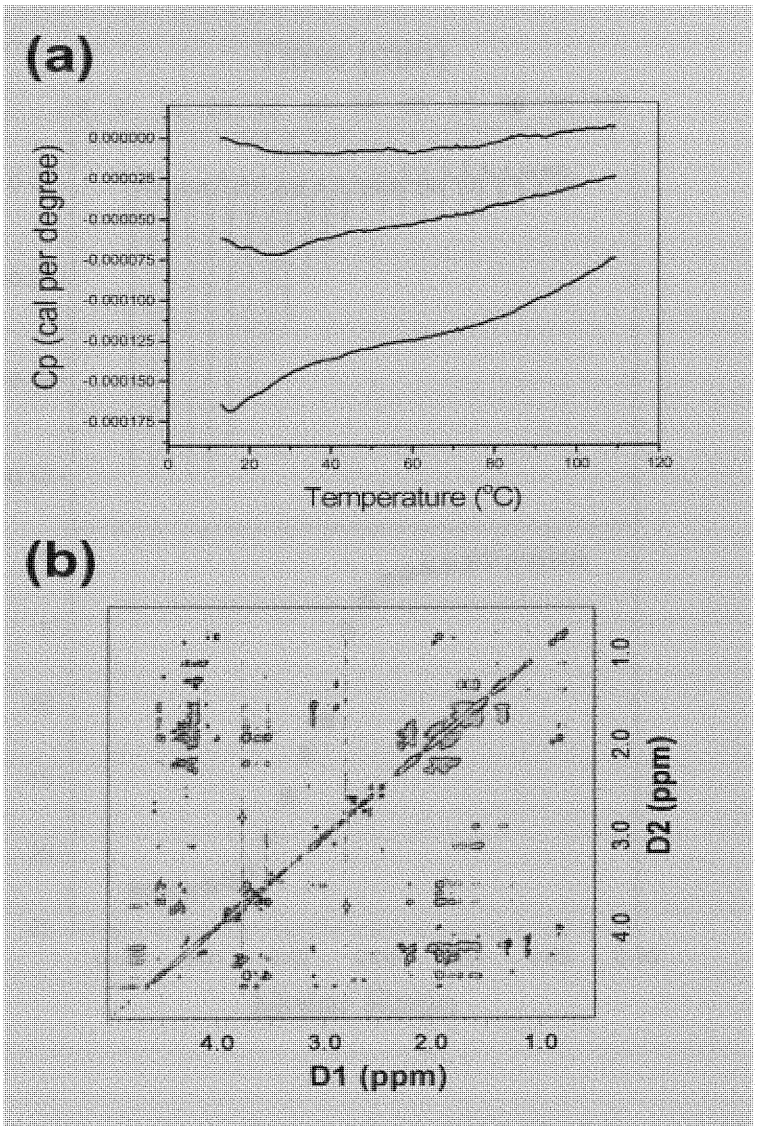


Figure 6. (a) Results of differential scanning calorimetry studies of HMGA2 in BPE plus 50 mM NaCl. Increasing concentrations of HMGA2 (37, 53, and 98 μM from top to bottom) are used and the buffer-buffer baseline was substracted from the thermograms. (b) NMR measurement: 600-MHz phase sensitive ^{1}H NOESY spectra of ~2 mM HMGA2 in D_2O were taken with a mixing time of 350 ms. The spectrum was collected at 293 K in 0.1 M phosphate buffer containing 5 mM mercaptoethanol at pH 7.0.

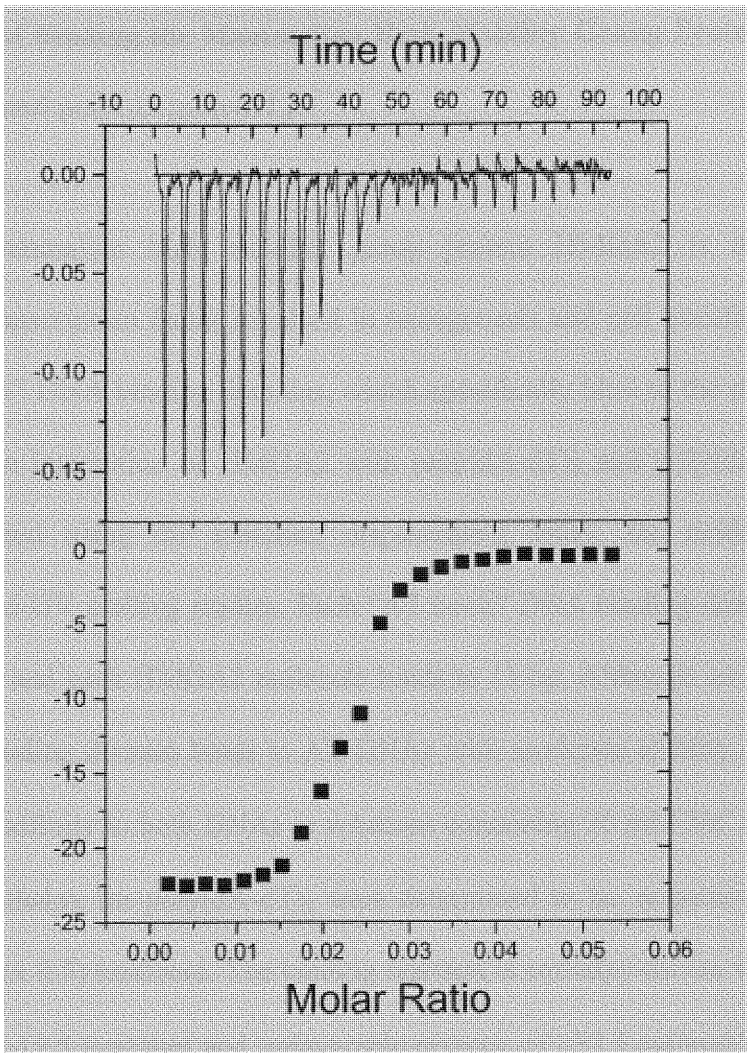


Figure 7. Sample raw data for the titration of HMGA2 into poly(dA-dT)$_2$ at 25 °C in BPE buffer containing 100 mM Na^+. Top, each peak shows the heat produced by injection of an aliquot of 15 µl of HMGA2 (50 µM) into DNA solution (1.7 ml of 100 µM (bp)). Bottom, binding isotherm resulting from integration with respect to time with appropriate dilution correction.

4. Energetics of Binding the Mammalian High Mobility Group Protein HMGA2 to poly(dA-dT)$_2$ and poly(dA)poly(dT)

4.1 Abstract

The mammalian high mobility group protein HMGA2 is a chromosomal architectural transcription factor involved in oncogenesis and cell transformation. It has three "AT-hook" DNA binding domains, which specifically bind to the minor groove of AT DNAs. The interaction of HMGA2 with poly(dA-dT)$_2$ and poly(dA)poly(dT) has been investigated using the ethidium displacement assay, the isothermal titration calorimetry, and UV melting studies. Each "AT hook" DNA binding domain was found to bind to 5 bp and each HMGA2 molecule binds to 15 bp. Although an individual "AT hook" DNA binding domain binds to AT DNAs with moderate affinity, HMGA2 binds with very high affinity to both DNAs in solutions containing 20 mM Na^+ at 25°C. The K_a and binding enthalpy for poly(dA-dT)$_2$ were determined to be, respectively, 1.9×10^{14} M^{-1} and –29.1 (± 0.5) kcal/mol. The binding reaction is enthalpy-driven with a favorable free energy of –19.5 kcal/mol and unfavorable entropy of –32.5 cal/mol/K ($-T\Delta S$ = +9.7 kcal/mol) at a 1 M reference state. Interestingly, although HMGA2 binds to poly(dA)poly(dT) with a binding constant of 9.6×10^{12} M^{-1}, the binding reaction is entropy-driven with an unfavorable enthalpy of +0.6 kcal/mol, a free energy of –17.7 kcal/mol and an entropy of +61.4 cal mol/K ($-T\Delta S$ = -18.3 kcal/mol) at the 1 M state. The enthalpy-entropy compensation is similar to that of several minor groove-binding drugs such as netropesin, distamycin A and hoechst33258 and may be a reflection of dehydration difference of different ligand-DNA complexes. The salt-dependence of the

binding constant of HMGA2 with both DNAs showed that electrostatic interaction is a dominant force for the binding reactions. The temperature dependence of binding enthalpy for poly(dA-dT)$_2$ indicates a large heat capacity of binding of –705 (± 113) cal/mol/K, consistent with an important role of solvent displacement in the linked folding/binding processes in this system.

4.2 Introduction

The mammalian high mobility group protein HMGA2 (old name: HMGI-C) is a non-histone chromosomal protein that is involved in oncogenesis and cell transformation (Goodwin, 1998; Bustin, 1999; Reeves, 2001, 2003; Sgarra et al., 2004). It was first identified as a nuclear protein in proliferating fibroblasts and embryos, and also in transformed thyroid cells infected by oncogenic viruses (Goodwin et al., 1985; Giancotti et al., 1985). Aberrant expression of HMGA2 gene and its mutants is directly linked to tumorigenesis of different benign tumors of mesenchymal origin such as lipomas, uterine leiomyomas, and fibroadenomas (Schoenmakers et al., 1995; Klotzbucher et al., 1999; Ligon & Morton, 2000; Fedele et al., 2001). In many cases, rearrangements of chromosomal bands 12q13-15 where HMGA2 gene is located cause disruptions of HMGA2's normal functions (Tallini & Dal, 1999; Aman, 1999). Several types of malignant tumors such as lung cancer, breast cancer, and leukemia were also found to be related to the expression of HMGA proteins including HMGA2 (Kazmierczak et al., 1996; Rogalla et al., 1998; Reeves & Beckerbauer, 2001; Kayser et al., 2003; Evans et al., 2004). The expression level is correlated with the degrees of malignancy and metastasis potential of the transformed cells (Abe et al., 1999; Tallini & Dal, 1999;

Evans et al., 2004). Therefore, HMGA proteins may serve as a biomarker for diagnosing the neoplastic transformation and the metastasis potential of many cancers (Giancotti et al., 1991; Rogalla et al., 1998; Reeves & Beckerbauer, 2003; Sgarra et al., 2004). HMGA2 is only expressed in proliferating, undifferentiated mesenchymal cells and is undetectable in normal fully differentiated adult cells (Zhou et al., 1995; Gattas et al., 1999). Disruption of its normal expression patterns causes deregulations of cell growth and differentiation. Zhou et al. showed that inactivation of Hmga2 gene results in the pygmy phenotype in transgenic Hmga2 (-/-) mice (Zhou et al., 1995). These mutant mice showed a severe deficiency of fat cells (almost 20-fold decrease) and a reduction of mesenchymal tissues. Microinjection of HMGA1a, another member of the HMGA family, into mouse embryos induces early activation of gene transcription (Beaujean et al., 2000) and causes gigantism and tumor formation (Battista et al., 1999; Baldassarre et al., 2001). It has been demonstrated that these *in vivo* effects are mediated by the binding of HMGA proteins to their target DNA sequences (Beaujean et al., 2000).

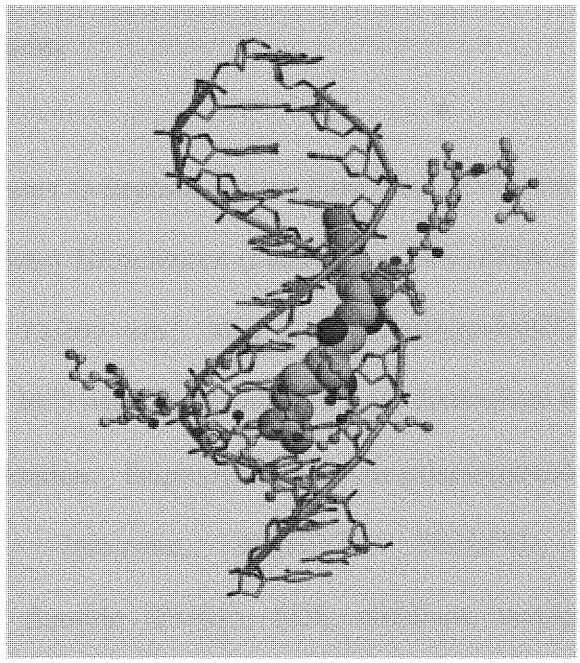
HMGA2 is a member of the HMGA family, which consists of three proteins, HMGA1a (old name: HMGI), HMGA1b (old name: HMGY), and HMGA2. HMGA1a and HMGA1b are splice variants of the same gene, HMGA1 gene (Friedmann et al., 1993). HMGA2 is the product of a separate gene, HMGA2 gene (Chau et al., 1995). The unique feature of HMGA proteins is that all have three "AT-hook" DNA binding domains, which contains a consensus palindromic sequence, PRGRP, flanked on each side by one or two positively charged amino acid residues (arginine or lysine). Circular dichroism (CD) and nuclear magnetic resonance (NMR) studies have shown that the

"AT-hook" DNA binding domain, in the absence of DNA, has no defined structure (Lehn et al., 1988; Huth et al., 1997). However, when it binds to the minor groove of AT base pairs, this domain adopts a defined conformation (Huth et al., 1997). The central core, RGR, deeply penetrates into the minor groove of AT base pairs with two arginine residues forming extensive electrostatic and hydrophobic contacts with the floor of the minor groove (Huth et al., 1997). This property is a signature of several small minor-groove binders, such as netropsin and Hoechst33258 (Zimmer & Wahnert, 1986; Neidle, 2001). All prefer binding to the minor groove of AT base pairs without causing a large disruption of the DNA conformation (Kopka et al., 1985; Pjura et al., 1987). In the NMR structure of the "AT-hook" peptide-DNA complex (Huth et al., 1997), two prolines on each side of the RGR core direct the protein away from the floor of the minor groove and position the positively charged arginine or lysine near the negatively charged phosphate backbone, thereby making further contacts (Figure 8(a)). This disordered to ordered conformational change significantly increases HMGA proteins' adaptability and allows them to participate in a variety of nuclear activities such as gene transcription, DNA replication, and chromatin remodeling (Goodwin, 1998; Reeves, 2003).

HMGA proteins are general transcriptional factors, proteins that regulate transcriptional activities by modulating DNA conformation and providing a framework for organizing functional transcriptional machinery (Wolffe, 1994). They prefer binding to appropriately spaced two to three runs of 5 AT base pairs (Maher & Nathans, 1996). However, for different promoter regions, they bind to different AT rich DNA sequences. For example, HMGA1a binds to a 27 A-track, i.e. 5'-

AAAAAAAAAAAAAAAAAAAAAAAAAA-3' within the insulin receptor gene promoter E3 region to modulate transcriptional activities (Foti et al., 2003). For other promoter regions such as human IFN β promoter (Whitley et al., 1994), human HIV-1 proviral promoter (Henderson et al., 2000) and human interleukin-15 promoter region (Baldassarre et al., 2001), HMGA proteins bind to the alternate AT base pairs. These AT rich sequences do not share a consensus sequence. Clearly, HMGA proteins recognize the minor groove of different AT base pairs. An understanding of the thermodynamics of the interactions between HMGA proteins and AT DNAs can provide insight into the physical basis of HMGA specificity. In this study, we have used a combination of isothermal titration calorimetry (ITC) and DNA UV melting experiments to characterize the interactions of HMGA2 with poly(dA-dT)$_2$ and poly(dA)poly(dT). Poly(dA-dT)$_2$ has alternate AT base pairs and poly(dA)poly(dT) contains A tracks only. ITC directly measures the size of the DNA binding site, the enthalpy, and the heat capacity change for binding. The UV melting curves obtained as a function of increasing ligand concentration can be analyzed by the rigorous statistical mechanical models of Crothers (Crothers, 1971) and McGhee (McGhee, 1976) to obtain the protein- DNA binding constant. Our results showed that HMGA2 binds very strongly to both poly(dA-dT)$_2$ and poly(dA)poly(dT). Interestingly, the binding reaction of HMGA2 to poly(dA-dT)$_2$ was enthalpy driven and the binding reaction of HMGA2 with poly(dA)poly(dT) was entropy-driven.

(a)



(b)

M S A R G E G A G Q P S T S A Q G Q
P A A P V P Q K R G R G R P R K Q Q
Q E P T C E P S P K R P R G R P K G S
K N K S P S K A A Q K K A E T I G E
K R P R G R P R K W P Q Q V V Q K K
P A Q E T E E T S S Q E S A E E D

(c) K R P R G R P R K W

Figure 8. (a) The NMR structure of the complex of a "AT hook" DNA binding domain with DNA determined by Huth et al. (b) The amino acid sequence of the mouse high mobility group protein HMGA2. (c) The amino acid sequence of a synthetic peptide containing the third "AT hook" DNA binding domain of HMGA2 (the ATHP).

4.3 Materials and Methods

4.3.1 Materials

Ion exchange resins Q-Sepharose Fast Flow and SP-Sepharose Fast Flow were purchased from Amersham Biosciences (Piscataway, NJ). DNA poly(dA-dT)$_2$ (molar extinction coefficient, 13,200 cm^{-1} M^{-1}) and poly(dA)poly(dT) (molar extinction coefficient, 12,000 cm^{-1} M^{-1}) were also purchased from Amersham Biosciences and used without further purification. Isopropyl-beta-D-thiogalactopyranoside (IPTG), DL-dithiothreitol (DTT), lysozyme, Na_2HPO_4, NaH_2PO_4, NaCl, glycerol, phenylmethanesulfonyl fluoride (PMSF), SDS, ethidium, netropsin, distamycin A, ampicillin, kanamycin, acrylamide, and N,N'-methylenebisacrylamide were from Sigma-Aldrich corporation (St. Louis, MO). All restriction enzymes were obtained from New England Biolabs (Beverly, MA). pfu Turbo DNA polymerase was purchased from Stratagene Corporation (La Jolla, CA). Plasmid pET30a was purchased from Novagen (Madison, WI). Synthetic deoxyoligonucleotides were purchased from MWG-Biotech, Inc. (High Point, NC). A peptide containing the third "AT hook" DNA-binding domain (KRPRGRPRKW), therefore designated the "AT hook" peptide (ATHP), was custom synthesized by advanced ChemTech, Inc. (Louisville, KY) with 95% purity and used without further purification.

4.3.2 Plasmid construction and protein purification

A plasmid, pET21c(+)-Hmga2 that contains a full length of murine Hmga2 cDNA and expresses a his-tag HMGA2, was kindly provided by Dr. Alfredo Fusco at Universita di Catanzaro, Italy. In this study, we used a PCR-based method to create a

plasmid pMGM1 to express wild type HMGA2 (without the C-terminal his-tag). The sequence of Hmga2 gene was confirmed by DNA sequencing. Since HMGA2 has a basic N-terminus and an acidic C-terminus, it should bind to both anion exchange column, such as Q-Sepharose Fast Flow column, and cation exchange column, such as SP-Sepharose Fast Flow column. Indeed, this unique property allowed us to adopt a simple procedure to purify large quantities of pure wtHMGA2. *E. coli* strain BLR (DE3)/pMGM1 was grown in Terrific Broth containing 50 μg/ml of the antibiotic kanamycin. The cell growth was monitored by measuring OD_{595} every hour using an Amersham Ultrospec 2000 UV-VIS spectrophotometer. When OD_{595} reached 0.6-0.7, HMGA2 expression was induced by the addition of 1 mM of IPTG to the cell culture. After a three-hour incubation or when the OD_{595} reached approximately 1.5, the cells were harvested by centrifugation at 4,000 rpm for 25 minutes at 4°C. The supernatant was discarded and the cell pellet was re-suspended in an ice-cold lysis buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 0.5 mM PMSF, 0.1 mM DTT, 1 mg/ml lysozyme) at 3.5 ml per gram of cells. The cells in the lysis buffer were incubated on ice for one hour and then the sample was frozen in liquid nitrogen and stored in a -80°C freezer. In the following day, the frozen cells were thawed and the salt concentration was adjusted to 1 M by adding solid NaCl. The solution was sonicated on ice for a total of 8 times at 300 W each time with a 10 second interval between each sonication. This was followed by centrifugation of the solution at 18,000 rpm for 20 minutes at 4°C. After centrifugation, the supernatant was dialyzed against buffer A (50 mM sodium phosphate, pH 8.0, 0.5 mM PMSF, 0.1 mM DTT, 10% glycerol) plus 200 mM NaCl at

4°C overnight. After dialysis, the solution was loaded onto a 40 ml SP-Sepharose Fast Flow cation exchange column pre-equilibrated with buffer A plus 200 mM NaCl. The loaded column was washed with 120 ml of buffer A plus 300 mM NaCl. Then the protein was eluted with a 300 ml NaCl gradient of 0.3 to 0.8 M NaCl in buffer A. HMGA2 was eluted at approximately 0.5 M NaCl. Peak fractions containing electrophoretically identified HMGA2 were pooled and dialyzed against buffer B (50 mM sodium phosphate pH 8.0, 10% glycerol, 0.1 mM DTT) plus 20 mM NaCl. The dialyzed protein was loaded onto a 40 ml Q-Sepharose Fast Flow anion exchange column pre-equilibrated with buffer B plus 20 mM NaCl. Then the protein was eluted with a 300 ml NaCl gradient 20 to 300 mM NaCl in buffer B. HMGA2 eluted at approximately 190 mM NaCl. The pooled Q-Sepharose fractions contained HMGA2 with purity greater than 98% judged from SDS-PAGE gels, and are free of nucleases and nucleic acids.

4.3.3 Ethidium displacement assay

Binding stoichiometries of HMGA2 or ATHP to DNA (bp) were determined by ethidium displacement assays. An aqueous solution containing 10 μM (bp) either poly(dA-dT)$_2$ or poly(dA)poly(dT) and 5 μM ethidium was prepared in BPE buffer (6 mM Na_2HPO_4, 2 mM NaH_2PO_4, 1 mM Na_2EDTA, pH 7.0) containing 4 mM NaCl. A sample (3 ml) of this solution in a fluorescence quartz cuvette, was titrated by adding 10 μl aliquots of 25 μM HMGA2 or the ATHP stock solution until the fluorescence intensity no longer changed. The fluorescence intensity was measured on a Jobin Yvon

FluoroMax-3 spectrofluorometer with λ_{ex} = 520 nm and λ_{em} = 600 nm. The percentage of the fluorescence quenched was plotted against the binding ratio of protein to DNA, which was used to determine the apparent DNA binding site size.

4.3.4 Isothermal titration calorimetry

Calorimetric experiments were carried out using a VP-ITC titration calorimeter (Microcal Inc., Northampton, MA, USA) interfaced to a PC. Origin 5.0 software, supplied by the manufacturer was used for data acquisition and analysis. Samples were extensively dialyzed against BPE buffer containing 4 mM NaCl. Typically, the titration was set up so that 15 μl of a 50 μM HMGA2 sample was injected every 200 seconds, up to a total of 18 injections, into a DNA sample (1.7 ml of 100 μM (bp)) 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 (usually the heat from the addition of buffer into DNA is very small and was ignored). The net enthalpy for each protein-DNA interaction was determined by subtraction of the component heats of dilution.

4.3.5 DNA UV melting studies

Ultraviolet DNA melting curves were determined using a Beckman DU-600 UV/VIS spectrophotometer equipped with a thermoelectric temperature controller. Poly(dA-dT)$_2$ and poly(dA)poly(dT) in BPE buffer plus 4 mM NaCl were used for melting studies. DNA samples (final concentration of 2.0 x 10^{-5} M bp) were prepared by direct mixing with aliquots from a HMGA2 stock solution, followed by incubation for several hours at 24°C to ensure equilibration. Samples were heated at a rate of 1°C min^{-1}, while continuously monitoring the absorbance at 260 nm. Primary data were transferred to the graphics program Origin (MicroCal, Inc., Northampton, MA) for plotting and analysis.

4.3.6 Determination of DNA binding constants

The DNA binding constant was determined with DNA UV melting studies. According to Crothers' (Crothers, 1971) and McGhee's theories (McGhee, 1976), in the presence of saturated binding ligands, DNA melting temperature is a function of the DNA binding constant, the binding site size, and the ligand concentration described by the following equation with an assumption of no interaction of the ligand with single-stranded DNA:

$$\left(\frac{1}{T_m^0} - \frac{1}{T_m}\right) = \frac{R}{\Delta H_m} \ln(1 + K_{T_m} L)^{\frac{1}{n}} \qquad (1)$$

where T_m^0 is the melting temperature of the DNA alone, T_m is the melting temperature in the presence of saturating amounts of ligand, ΔH_m is the enthalpy of DNA melting (per

mol bp), R is the gas constant, K_{Tm} is the ligand binding constant with double stranded DNA at T_m, L is the free ligand concentration (approximated at T_m by the total ligand concentration), and n is the ligand site size for double stranded DNA.

The DNA binding constants at lower temperatures was determined by the van't Hoff equation:

$$\ln(\frac{K}{K_{Tm}}) = -\frac{\Delta H_b}{R}(\frac{1}{T} - \frac{1}{T_m}) \qquad (2)$$

where K is the DNA binding constant at temperature T (Kelvin) and ΔH_b is the ligand's DNA binding enthalpy.

The DNA binding constant of HMGA2 and the ATHP was also determined by analysis of complete UV melting curves at less than saturating ligand concentrations, using McGhee's theory of DNA melting in the presence of ligands. A detailed description of the theory is given in the original paper (McGhee, 1976). Briefly, if HMGA2 and the ATHP are assumed not to bind to single-stranded DNA, complete melting curves at a given ligand concentration may be calculated by McGhee's algorithm from the parameters T_m°, ΔH_m, s, σ, ω_h, ΔH_b, K, and n. The parameters are defined as follows: T_m° is the melting temperature of the DNA in the absence of ligand; ΔH_m is the enthalpy for DNA melting in the absence of ligand; s is the equilibrium constant for forming a helix base pair from two coil nucleotides; σ is the nucleation parameter for forming a single stranded base pair within a stretch of helix; ω_h is the cooperative parameter for the ligand binding to helical base pairs; ΔH_b is the enthalpy for the ligand

binding to helical base pairs; K is the DNA binding constant; and n is the neighbor exclusion parameter, the number of DNA base pairs in the binding site. In order to generate melting curves, T_m°, ΔH_m, s, ω_h, and ΔH_b were independently determined and constrained. The parameters K, σ, and n were estimated by successive approximation. Each parameter was systematically adjusted to produce "best fit" curves that gave a minimum value of the sum-of-squares of residuals (SSR), the differences between observed and calculated values of melting curves, i.e., $SSR = \Sigma(y_{obs.} - y_{calc})^2$ (Leng et al., 1998). In this fitting process, one parameter (K, n, σ) was varied to construct different melting curves while all other parameters were constrained to a given value. The data derived from these calculated melting curves were used to generate SSR profiles. A FORTRAN program for calculating the DNA melting curves according to McGhee's theory was kindly provided by Dr. James McGhee and was edited and recompiled by Dr. Susan Wellman.

The DNA binding constants of HMGA2 and the ATHP at different salt concentrations were determined by DNA UV melting studies and were corrected to 20°C by the method described above. The data were plotted as logK against $\log[Na^+]$. The slope of this graph gives an estimate of SK of $d\log K/d\log[Na^+]$. This parameter can then be used to dissect the binding free energy into its polyelectrolyte and nonpolyelectrolyte components using polyelectrolyte theory (Leng et al., 1998).

4.4 Results

4.4.1 Determining the apparent DNA binding site size for the ATHP and HMGA2 by the ethidium displacement assay

Mouse HMGA2 is a 108 amino acid residue protein with a single tryptophan residue located at position 84 in the C-terminal side of the third "AT hook" DNA binding domain (Figure 8(b)). Therefore, it has a near UV absorption spectrum with maximum absorbance at 280 nm, which can be used to precisely determine the protein concentration (an extinction coefficient of 5,810 M-1 cm-1 at 280 nm was determined by Gill and von Hippel's method (Gill & von Hippel, 1989)). It also has a fluorescence emission spectrum with maximum near 360 nm, indicating that the tryptophan residue has a substantial contact with aqueous phase. The tryptophan fluorescence may be quenched upon forming HMGA2-DNA complex, which could be used to determine the apparent DNA binding site size and other thermodynamic parameters (Lohman & Overman, 1985). However, the fluorescence change is too complicated to derive useful data (unpublished results). Therefore, an ethidium displacement assay (Boger et al., 2001) was used to derive the apparent DNA binding site size for the ATHP and HMGA2 . As an initial step, the apparent DNA binding site size for several well-characterized DNA minor groove binders, such as distamycin A, netropsin and hoechst33258, binding to both poly(dA-dT)$_2$ and poly(dA)poly(dT) were examined. It was found that these minor groove binders bind to 4 to 5 AT base pairs. These results are consistent with the previously published observations (Boger et al., 2001).

Figure 9(a) shows the results of the ATHP titrating into a solution containing poly(dA)poly(dT)-ethidium complex. The inflection point between the pre- and post-saturation corresponds to a peptide to DNA mole ratio of 0.2. This is equivalent to a stoichiometry of 5 mol of bp per mole of the ATHP, the stoichiometry expected for binding of one "AT hook" DNA-binding domain to the minor groove of the AT base pairs (Huth et al., 1997). Similar results were also obtained for the ATHP binding to poly(dA-dT)$_2$; however, the titration curve is rather gradual and the inflection point is less obvious for the pre- and post-saturation (data not shown). This may result from the properties of poly(dA-dT)$_2$ or the experimental conditions used here. Nevertheless, the stoichiometry of 5 bp per peptide molecule was confirmed in our ITC experiments (Figure 10) and our UV melting studies (Table 1) for both poly(dA)poly(dT) and poly(dA-dT)$_2$.

Figure 9(b) and 9(c) show the results of HMGA2 titrating into solutions containing poly(dA-dT)$_2$-ethidium complex and poly(dA)poly(dT)-ethidium complex respectively. Both plots show an inflection point between the pre- and post-saturation corresponding to HMGA2 to DNA mole ratio of 0.067. These results suggest that HMGA2 binds to 15 base pairs of the AT DNAs. These binding stoichiometris are consistent with our results obtained from ITC and the UV melting studies (Table 1; see below for details).

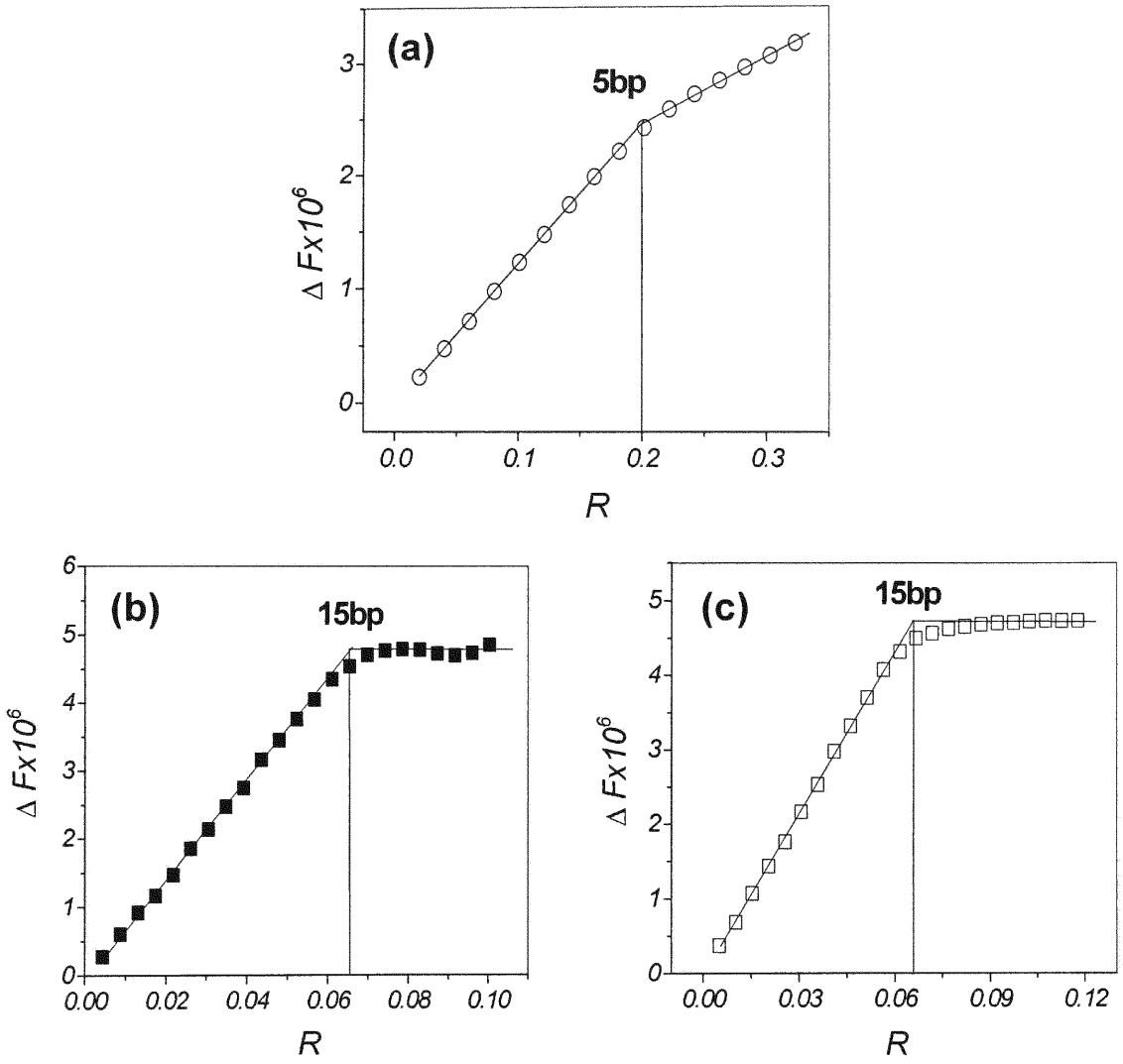


Figure 9. Ethidium displacement assays were performed according to conditions as described under "Materials and Methods". (a) The ATHP was titrated into a solution containing poly(dA)poly(dT)-EB complex. (b) HMGA2 was titrated into a solution containing poly(dA-dT)$_2$-EB complex. (c) HMGA2 was titrated into a solution containing poly(dA)poly(dT)-EB complex.

4.4.2 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) was used to investigate the DNA binding characteristics of the ATHP and HMGA2. With ITC, we directly measure the DNA binding enthalpy and the DNA binding stoichiometry. Figure 10(a) shows results from a typical ITC experiment in which the ATHP was titrated into the poly(dA-dT)$_2$ solution. There were two distinct binding processes in this titration experiment. The first results in a binding enthalpy of -6.8 (± 0.3) kcal mol^{-1} and an apparent binding site size of 5 bp per ligand molecule; the second has a binding enthalpy of -3.0 (± 0.1) kcal mol^{-1} and a binding stoichiometry of 2.5 bp per peptide. We observed precipitation during the second binding process, suggesting that high concentrations of the ATHP may cause DNA condensation. Figure 10(b) shows results from the titration experiment in which the ATHP was titrated into the poly(dA)poly(dT) solution. Interestingly, a positive DNA binding enthalpy (+3.1 (± 0.1) kcal mol^{-1}) was obtained with an apparent DNA binding size of 5 bp per peptide molecule. These differences in binding enthalpy for the two DNA molecules are discussed below. Nevertheless, the binding stoichiometry obtained from the ITC experiments is the same as that determined from the ethidium displacement assays, i.e. 5 bp per the ATHP. Since we used high concentrations of the ATHP (2.0 × 10^{-4} M) in the titration experiments, the binding is nearly stoichiometric and it is inappropriate to derive the DNA binding constant from such experiments. All other thermodynamic parameters will be obtained from UV melting studies.

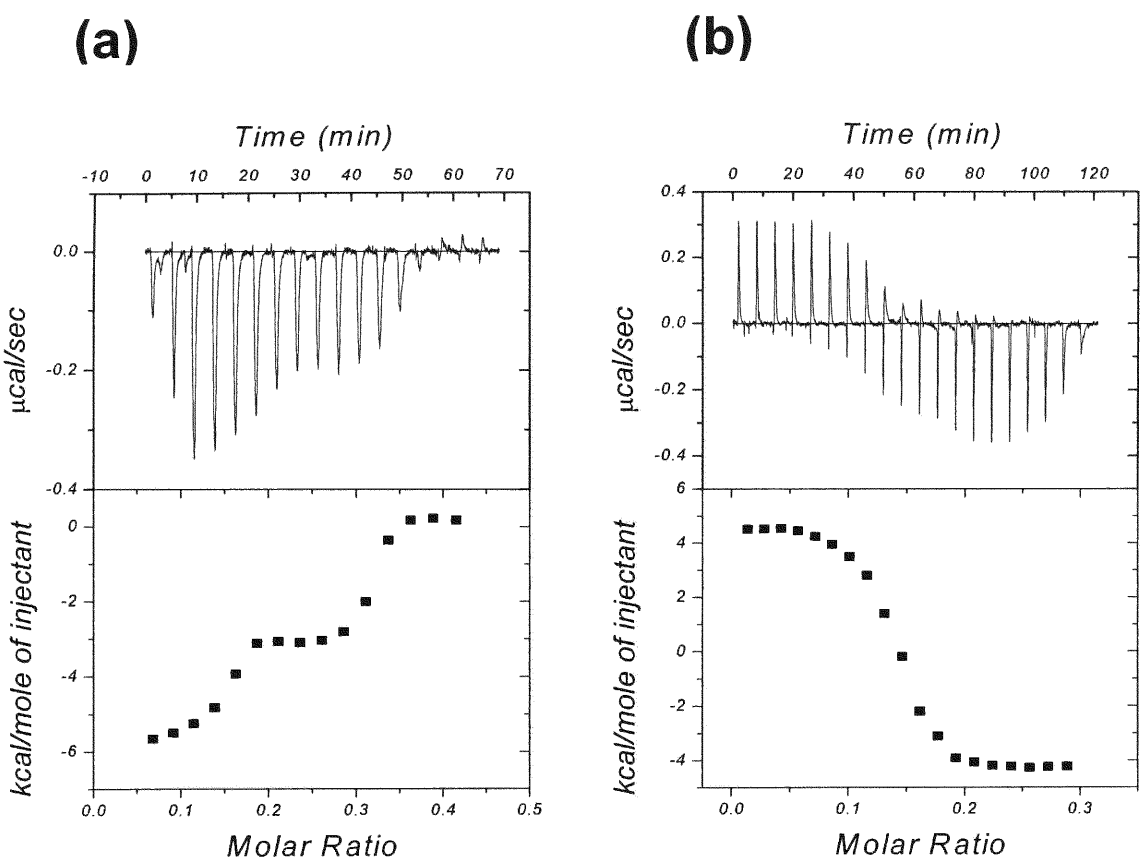


Figure 10. Sample raw data for the titration of the ATHP into (a) poly(dA-dT)$_2$ and (b) poly(dA)poly(dT) at 25°C in BPE buffer plus 4 mM NaCl (total 20 mM Na^+). Top, each peak shows the heat produced by injection of an aliquot of 13 µl of the ATHP (150 µM) into DNA solution (1.7 ml of 100 µM (bp)). Bottom, binding isotherm resulting from integration with respect to time with appropriate dilution correction.

Figure 11 shows results from the ITC experiment in which HMGA2 was titrated into a solution containing poly(dA-dT)$_2$ and 20 mM Na^+. Binding HMGA2 to poly(dA-dT)$_2$ is an exothermic reaction yielding a large negative enthalpy of -29.1 (± 0.5) kcal mol^{-1}. The binding stoichiometry was determined to be 15 base pairs per HMGA2

molecule, in agreement with the binding stoichiometry measured by the ethidium displacement assay (Figure 9). Since HMGA2 binds avidly to poly(dA-dT)$_2$ (see below for details) and the binding is nearly stoichiometric, we could not determine the DNA binding constant by ITC. HMGA2 was also used to titrate into a solution containing poly(dA)poly(dT), which produced a small positive enthalpy of +0.6 (± 0.1) kcal mol^{-1}. Because of the small magnitude of heat, we could not obtain a reliable titration curve and measure the binding stoichiometry.

4.4.3 Determination of the DNA binding constant of the ATHP and HMGA2 by UV melting studies

Since the DNA binding enthalpy and the DNA binding stoichiometry of the ATHP and HMGA2 have been determined, DNA UV melting studies were used to derive the K_a values for the interactions of the ATHP and HMGA2 with poly(dA-dT)$_2$ and poly(dA)poly(dT) in BPE buffer plus 4 mM NaCl (20 mM total Na^+). The T_m of poly(dA-dT)$_2$ and poly(dA)poly(dT) alone were 45.9°C and 52.9°C respectively. In the presence of 5 μM of the ATHP, a concentration sufficient to saturate the DNA lattice, the T_m values of poly(dA-dT)$_2$ and poly(dA)poly(dT) were elevated to 63.9°C and 76.9°C respectively. In the presence of 2 μM HMGA2, a concentration also sufficient to saturate the DNA lattice, the T_m of poly(dA-dT)$_2$ and poly(dA)poly(dT) were increased to 68.9°C and 81.9°C respectively. The DNA melting enthalpy of poly(dA-dT)$_2$ and poly(dA)poly(dT) have been previously determined by differential scanning calorimetry to be 8.4 and 8.9 kcal/mol respectively (Haq et al., 1997). Equation 1 can be used to

calculate the DNA binding constants at the melting temperature. For the ATHP, the DNA binding constant of poly(dA-dT)$_2$ and poly(dA)poly(dT) were 6.7 × 10^6 M^{-1} (at 63.9°C) and 2.2 × 10^7 M^{-1} (at 76.9°C) respectively, while for HMGA2, the DNA binding constant for poly(dA-dT)$_2$ and poly(dA)poly(dT) were 3.5 × 10^{11} M^{-1} (at 68.9°C) and 1.1 × 10^{13} M^{-1} (at 81.9°C) respectively. By application of the standard van't Hoff equation (equation 2), the binding constants for the ATHP & poly(dA-dT)$_2$, the ATHP & poly(dA)poly(dT), HMGA2 & poly(dA-dT)$_2$, and HMGA2 & poly(dA)poly(dT) were calculated to be 2.5 × 10^7, 1.0 × 10^7, 1.9 × 10^{14}, and 9.6 × 10^{12} M^{-1} respectively. Knowledge of the binding constant and binding enthalpy allowed us to construct the complete thermodynamic profiles for binding both the ATHP and HMGA2 to poly(dA-dT)$_2$ and poly(dA)poly(dT). The free energy is obtained from the standard relation ΔG° = -RTlnK and the entropy is computed with the equation $-T\Delta S = \Delta G - \Delta H$. The values of these parameters are summarized in Table 1.

The DNA binding constant for the ATHP binding to poly(dA-dT)$_2$ and poly(dA)poly(dT) were also determined by fitting experimental UV melting curves to McGhee's model of the helix-coil transition of DNA in the presence of ligands (McGhee, 1976).It was assumed that the ATHP does not bind to single-stranded DNA. For the ATHP binding to poly(dA-dT)$_2$, the following parameters were used and constrained in the fitting procedure: ΔH_m = 8.4 kcal mol^{-1} of bp, ΔH_b = -6.8 kcal mol^{-1}, [DNA] = 2.0 × 10^{-5} M bp, and ω_h = 1.0. The other four parameters (n, K, σ, and the effective [ATHP]) were adjusted to produce best-fit curves using methods described in

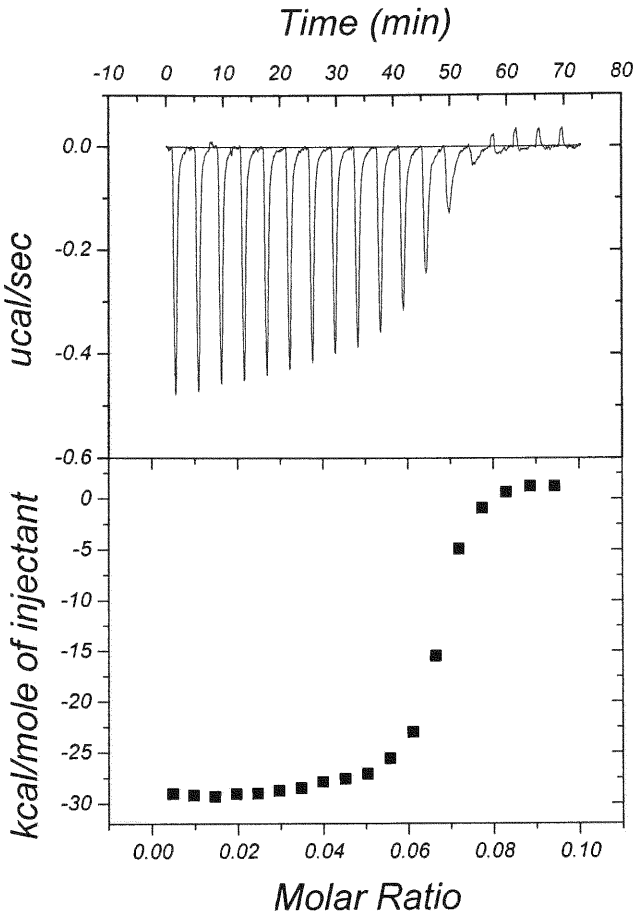


Figure 11. Sample raw data for the titration of HMGA2 into poly(dA-dT)$_2$ at 25°C in BPE buffer plus 4 mM NaCl (total 20 mM Na^+). Top, each peak shows the heat produced by injection of an aliquot of 15 µl of HMGA2 (50 µM) into DNA solution (1.7 ml of 100 µM (bp)). Bottom, binding isotherm resulting from integration with respect to time with appropriate dilution correction.

Table 1. Comparison of Thermodynamic Parameters for the ATHP and HMGA2 Binding to Poly(dA-dT)$_2$ and Poly(dA)poly(dT)[a]

Ligand	DNA	n	K_{obs}	ΔG_{obs}	SK	ΔG_{pe}	ΔG_t	ΔH	$T\Delta S$
ATHP[b]	poly(dA-dT)$_2$	5.0	2.5×10^7	-10.1	2.66	-6.2	-3.9	-6.8 ± 0.3	+3.3
ATHP[b]	poly(dA)poly(dT)	5.0	1.0×10^7	-9.6	2.54	-5.9	-3.7	+3.1 ± 0.1	+12.7
HMGA2	poly(dA-dT)$_2$	15.0	1.9×10^{14}	-19.5	5.91	-13.7	-5.8	-29.1 ± 0.5	-9.7
HMGA2	poly(dA)poly(dT)	15.0	9.6×10^{12}	-17.7	5.61	-13.0	-4.7	+0.6 ± 0.1	+18.3

[a]K_{obs} (M^{-1}) is the binding constant for the interaction of a ligand molecule with DNA and refers to solutions containing 0.020 M NaCl at 25°C. ΔG_{obs} (kcal mol^{-1}) is the binding free energy calculated from the equation $\Delta G_{obs} = -RTlnK_{obs}$. The parameter SK is the slope of the graphs shown in Figure 14. ΔG_{pe} and ΔG_t (kcal mol^{-1}) are the polyelectrolyte and the nonpolyelectrolyte contributions to the binding free energy, respectively. The polyelectrolyte contribution was calculated from equation $\Delta G_{pe} = SKRTln[Na^+]$, evaluated at $[Na^+] = 0.02$ M. The nonpolyelectrolyte potion of the free energy was calculated by subtraction. The apparent DNA binding site size n (bp) was determined by ethidium displacement assay or isothermal titration calorimetry experiments. The DNA binding enthalpy ΔH was determined by isothermal titration calorimetry experiments as well. $T\Delta S$ was calculated by subtraction, $-T\Delta S = \Delta G - \Delta H$.

[b]ATHP refers to the "AT hook" peptide.

Materials and Methods and also in our previous publications (Leng et al., 1998). The experimental and the best fit melting curves are compared in Figure 12(a), and the resulting optimized parameters are summarized in Table 2. The analysis of five melting curves at varying the ATHP concentrations yielded an estimate for K_a = 1.7 (± 0.2) × 10^7 M^{-1} and n = 5.0 (± 0.1) bp at 25°C. The DNA binding constant is in excellent agreement with that based on T_m measurements at saturating concentrations of the ATHP described above; the DNA binding site size equals the values determined by ethidium displacement assays and ITC experiments. Melting curves of poly(dA)poly(dT) in the presence of the ATHP were also fitted to McGhee's Model by constraining the following parameters in the fitting process: ΔH_m = 8.9 kcal mol^{-1} (bp), ΔH_b = +3.1 kcal mol^{-1}, [DNA] = 2.0 × 10^{-5} M bp, and ω_h = 1.0 (Figure 12(b) and Table 3). The resulting DNA binding constant (9.6 (± 0.4) × 10^6 M^{-1}) and binding site size (5.1 (± 0.2) bp per ligand) are also in excellent agreement with the values determined as described above.

We also attempted to fit the melting curves of poly(dA-dT)$_2$ and poly(dA)poly(dT) in the presence of HMGA2 to McGhee's model. The experimental and the best fitting melting curves are compared in Figures S1 and S2 (Appendix 1), and the resulting optimized parameters are summarized in Tables S1 and S2 (Appendix 1). The melting curves in the presence of unsaturating HMGA2 concentrations did not fit well to the experimental curves. Nevertheless, the DNA binding constants and the binding site size are in good agreement with the values determined as described above.

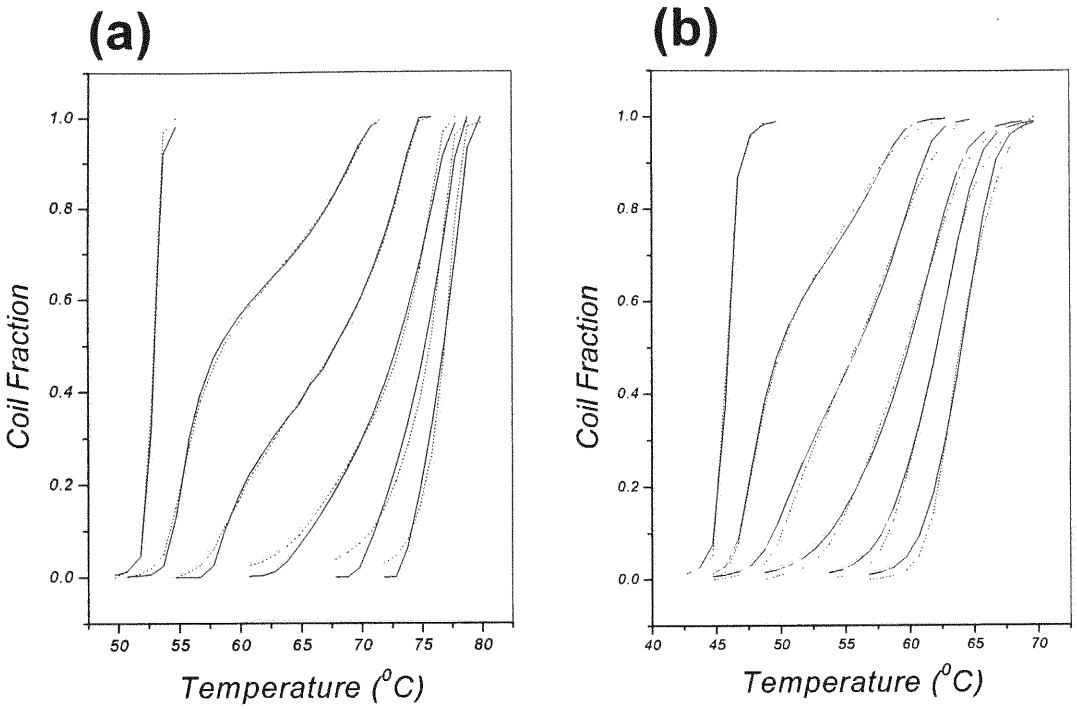


Figure 12. Fits of (a) poly(dA-dT)$_2$ and (b) poly(dA)poly(dT) DNA melting curves in the presence of various amounts of the ATHP to McGhee's theory. The dotted lines are experimental data, and the solid lines are theoretical curves with the parameters listed in Tables 2 and 3.

4.4.4 Salt dependence of the binding constants

The salt dependence of the ATHP and HMGA2 binding to poly(dA-dT)$_2$ and poly(dA)poly(dT) was investigated by UV melting experiments. Differences between the DNA melting temperatures of DNA alone and in the presence of saturating amounts of the ATHP or HMGA2 at different NaCl concentrations were used to calculate DNA binding constants of the ATHP and HMGA2 at the melting temperature. The DNA binding enthalpy of the ATHP and HMGA2 binding to poly(dA-dT)$_2$ was independently

Table 2. Binding parameters from melting of poly(dA-dT)$_2$ in the presence of the ATHP[a]

[DNA] (M bp)	[Peptide] (M)	K (M^{-1})	n (bp)	$[Peptide]_{sim}$ (M^{-1})	σ
2.0×10^{-5}	0	0	0	0	1.2×10^{-4}
2.0×10^{-5}	1.0×10^{-6}	1.5×10^{7}	4.8	1.0×10^{-6}	1.1×10^{-4}
2.0×10^{-5}	2.0×10^{-6}	1.5×10^{7}	4.9	1.9×10^{-6}	2.3×10^{-4}
2.0×10^{-5}	3.0×10^{-6}	1.7×10^{7}	5.0	2.9×10^{-6}	1.0×10^{-3}
2.0×10^{-5}	4.0×10^{-6}	1.8×10^{7}	5.0	3.9×10^{-6}	8.4×10^{-4}
2.0×10^{-5}	5.0×10^{-6}	2.1×10^{7}	5.0	4.9×10^{-6}	6.2×10^{-4}

[a][DNA] is the concentration of poly(dA-dT)$_2$. [Peptide] is the concentration of the ATHP used in the experiments. K is the DNA binding constant at 25°C. n is the DNA binding site size. $[Peptide]_{sim}$ is the concentration of the ATHP used in the simulation. σ is the nucleation parameter used in the simulation. The following parameters are constrained and used in the simulation: ΔH_m = 8400 cal mol^{-1} of bp, ΔH_b = -6.8 kcal mol^{-1}, and ω_h = 1.0.

Table 3. Binding parameters from melting of poly(dA)poly(dT) in the presence of the ATHP[a]

[DNA] (M bp)	[Peptide] (M)	K (M^{-1})	n (bp)	$[Peptide]_{sim}$ (M^{-1})	σ
2.0×10^{-5}	0	0	0	0	5.7×10^{-5}
2.0×10^{-5}	1.0×10^{-6}	9.3×10^{6}	5.4	1.1×10^{-6}	3.5×10^{-5}
2.0×10^{-5}	2.0×10^{-6}	9.8×10^{6}	5.1	2.1×10^{-6}	5.8×10^{-7}
2.0×10^{-5}	3.0×10^{-6}	1.0×10^{7}	5.0	3.1×10^{-6}	2.6×10^{-5}
2.0×10^{-5}	4.0×10^{-6}	9.8×10^{6}	5.0	4.0×10^{-6}	4.0×10^{-7}
2.0×10^{-5}	5.0×10^{-6}	9.0×10^{6}	5.0	5.1×10^{-6}	4.0×10^{-7}

[a][DNA] is the concentration of poly(dA)poly(dT). [Peptide] is the concentration of the ATHP used in the experiments. K is the DNA binding constant at 25°C. n is the DNA binding site size. $[Peptide]_{sim}$ is the concentration of the ATHP used in the simulation. σ is the nucleation parameter used in the simulation. The following parameters are constrained and used in the simulation: ΔH_m = 8,900 cal mol^{-1} (bp), ΔH_b = +3.1 kcal mol^{-1}, and ω_h = 1.0

determined by ITC. Both binding reactions are exothermic over a wide range of NaCl concentrations (Figure 13) and the DNA binding enthalpy decreases greatly with increasing NaCl concentration. Since the magnitude of the heat produced by the binding of the ATHP and HMGA2 to poly(dA)poly(dT) is so small, the DNA binding enthalpy at different NaCl concentrations could not be determined. It was assumed that the binding enthalpy of the ATHP and HMGA2 binding to poly(dA)poly(dT) is independent of salt concentration.

With the knowledge of the DNA binding enthalpy, the K_a values for DNA binding by the ATHP and HMGA2 to both poly(dA-dT)$_2$ and poly(dA)poly(dT) were corrected to 25°C by applying the standard van't Hoff equation (equation 2). Figure 14 shows the dependence of K_a at 25°C on the concentration of Na^+ as determined by UV melting studies. The K_a for both the ATHP and HMGA2 decrease with increasing salt concentrations, due to counterion release that accompanies the binding of the charged ligands to DNA. Based on the polyelectrolyte theory of Record et al. (Record et al., 1978), the slopes of the lines in Figure 14 are equal to

$$SK = \frac{\partial \log K}{\partial \log[Na^+]} = -Z\psi$$

where Z is the charge on the ligands and ψ is the fraction of counterions associated with each DNA phosphate (ψ= 0.88 for the double-stranded B-form of DNA). The quantity SK is equivalent to the number of the counterions released upon binding of a ligand with net charge Z. For the ATHP binding to poly(dA-dT)$_2$ and poly(dA)poly(dT), it was found that 2.66 and 2.54 counterions, respectively, are liberated from the DNA. For

HMGA2 binding to poly(dA-dT)$_2$ and poly(dA)poly(dT), 5.91 and 5.61 counterions, respectively, are released from the DNA. From these values of Zψ, the charges on each ligand are as follows: +3.01 for the ATHP binding to poly(dA-dT)$_2$, +2.89 for the ATHP binding to poly(dA)poly(dT), +6.72 for HMGA2 binding to poly(dA-dT)$_2$, and +6.38 for HMGA2 binding to poly(dA)poly(dT).

4.4.5 Determination of the heat capacity change (ΔC_p) for HMGA2 binding to poly(dA-dT)$_2$

ΔC_p for the interaction is $\delta\Delta H/\delta T$, the temperature dependence of the enthalpy of binding. ΔH values for the interaction between HMGA2 and poly(dA-dT)$_2$ were determined at different temperature (from 20°C to 30°C) by ITC. Figure 15 shows that the slope of a linear least-squares fit gives a value for ΔCp of -705 (±131) cal mol^{-1} K^{-1}. The free energy of DNA binding ($\Delta G = -RTlnK$) was calculated from UV melting studies. It should be noted that the ΔG of binding is concentration dependent. By convention it is normally calculated at a standard 1 M reference state. However, as discussed previously, ΔG values under the conditions used in our experiments are less negative than at the 1 M state. For example, ITC studies were conducted at concentrations in the 10 μM range where ΔG of binding will be 6.82 kcal mol^{-1} less negative than the values given in Table 1. For the ATHP binding to poly(dA-dT)$_2$, under these conditions, $\Delta G_{1,10\mu M}$ is -3.3 kcal mol^{-1} and $T\Delta S$, -3.5 kcal mol^{-1}. Interestingly, the free energy change is less sensitive to temperature than ΔH. The values for the binding enthalpy and entropy seem to be linearly correlated; the large heat capacity change is

compensated by changes in the entropy to produce relatively small changes in the free energy.

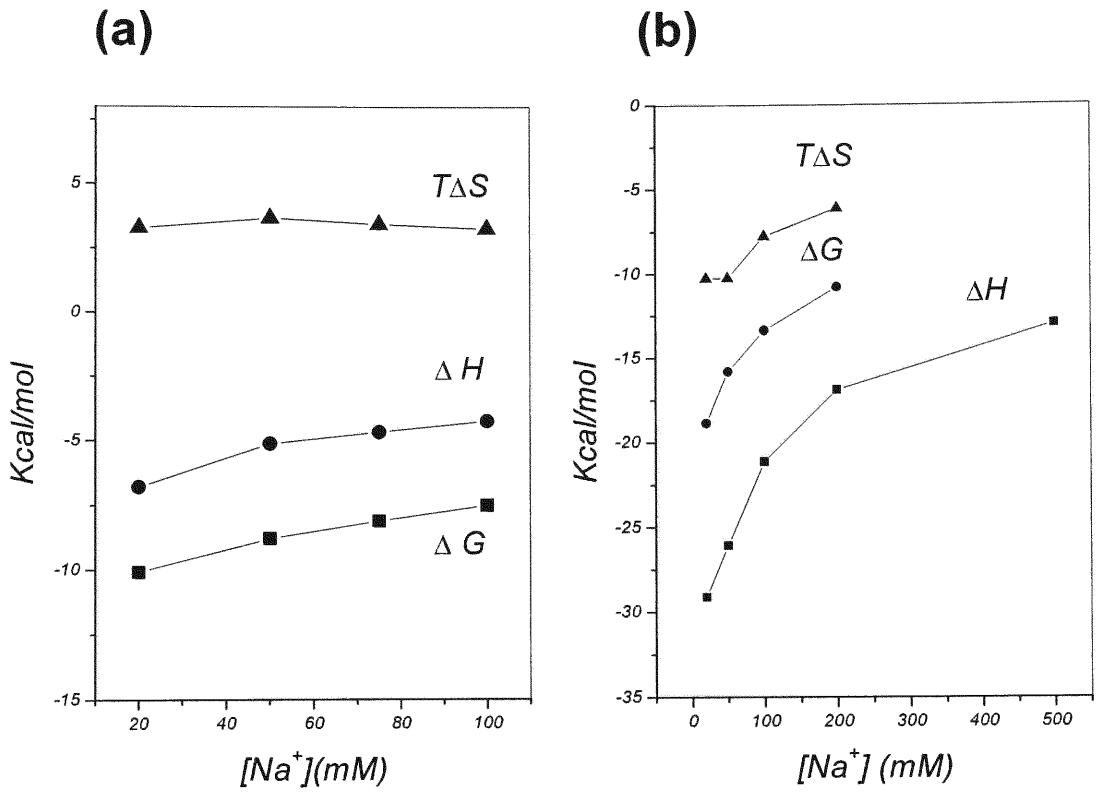


Figure 13. Ionic strength dependence of the Gibbs free energy (ΔG), enthalpy (ΔH), and entropy (TΔS) for the binding reactions between (a) the ATHP and poly(dA-dT)$_2$ or (b) HMGA2 and poly(dA-dT)$_2$. The Gibbs free energy (ΔG) was calculated according to equation ΔG = -RTlnK where K is the DNA binding constant determined by UV melting studies. ΔH was measured directly by ITC experiments. ΔS was computed by equation -TΔS = ΔG – ΔH. The standard deviation of ΔH for the ATHP binding to poly(dA-dT)$_2$ was estimated to be 0.2 (20 mM Na^+), 0.1 (50 mM Na^+), 0.1 (75 mM Na^+), and 0.1 (100 mM Na^+) kcal mol^{-1} respectively; the standard deviation of ΔH for HMGA2 binding to poly(dA-dT)$_2$ was estimated to be 0.2 (20 mM Na^+), 0.2 (50 mM Na^+), 0.1 (100 mM Na^+), 0.1 (200 mM Na^+), and 0.1 (500 mM Na^+) kcal mol^{-1} respectively.

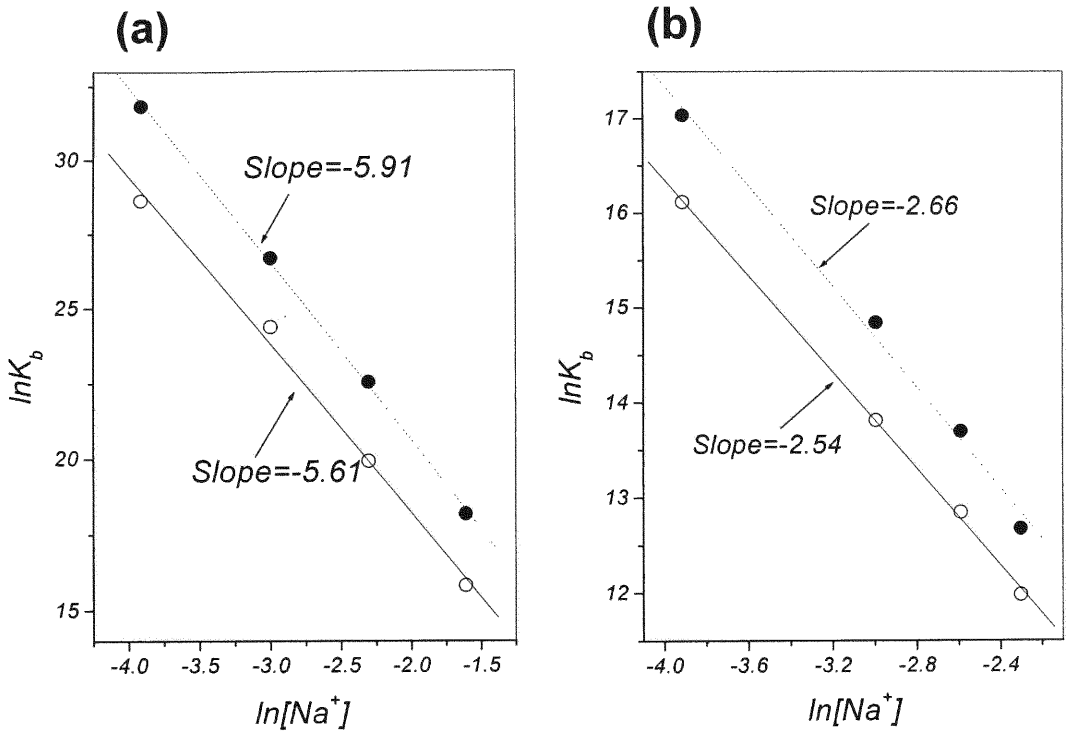


Figure 14. Dependence of DNA binding constants at 25°C on the salt concentration. (a) HMGA2 binding to poly(dA-dT)$_2$ (solid circles) and poly(dA)poly(dT) (open circles) in the presence of various NaCl concentrations. (b) The ATHP binding to poly(dA-dT)$_2$ (solid circles) and poly(dA)poly(dT) (open circles) in the presence of various NaCl concentrations. The linear least-squares fits to the data are shown.

4.5 Discussion

The results presented in this article provide a detailed thermodynamic profile (ΔG, ΔH, ΔS, n, and ΔC_p) for the interactions of two minor groove binding ligands, the ATHP and the mammalian high mobility group protein HMGA2, with DNA. The ATHP

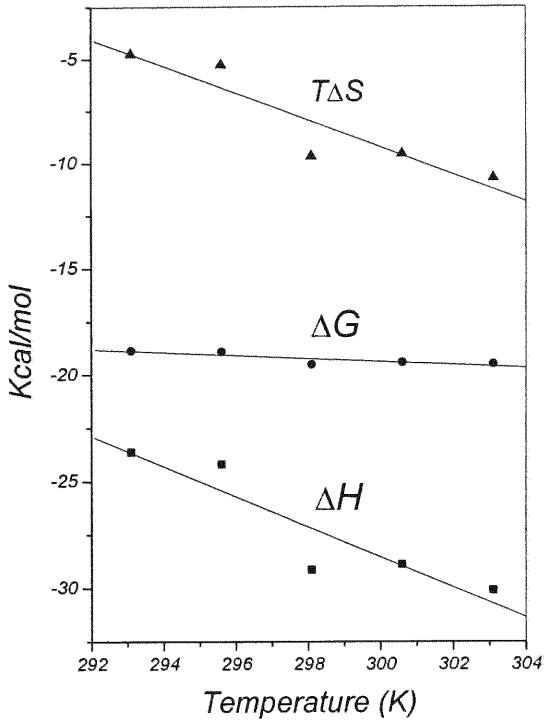


Figure 15. Heat capacity change (ΔCp) for binding HMGA2 to poly(dA-dT)$_2$. Linear least-squares fitting of the enthalpy data (squares) determined by ITC gives a ΔCp value of -705 cal mol^{-1} K^{-1}. The dependence of ΔG (circles) and TΔS (triangles) on temperature are shown. The standard deviation of ΔH at different temperature was estimated to be 0.1 (20°C), 0.1 (22.5°C), 0.2 (25°C), 0.1 (27.5°C), and 0.1 (30°C) kcal mol^{-1} respectively.

is a synthetic peptide containing one "AT hook" DNA binding domain whereas HMGA2 has three "AT hook" DNA binding domains. These results provide an opportunity to compare the energetics of multidentate binding by HMGA2 and its single domain counterpart, the ATHP, to two different AT DNAs: poly(dA-dT)$_2$ and poly(dA)poly(dT).

4.5.1 Binding stoichiometry

Binding stoichiometry is a key parameter for describing and understanding a ligand-DNA interaction. The NMR structure (Huth et al., 1997) indicates that one "AT hook" DNA binding domain occupies 5 AT base pairs. This is clearly confirmed by our ethidium displacement assays (Figure 9(a)) and ITC experiments (Figure 10) with the ATHP (Figure 16(a)). Also, our results showed that HMGA2 binds to 15 AT base pairs. Since HMGA2 contains three "AT hook" DNA binding domains, it is possible that all three "AT hook" domains simultaneously bind to the minor groove of the AT DNAs (Figure 16(c)) so that the whole protein will wrap around 1.5 turns of the DNA double helix. Alternatively, two of the three "AT hook" DNA binding domains could bind to the minor groove of the AT base pairs with each DNA binding domain binding to 5 base pairs with a 5 base pair space in the middle position (Figure 16(b)). In such a situation, HMGA2 binds to DNA only on one side. This binding mode is consistent with previous studies by Schwanbeck et al (Schwanbeck et al., 2000) who demonstrated that only the first and second DNA binding domains bind to the minor groove of a fragment of the beta-interferon promoter. However, all three AT hooks are involved in binding to long AT stretches of DNA and this is likely to be the situation in our experiments because the ATHP corresponds to the third DNA binding domain of HMGA2. Therefore, our results show that a stretch of five AT base pairs is the basic binding unit for the "AT hook" DNA binding domain and HMGA2 uses multiple "AT hook" DNA binding domains (probably three) to bind to the target DNA sequences, which is consistent with previous results (Solomon et al., 1986; Maher & Nathans, 1996).

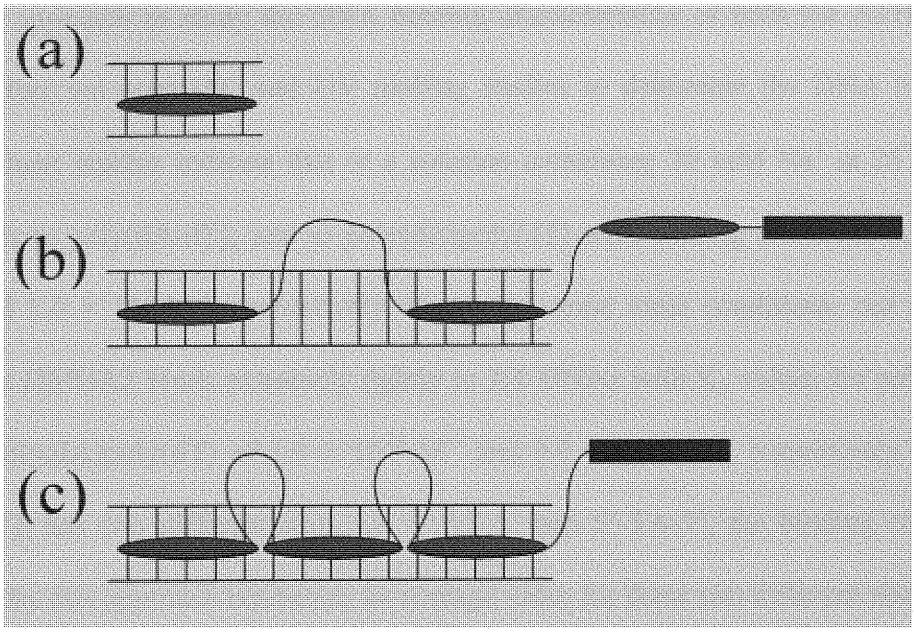


Figure 16. Possible binding modes for the interaction of HMGA2 with DNA. Symbols: red oval, one "AT hook" DNA binding domain; blue rectangle, the negative charged C-terminal motif. (a) One "AT hook" DNA binding domain binds to 5 AT base pairs. (b) HMGA2 binds to 15 AT base pairs using only two "AT hook" DNA binding domains. (c) HMGA2 binds to 15 AT base pairs using all three "AT hook" DNA binding domains.

4.5.2 Binding affinity

Since the ATHP does not have a defined structure when free in aqueous solution (Huth et al., 1997), it is reasonable to assume that increasing temperature would not significantly affect its conformation. The far and near UV CD spectrum of HMGA2 suggests that it has very little secondary structure (data not shown; a temperature dependent CD study showed little change of HMGA2's CD spectrum upon heating) and we assume in our analysis that it is also unfolded in the absence of DNA in the temperature range of our experiments. Therefore, DNA UV melting studies should be an

appropriate approach to study the DNA binding characteristics of both the ATHP and HMGA2. Here we used ITC and UV melting studies to determine the DNA binding constants; these studies indicate that in a solution containing 20 mM Na^+ at 25°C, the ATHP binds to poly(dA-dT)$_2$ and to poly(dA)poly(dT) with binding constants of 2.5 × 10^7 M^{-1} and 1.0 × 10^7 M^{-1}, respectively. Under the same ionic conditions, our results showed that HMGA2 binds extremely tightly to AT DNAs with K_a values for poly(dA-dT)$_2$ and poly(dA)poly(dT) of 1.9 × 10^{14} M^{-1} and 9.6 × 10^{12} M^{-1}, respectively. Since poly(dA-dT)$_2$ has two different sites, i.e. 5'-AT-3' and 5'-TA-3', the binding constant determined by this approach represents an average value for these two different sites.

Melting curves of poly(dA-dT)$_2$ and poly(dA)poly(dT) in the presence of increasing concentrations of the ATHP were fitted to McGhee's model of the helix to coil transition in the presence of ligands (McGhee, 1976). This method provides an independent verification of the binding site size and binding constant, along with a more reliable estimate of the uncertainty in the parameters. Figure 12 shows that the agreement between the calculated curves and the experimental results are excellent. In particular, the shapes and biphasic of the experimental melting curves are matched very well by the calculated curves. Our best fit produces the DNA binding site size of 5 base pairs and binding constants of 1.7 × 10^7 M^{-1} for poly(dA-dT)$_2$ and 9.6 × 10^6 M^{-1} for poly(dA)poly(dT). The DNA binding site size is exactly the same one determined independently by ITC experiments (Figure 10) and also by ethidium displacement assays (Figure 9(a)); it also matches the value determined by NMR studies in which an "AT hook" DNA binding domain occupies 5 AT base pairs (Huth et al., 1997). The DNA

binding constant is in good agreement with the one calculated according to equations 1 and 2 at saturating concentrations of the ATHP.

The same method was also used to fit DNA melting curves in the presence of HMGA2 to McGhee's model of the helix to coil transition in the presence of ligands. The best fits are shown in Figures S1 and S2 (Appendix 1). The melting curves in the presence of sub-saturating concentrations of HMGA2 do not match exactly with the computed curves. This deviation may partially reflect the conformational transition of HMGA2 upon DNA denaturation. Alternatively, since HMGA2 uses multiple "AT hook" DNA binding domains to interact with a total of 15 AT base pairs, individual "AT hook" DNA binding domains may bind to sections of 5 AT base pairs with different binding affinities. On heating, the different DNA segments that associate with the different "AT hook" DNA binding domains, may not melt in a single cooperative transition but at different temperatures, thus causing the deviation (McGhee's model neglects such a situation). Despite this uncertainty in interpretation, this analysis yields a DNA binding site size of 15 base pairs for HMGA2 binding to AT DNAs, in agreement with those determined independently by ITC (Figure 11) and ethidium displacement (Figures 9(b) and 9(c)).

The DNA binding constants determined in this study are in excellent agreement with previously published results. Using electrophoretic mobility shift assay (EMSA) Maher and Nathans showed that HMGA proteins bind tightly to a DNA binding site containing two to three 5 bp AT tracks with appropriate spacing (Maher & Nathans,

1996). Their data indicate that the DNA binding constant for HMGA2 binding to a high affinity binding site is 2×10^{10} M^{-1} in a solution containing 50 mM K^+ at 22°C, a value comparable to that determined in the present study. Furthermore, Piekielko et al demonstrated that HMGA2 binds tightly to a 34 bp fragment of the promoter of the IFNβ gene with a DNA binding constant of 1.7×10^7 M^{-1} in a solution containing 180 mM K^+ and 1 mM Mg^{2+} at 20°C (Piekielko et al., 2001). This value is also in the same order of magnitude as our DNA binding constant under similar ionic conditions (Figure 14(a)). Reeves and Nissen reported that two other members of the HMGA family, HMGA1a and 1b bind to an AT rich 3'-UTR of the bovine interleukin-2 cDNA with a binding constant of 9.7×10^8 M^{-1} and 7.2×10^8 M^{-1} respectively in a solution containing 52 mM Na^+ at 25°C (Reeves & Nissen, 1990). They used a 300 bp DNA fragment containing multiple binding sites for HMGA proteins. It is not surprising that their values are slightly lower than those determined in this paper and they may represent an average value for different binding sites. Recently, Dragan et al (Dragan et al., 2003) showed that in a solution containing about 120 mM K^+ at 20°C, one "AT hook" DNA binding domain binds to AT base pairs with a binding constant of 1.4×10^4 M^{-1}. This value is about 5 times lower than the value we obtained here (Figure 14(b)), possibly reflecting the fact that these authors used a short, 12 bp, oligomer for their studies. Short oligomers usually have different thermodynamic properties from their polymeric counterparts, especially with respect to end effects which may significantly weaken the ligand-oligomer interactions (Ballin et al., 2004).

4.5.3 Salt dependence of binding

The ATHP and HMGA2 are both cations, so their binding to DNA is thermodynamically linked to Na^+ binding to DNA, and their DNA binding constants (K_a values) depend on the total Na^+ concentration. Polyelectrolyte theories based on Manning's counterion condensation model provide a basis for interpreting our data (Record et al., 1978; Manning et al., 1978). Our calculations indicate that the interaction of the ATHP (one DNA binding domain) uses only three out of six positive charges at neutral pH binding to AT DNAs. The NMR structure (Huth et al., 1997) (Figure 8), suggests that two positive charges from the central RGR motif, which deeply penetrates into minor groove, may fully participate the binding process while the other four positive charges partially contribute to the DNA binding. With HMGA2, which has 25 positive charges at neutral pH, our results (Figure 14) indicate that only 6 to 7 charges contribute to the DNA binding process. It is likely that all these charges are from the "AT hook" DNA binding domains. Our results are in excellent agreement with previous published studies of the ATHP binding to synthetic oligonucleotides (Dragan et al., 2003).

From the dependence of the binding constant on salt concentration, the observed binding free energy (ΔG^0_{obs}) can be partitioned into two contributions:

$$\Delta G^0_{obs} = -RT \ln K = \Delta G_t + \Delta G_{pe}$$

where ΔG_t is the nonpolyelectrolyte contribution to the binding free energy and ΔG_{pe} is the polyelectrolyte contribution. The latter term may be calculated from experimentally determined quantity ($\frac{\partial \ln K}{\partial \ln[Na^+]} = SK$). Ballin et al (Ballin et al., 2004) have shown

that $\Delta G_{pe} = (SK)RT \ln[MX]$, where MX is the monovalent salt concentration. Values for parameters calculated using these equations are give in Table 1. The magnitude of ΔG_t provides a measure of the nonpolyelectrolyte forces that stabilize the DNA-protein complexes (nonelectrolyte forces include all other type of molecular interactions and transfer processes other than polyelectrolyte effects). ΔG_{pe} is the free energy contribution arising from coupled polyelectrolyte effects, especially the release of the condensed counter-ions from the DNA helix upon binding the charged ligands. Although non-polyelectrolyte forces contribute substantially to the stability of ligand-DNA complexes, the polyelectrolyte interactions provide most of the free energy of association.

4.5.4 Comparison of the ligands binding to poly(dA-dT)$_2$ and poly(dA)poly(dT): enthalpy and entropy compensation

The ATHP binds to poly(dA-dT)$_2$, at 25°C, with a ΔH of -6.8 kcal mol^{-1}, a ΔG of -10.1 kcal mol^{-1}, and a small ΔS of +11.1 cal mol^{-1} K^{-1} (-TΔS, -3.3 kcal mol^{-1}). This binding reaction is therefore overwhelmingly enthalpy-driven. Surprisingly, the binding of the ATHP to poly(dA)poly(dT), at 25°C, is entropy-driven with a small unfavorable ΔH of +3.1 kcal mol^{-1}, a favorable ΔG of -9.6 kcal mol^{-1}, and a large favorable TΔS of +42.6 cal mol^{-1} K^{-1} (-TΔS = -12.7 kcal mol^{-1}). Similar results were obtained for HMGA2 binding to AT DNAs. HMGA2 binds to poly(dA-dT)$_2$, at 25°C, through an enthalpically-driven interaction with a large favorable ΔH of -29.1 kcal mol^{-1}, a ΔG of -19.5 kcal mol^{-1}, and an unfavorable -TΔS = +9.7 kcal mol^{-1}. In contrast, the interaction of HMGA2 with poly(dA)poly(dT), at 25°C is entropically-drive with a small unfavorable ΔH of

+0.6 kcal mol^{-1}, a favorable ΔG of -17.7 kcal mol^{-1}, and a large favorable $-T\Delta S$ of -18.3 kcal mol^{-1}. This is a typical example of enthalpy/entropy compensation for the two ligands binding to different DNA substrates with similar binding free energies (Breslauer et al., 1987; Jen-Jacobson et al., 2000). This DNA binding property is a hallmark for several minor groove binding drugs, such as netropsin, binding to different DNA substrates (Breslauer et al., 1987). As demonstrated by Breslauer and his co-workers (Marky et al., 1987), the binding of netropsin to polyd(dA-dT)$_2$ is overwhelmingly driven by enthalpy but the binding to poly(dA)poly(dT) is driven by entropy. They attributed this difference (enthalpy/entropy compensation) to the formation of a more dehydrated poly(dA)poly(dT)-netropsin complex than the corresponding poly(dA-dT)$_2$-drug complex (Chalikian et al., 1994). Specifically, netropsin binding to poly(dA)poly(dT) releases 18 more water molecules than when it binds to poly(dA-dT)$_2$ (Chalikian et al., 1994). The ATHP and HMGA2 may have analogous binding properties to these small minor groove binding antibiotics; their binding to poly(dA)poly(dT) may release more water molecules than their binding to poly(dA-dT)$_2$. Another possibility of the unusual thermodynamic profile of HMGA2 binding to poly(dA)poly(dT) may result from the conformational transition in poly(dA)poly(dT) coupled to the binding reaction. As demonstrated previously (Herrera & Chaires, 1989), certain small DNA binding ligands binding to poly(dA)poly(dT) is coupled to the DNA conformational transition. The positive binding enthalpy may come from a helix to helix transition of poly(dA)poly(dT). Nevertheless, the coupled conformational transition is directly linked to the disruption of hydration in poly(dA)poly(dT) (Herrera & Chaires, 1989).

The binding of both the ATHP and HMGA2 to poly(dA-dT)$_2$ is associated with large favorable enthalpy contributions. This favorable enthalpy probably arises from the formation of different non-covalent bonds such as hydrogen bond, electrostatic interactions and nonpolar contacts between proteins and DNA. Interestingly, the DNA binding enthalpy for both ligands is strongly dependent on the ionic strength (Figure 13), suggesting that ionic dependent enthalpy significantly contributes to ligand-DNA interaction. These results contradict with Manning's counterion condensation theory where ligand-DNA interactions, including specific and nonspecific interactions, are driven by the entropic release of counterions associated with the DNA backbone (Record et al., 1978; Manning, 1978). More recently, Honig and his co-workers used the nonlinear Poisson-Boltzmann (PB) equation to evaluate the salt dependent contribution to the electrostatic binding free energy of several small minor groove binders (Misra et al., 1994). They found that the electrostatic free energy of ligand-DNA interactions has significant enthalpic and entropic contributions. In their theory, the counterion atmosphere around a polyion (such as DNA) is modeled as a single population described by a continuously distributed salt dependent ion atmosphere. In such a case, the electrostatic enthalpy is strongly dependent on the bulk salt concentration. For the ATHP and HMGA2 binding to poly(dA)poly(dT), the favorable enthalpy gained from electrostatic ion-macromolecule or ion-ion interactions may be used up to release more water molecules from ligand-DNA complexes, which results in an increase in entropy (enthalpy-entropy compensation). Figure 17 summarizes the enthalpic and entropic contributions to the binding free energy for the ATHP and HMGA2 binding to poly(dA-dT)$_2$ and poly(dA)poly(dT).

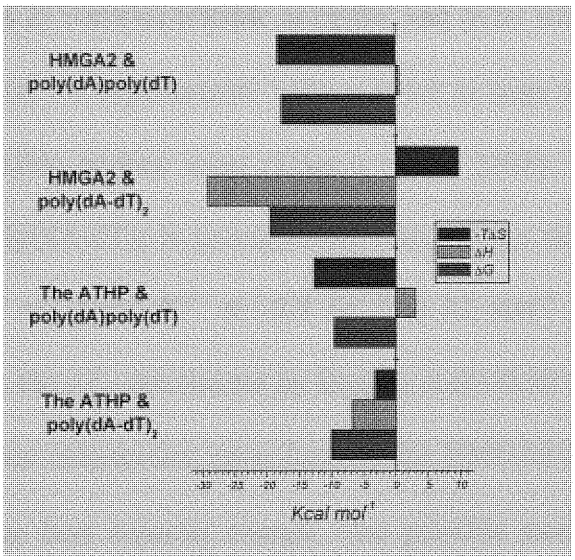


Figure 17. Comparison of the thermodynamic profiles for the ATHP and HMGA2 binding to poly(dA-dT)$_2$ and poly(dA)poly(dT).

4.5.5 Temperature dependence of binding enthalpy

As demonstrated by Dragan et al (Dragan et al., 2003), the binding of AT hook peptides, which contain one "AT hook" DNA binding domain, to a DNA duplex 5'-GGGAAATTCCGC-3' (top strand) is accompanied by a negative heat capacity change (ΔC_p). In their case, the AT hook peptides bind to DNA with an unfavorable positive enthalpy and a favorable entropy contribution. The binding reaction is entropy-driven, which is similar to the binding process between our ATHP or HMGA2 and poly(dA)poly(dT). Furthermore, they suggested that the heat capacity change is a result of dehydration of the groups forming the interface in the complexes (As shown previously (Spolar & Record, 1994), a large heat capacity change is associated with the formation of a highly complementary interface between protein and DNA; ΔC_p can be

calculated from changes in water-accessible nonpolar and polar surface area (Spolar & Record, 1994)). However, the experimental value is twice as large as the ΔC_p calculated from their water-accessible surface areas (Dragan et al., 2003). We also found that HMGA2 binding to poly(dA-dT)$_2$ is accompanied with a large negative heat capacity change (-705 (± 113) cal mol^{-1} K^{-1} or -2.9 (± 0.5) kJ mol^{-1} K^{-1}). This value is significantly larger than the ΔC_p determined by Dragan et al (-0.30 to -0.6 kJ mol^{-1} K^{-1}). These results suggest that dehydration upon forming the HMGA2-DNA complex alone cannot explain the large heat capacity change. One possible mechanism for the ΔC_p discrepancy is protein folding induced by DNA binding (Spolar & Record, 1994). Our CD studies suggest that, in the absence of DNA, HMGA2 has little secondary structure and may be unfolded. NMR studies show that on binding to DNA, AT hook domains adopt a crescent-shaped conformation (Huth et al., 1997). This induced disorder to order conformational transition may significantly contribute to ΔC_p.

Another explanation for ΔC_p discrepancy stems from the charge-charge interactions between highly positively charged HMGA2 and highly negatively charged DNA. As described by Sturtevant (Breslauer & Sturtevant, 1977), creation of a positive and negative pair of charges in aqueous solution leads to ΔC_p = -20 to -50 cal mol^{-1} K^{-1}. Inspection of the NMR structure of a truncated form of HMGA1 and DNA showed several well defined electrostatic contacts between the side chains of Arg and Lys and the DNA phosphates (Huth et al., 1997). Particularly, two Arg residues in the RGR core of the "AT hook" DNA binding domain deeply penetrate into the narrow minor groove.

Since the minor groove has a high negative potential (Lavery & Pullman, 1981) and a low dielectric constant (Jin & Breslauer, 1988), the contribution of the electrostatic interaction (Columbic forces) in the minor groove to the ΔG, ΔH and ΔC_p of binding should be substantial. Recently, Holbrook et al demonstrated that binding of integration host factor (IHF) of *E. coli* (another minor groove binding protein) to its cognate sequence from λ phage DNA produces a large negative binding enthalpy and a negative heat capacity change; the magnitudes of enthalpy and heat capacity changes decrease strongly with increasing $[K^+]$ (Holbrook et al., 2001). They interpreted these data by hypothesizing that 22 of the 23 IHF cationic side chains contacting with DNA phosphate oxygen atoms in the IHF-DNA complex are masked in the absence of DNA by pairing with negatively charged carboxylate groups, i.e. forming dehydrated ion-pairs. IHF binding to DNA results in disruption of multiple salt bridges on the protein surface followed by hydration of the unmasked charged groups and formation of interactions between positively charged groups and DNA. Interestingly, the charge distribution of HMGA2 is asymmetrical with the positively charged residues mainly concentrated in the center of the sequence and the negatively charged residues at the C-terminus. We found that the negatively charged C-terminus interacts with the positively charged "AT hook" DNA binding domains in the absence of DNA by forming multiple salt bridges (unpublished results). Potentially, these salt bridges could be disrupted when HMGA2 binds to DNA. Further studies are required to fully understand the high ΔC_p.

4.5.6 Binding selectivity

It has been suggested that the "AT hook" proteins, HMGA1a, HMGA1b, and HMGA2, only recognize the minor groove of AT DNAs and do not recognize specific DNA sequences (Solomon et al., 1986). However, the solution structure of an ATHP-DNA complex showed that the "AT hook" DNA binding domain binds to the DNA sequence 5'-A_4AATT_8-3' (numbers 4 and 8 represent the positions of the bases in the DNA dodecamer used for the structure determination) in a specific orientation with the N-terminal Arg of the core RGR located near base pair 8 and C-terminal Arg located near base pair 4 (Figure 8(a)) (Huth et al., 1997). There are extensive hydrophobic and polar contacts between the "AT hook" DNA binding domain and DNA. We showed here that both the ATHP and HMGA2, at 25°C, bind to poly(dA-dT)$_2$ with a higher affinity than for poly(dA)poly(dT) (Table 1). These studies suggest that the "AT hook" DNA binding domain does have sequence selectivity. Similar to other small minor groove binders, the "AT hook" DNA binding domain may recognize all 5 bp AT sequences but with different binding affinities. The sequence selectivity of the ATHP is currently under investigation in our laboratory.

4.6 Conclusions

The following are major conclusions drawn from this study: (1) One "AT hook" DNA binding domain binds to 5 AT base pairs and one HMGA2 molecule binds to 15 AT base pairs. (2) HMGA2 binds with very high affinity to both poly(dA-dT)$_2$ and poly(dA)poly(dT) and electrostatic interactions are a dominant force in their interactions with DNA. (3) The interaction of HMGA2 and poly(dA-dT)$_2$ is enthalpy driven. In

contrast, the interaction between HMGA2 and poly(dA)poly(dT) is entropy driven. This enthalpy-entropy compensation property is a signature for DNA minor groove binders, and may be a reflection of dehydration difference of different ligand-DNA complexes. (4) The binding reaction between HMGA2 and poly(dA-dT)$_2$ was temperature-dependent with a ΔC_p of -705 (± 113) cal $mol^{-1}K^{-1}$.

5. Specific Recognition of AT-Rich DNA Sequences by the Mammalian High Mobility Group Protein AT-hook 2: A SELEX study

5.1 Abstract

The Mammalian high mobility group AT-hook 2 (HMGA2) is a transcriptional factor involved in cell differentiation and transformation. Disruption of its normal expression pattern is directly linked to oncogenesis and obesity. HMGA2 contains three "AT-hook" DNA binding domains, which specifically bind to the minor groove of AT-rich sequences. Using a PCR-based systematic evolution of ligands by exponential enrichment (SELEX) procedure, we have identified two consensus sequences for HMGA2: 5'ATATTCGCGAWWATT3' and 5'-ATATTGCGCAWWATT-3' where W represents A or T. These two consensus sequences have a unique and interesting feature: the first five base pairs are AT-rich, the middle four base pairs are GC-rich, and the last six base pairs are AT-rich. Our results showed that all three of these segments are critical for high affinity binding of HMGA2 to DNA. For example, if one of the AT-rich sequences is mutated to a non-AT-rich sequence, the DNA binding affinity of HMGA2 is reduced at least 100-fold. Intriguingly, if the GC-segment is replaced by an AT-rich segment, the binding affinity of HMGA2 is reduced approximately 5-fold. Identification of the consensus sequences for HMGA2 represents an important step towards finding its binding sites within the genome.

5.2 Introduction

The mammalian high mobility group protein AT-hook 2 (HMGA2) is a transcriptional factor regulating mesenchymal cell development and differentiation (Wolffe et al., 1994; Zhou et al., 1995; Anand et al., 2000; Reeves, 2003). It was first discovered as a nuclear protein in proliferating fibroblasts and embryos, and also in transformed thyroid cells infected by oncogenic viruses (Goodwin et al., 1985; Giancotti et al., 1985). The aberrant expression of HMGA2 has been attributed to the formation of a variety of tumors, including benign tumors, such as lipomas (Ashar et al., 1995; Schoenmakers et al., 1995) and uterine leiomyomas (Kools et al., 1995; Schoenmakers et al., 1999; Klotzbucher et al., 1999), and malignant tumors, such as breast cancer (Rommel et al., 1997), lung cancer (Rogalla et al., 1998; Sarhadi et al., 2006; Meyer et al., 2007), and leukemia (Rommel et al., 1997; Kottickal et al., 1998). The expression level is correlated with the degrees of malignancy and metastatic potential of the transformed cells (Tallini et al., 1999), suggesting that HMGA2 can be used as a biomarker for diagnosing the neoplastic transformation and the metastatic potential of many cancers (Giancotti et al., 1991; Rogalla et al., 1998). HMGA2 is only expressed in proliferating, undifferentiated mesenchymal cells and is undetectable in normal fully differentiated adult cells (Zhou et al., 1995; Gattas et al., 1999). Disruption of its normal expression patterns causes deregulations of cell growth and differentiation. For example, *hmga2* knock-out mice developed the pygmy phenotype (Zhou et al., 1995). These mutant mice were severely deficient in fat cells and other mesenchymal tissues. Breaking of the *hmga2* gene caused a dramatic reduction in obesity of leptin-deficient mice (Lep^{ob}/Lep^{ob}) in a gene dosage dependent manner (Anand et al., 2000). These results

suggest that HMGA2 plays an important role in fat cell proliferation and is a potential target for the treatment of obesity (Anand et al., 2000).

HMGA2 is a member of the HMGA family, which also includes another two proteins, HMGA1a and 1b. HMGA1a and 1b are splice variants of the same gene, the HMGA1 gene (Friedmann et al., 1993). HMGA2 is the product of a separate gene, the HMGA2 gene (Chau et al., 1995). One unique feature of HMGA proteins is that all have three "AT-hook" DNA binding domains. These DNA binding domains contain a consensus sequence, PRGRP, flanked on each side by one or two positively charged amino acids (arginine or lysine). The "AT-hook" DNA binding domain, in the absence of DNA, is "unstructured" (Lehn et al., 1988; Huth et al, 1997). However, when it binds to the minor groove of AT-rich sequences, this domain adopts a defined conformation (Huth et al., 1997). The central core, RGR, deeply penetrates into the minor groove of AT base pairs with two arginine residues, forming extensive electrostatic and hydrophobic contacts with the floor of the minor groove (Huth et al., 1997). Two prolines on each side of the RGR core direct the protein away from the floor of the minor groove and position the positively charged arginine or lysine near the negatively charged phosphate backbone, thereby making further contacts. This disordered-to-ordered conformational change significantly increases HMGA proteins' adaptability and allows them to participate in a variety of nuclear activities such as gene transcription, DNA replication, chromatin remodeling, and DNA repair (Goodwin, 1998; Reeves, 2003).

Since their discovery, HMGA proteins were shown to preferentially bind to AT-rich DNA sequences in the promoter regions. The early qualitative footprinting studies showed that they could bind to any run of five to six AT base pairs with similar DNA binding affinity (Solomon et al., 1986). These studies also suggested that HMGA proteins only recognize the minor groove configuration of AT base pairs rather than specific sequences. However, more recent studies showed that HMGA proteins should have sequence specificity. For example, the high affinity binding of HMGA proteins requires two to three appropriately spaced AT-rich sequences as a single multivalent binding site (Maher et al., 1996). The DNA binding affinity of this multivalent binding is much higher than that of the single valent binding (Maher et al., 1996). More significantly, each HMGA protein always simultaneously binds to two to three runs of AT base pairs in the regulatory transcription regions, such as the human interferon β enhancer (Thanos & Maniatis, 1992; Du et al., 1993), the promoter regions of interleukin-2 gene (Baldassarre et al., 2001) and interleukin-2 receptor α-chain gene (John et al., 1995), and the promoter region of interleukin-15 gene (Reeves & Beckerbauer, 2001). These results suggest that HMGA proteins bind to specific DNA sequences as transcriptional factors. More evidence for sequence-specific DNA binding of HMGA proteins comes from the NMR structural studies (Huth et al., 1997). The structure of an HMGA1a-DNA complex showed that the "AT-hook" DNA binding domain binds to the 5'-AAATT-3' sequence in a fixed orientation with the N-terminal arginine of the core RGR sequence located near the 3'-end of the sequence, and the C-terminal arginine located near the 5'-end of the sequence. This binding orientation resulted from hydrophobic interactions of the arginine side chains of the RGR core with

the adenine bases of the DNA sequence, and hydrogen bonding between the "AT-hook" DNA binding domains and the bases of the binding site. These results suggest that HMGA proteins do not randomly bind to any AT-rich DNA sequences. In this study, we used an *in vitro* systematic evolution of ligands by exponential enrichment (SELEX) experiment to further investigate the DNA binding specificity of HMGA2. Our results showed that HMGA2 specifically recognizes a type of 15 bp AT-rich DNA sequence: the first 5 base pairs are AT-rich, the middle 4 or 5 base pairs are GC-rich, and the last 5 or 6 base pairs are AT-rich. Interestingly, all three segments are critical for high affinity binding of HMGA2 to DNA.

5.3 Materials and Methods

5.3.1 Preparation of ^{3}H-HMGA2 (^{3}H-labeled HMGA2)

E. coli strain BLR(DE3), which has plasmid pMGM1, was grown in 200 ml of a supplemented M9 medium containing 1 × M9 salts, 0.4% glucose, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.1 μg/ml thiamine, 40 μg/ml L-leucine, 150 μg/ml L-proline, 50 μg/ml each of the other 18 L-amino acids, and 50 μg/ml kanamycin. The cell growth was monitored by measuring OD_{595}. At $OD_{595} \approx 1$, cells were harvested by centrifugation at 4,000 rpm and 4 °C, and re-suspended in 100 ml of the supplemented M9 medium in the absence of L-lysine. After 40 min incubation at 37 °C, another 100 ml of the supplemented M9 medium (prewarmed to 37 °C) containing 10 mCi of ^{3}H-L-lysine (10 μg/ml) was added into the cell culture. When the OD_{595} reached ~0.8, HMGA2 expression was induced by addition of 1 mM of IPTG into the cell culture. After an additional three-hour incubation,

the cells were harvested by centrifugation at 4,000 rpm for 25 minutes at 4 °C. ^{3}H-HMGA2 was purified as described previously (Cui et al, 2007). The specific activity of ^{3}H-HMGA2 was determined to be 375 cpm/pmol.

5.3.2 Systematic evolution of ligands by exponential enrichment (SELEX) experiment

SELEX experiments started with a 61 bp DNA library that contains 15 random nucleotides. This DNA library was constructed through amplification of a 61 base synthetic oligodeoxynucleotide FL-250, 5'-CATGGTACCTCTAGAGGCTCGAG$(N)_{15}$GCTAGCTGGCATGCAAGCTTCAC-3'(N_{15} represents the 15 random nucleotides and underlined sequences are cleavage sites for Kpn I and Hind III, respectively), using two primers, FL-248 (5'-CATGGTACCTCTAGAGGC-3') and FL-249 (5'-GTGAAGCTTGCATGCCAG-3'), which correspond to the first 18 bases (top strand) and the last 18 bases (bottom strand). For the first round of selection, 80 nM of DNA was incubated with increasing amount of HMGA2 in 12 μl of 1 × DNA binding buffer containing 20 mM Tris-HCl (pH 8.0), 0.5 mM EDTA, 1 mM DTT, 0.5 mM $MgCl_2$, 50 mM NaCl, and 5% glycerol. After 30 min incubation at 22 °C, the DNA samples were loaded on a 12% native polyacrylamide gel to separate the bound and free species. Band shifts were detected by SYBR Gold staining. The bound DNA was excised and eluted in an elution buffer containing 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, and 100 mM NaCl, and purified by phenol extraction and ethanol precipitation. The purified DNA was dissolved in water and PCR amplified for the next round of SELEX selection. 20 cycles of PCR amplification were performed

for 1 min each at 95, 55, and 72 °C with an additional step of 3 min at 72 °C for the final cycle of the PCR reaction. The PCR products were purified by phenol extraction and ethanol precipitation. A total of 10 rounds of SELEX selection were performed. After the 8th round, the concentration of NaCl was increased to 200 mM, and 30 μM (bp) of poly(dG-dC)$_2$ was added to the binding reactions as a competitor to promote selection towards a high affinity and specificity pool of DNA fragments. After the 10th round of selection, the PCR-amplified products were cloned into Hind III-Kpn I sites of pUC18, which were used to transform *E. coli* strain DH5α. The plasmid DNA of each transformed colony was purified and sequenced. DNA sequences, corresponding to the 15 random nucleotides, were analyzed using programs ClustalX 1.81 (Thompson et al., 1997) and MEME 3.5.3 (Bailey et al., 2006).

5.3.3 Electrophoretic mobility shift assay (EMSA)

EMSA experiments were used to determine the apparent DNA binding constant of HMGA2. DNA oligomers containing AT-rich DNA sequences were labeled with ^{32}P at 5' termini by T4 polynucleotide kinase in the presence of γ-^{32}P-ATP. The protein-DNA complexes were formed by addition of appropriate amounts of the protein to a solution containing 1 nM of ^{32}P-labeled DNA in the 1 × DNA-binding buffer containing 20 mM Tris-HCl (pH 8.0), 200 mM NaCl, 1.5 μM (bp) poly(dG-dC)$_2$, 0.5 mM EDTA, 1 mM DTT, 0.5 mM $MgCl_2$, and 5% glycerol. After equilibration for 60 min at 22 °C, the samples were loaded on a 12% native polyacryamide gel in 0.5 × TBE buffer (0.045 M Tris-Borate (pH 8.3) and 1 mM EDTA) to separate free and bound DNA. The gels were

subsequently dried and visualized by autoradiography or quantitated using a Fuji FLA 3000 image analyzer. The radioactivity of the free and bound DNA was determined and used to calculate the binding ratio (R), which is equal to the ratio of the radioactivity of the bound DNA divided by the sum of the radioactivity of the bound and free DNA. The apparent DNA binding constant (K_{app}) was obtained by nonlinear-least-squares fitting the following equation using the program Scientist.

$$R = \frac{(a + x + 1/K_{app}) - \sqrt{(a + x + 1/K_{app})^2 - 4ax}}{2a} \quad (1)$$

where a and x represent the total DNA and the total protein concentration, respectively.

EMSA experiments were also used to determine binding stoichiometries of HMGA2 binding to DNA using ^{32}P-labeled ologonucleotides and ^{3}H-HMGA2. In these experiments, 200 nM of ^{32}P-labeled AT-rich DNA oligomers with a specific activity of ~700 cpm/pmol were mixed with 0.5 μM of ^{3}H-HMGA2 in 50 μl in the 1 × DNA binding buffer as described above. After incubation at 22 °C for 1 hour, the samples were loaded on a 12% native polyacryamide gel in 0.5 × TBE buffer to separate free and bound DNA. The gels were subsequently stained by SYBR Gold. DNA bands corresponding to the HMGA2-DNA complexes were excised and electro-eluted into a dialysis bag (8000 MWCO, BioDesign, NY) in 1 × TBE containing 0.1% SDS. The eluents in the dialysis bag were mixed with 15 ml of Aquasol 2 (PerkinElmer, MA), and counted in ^{32}P and ^{3}H channels in a liquid scintillation spectrometer to determine the molar ratio of HMGA2 to DNA oligomers. Alternatively, a method as previously described by Carey (Carey, 1988) was also used to determine the binding

stoichiometries. After EMSA experiments, the DNA bands were located by autoradiography and excised. Gel slices containing the HMGA2-DNA complex were solubilized by oxidation in 1 ml of 21% H_2O_2/17% $HClO_4$ in a tightly sealed scintillation vial at 65 °C for 24 hours. Vials were cooled to 22 °C, mixed with 15 ml of Aquasol 2, and then stored at 4 °C for 48 hours in dark. Radioactivities of ^{3}H and ^{32}P were counted in ^{3}H and ^{32}P channels using the method as described above. Background was determined by counting the control gel slices excised from the unused regions.

5.4 Results and Discussion

5.4.1 Determining the binding stoichiometries of HMGA2 binding to AT-rich DNA oligonucleotides

Mouse HMGA2 is a small basic protein containing 108 amino acid residues. One unique feature of HMGA2 is the asymmetric charge distribution of its primary structure (Figure 3). The positive charges are mainly concentrated in the three "AT hooks" and the negative charges at the C-terminus. In solution, HMGA2 may self-associate into homodimers or homo-oligomers through electrostatic interactions and may bind to AT-rich DNA as a homodimer. Thanos and colleagues showed that HMGA1a, another member of the HMGA family with similar physical properties, interacts with itself in a glutathione S-transferase (GST)-pull down experiment (Yie et al., 1999). Our preliminary results showed that free HMGA2 in solution may be a homodimer (data not shown). We therefore decided to determine whether HMGA2 binds to AT-rich DNA as a monomer or as a homodimer.

As demonstrated in our previous publication (Cui et al., 2005), each HMGA2 (monomer) binds to 15 AT base pairs in aqueous buffer solution. In this study, we used two DNA oligomers, FL-AT15 (5'-CATGGTACCTTCAGAGGCTCGAGAAAAAAAAAAAAAAAGCTGACTGGCATGCAAGCTG-3') and FL-AT30 (5'-CATGGTACCTTCAGAGGCTCGAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCTGACTGGCATGCAAGCTG-3'), to determine the stoichiometries of HMGA2 binding to AT-rich DNA sequences. FL-AT15 contains a 15 bp A-track (one HMGA2-binding site) and FL-AT30 contains a 30 bp A-track (two HMGA2-binding sites). As an initial step, the apparent DNA binding constants of HMGA2 (K_{app}) were determined using electrophoretic mobility shift assay (EMSA). Figure S3A shows the results of HMGA2 titrating into a solution containing FL-AT15. In this EMSA experiment, only one shift band was observed (the bands at the top, Figure S3A). The apparent DNA binding constant of HMGA2 binding to FL-AT15 was estimated to be 2.2×10^6 M^{-1} according to the method as described under "Materials and Methods." Since we used 200 mM of NaCl and 1.5 μM (bp) of poly(dG-dC)$_2$ (as a competitor) to reduce the non-specific binding between HMGA2 and the oligomers (see Figure S3 legend for details), the relatively low value of K_{app} may represent specific binding of HMGA2 to the AT-rich region of FL-AT15. Figure S3B shows results of HMGA2 titrating into a solution containing FL-AT30 in which two shift bands were generated. Because FL-AT30 contains two HMGA2-binding sites, it is reasonable to assume that these two shift bands represent HMGA2 occupying one and two sites on FL-AT30, respectively. Interestingly, the apparent binding constant of HMGA2 binding to the first site of FL-AT30 was

estimated to be 2.0×10^7 M^{-1}, 10-fold higher than K_{app} of FL-AT15. These results indicate that HMGA2 may cooperatively bind to DNA containing multiple AT-rich sites.

The binding stoichiometries of HMGA2-DNA complexes were measured directly in EMSA experiments using ^{3}H-HMGA2 and ^{32}P-labeled FL-AT15 or FL-AT30 (the double-label experiments). In these experiments, high concentrations of HMGA2 and DNA oligomers were used for stoichiometric binding (Figure 18). In addition, 200 mM of NaCl (physiologically relevant salt concentration) and 30 μM (bp) of poly(dG-dC)$_2$ (as a competitor for non-specific binding) were added to the binding reactions. The shifted bands were excised; the radioactivities of ^{3}H and ^{32}P in the protein-DNA complexes were determined by scintillation counting. As described under "Material and Methods," two methods were used to elute ^{3}H-HMGA2 and ^{32}P-labeled oligomers into solutions. We either electro-eluted ^{3}H-HMGA2 and the ^{32}P-oligomers into 1×TBE containing 0.1% SDS or dissolved the gel slices into a solution containing 21% H_2O_2 and 17% $HClO_4$ by oxidization. No difference was found between these two methods. Our results are summarized in Table 4. For the HMGA2-FL-AT15 complex, the shifted band gave a 1:1 molar ratio of HMGA2 per FL-AT15 molecule. For the HMGA2-FL-AT30 complex, the first shifted band yielded a 1:1 molar ratio of HMGA2 per FL-AT30 molecule and the second shifted band produced a 2:1 molar ratio of HMGA2 per FL-AT30 molecule. These results indicate that HMGA2 binds to AT-rich sites as a monomer under physiologically relevant salt conditions. Similar results were also obtained by using DNA oligomers containing one or two HMGA2 DNA binding sites determined by SELEX experiments (data not shown; see below for the SELEX experiments).

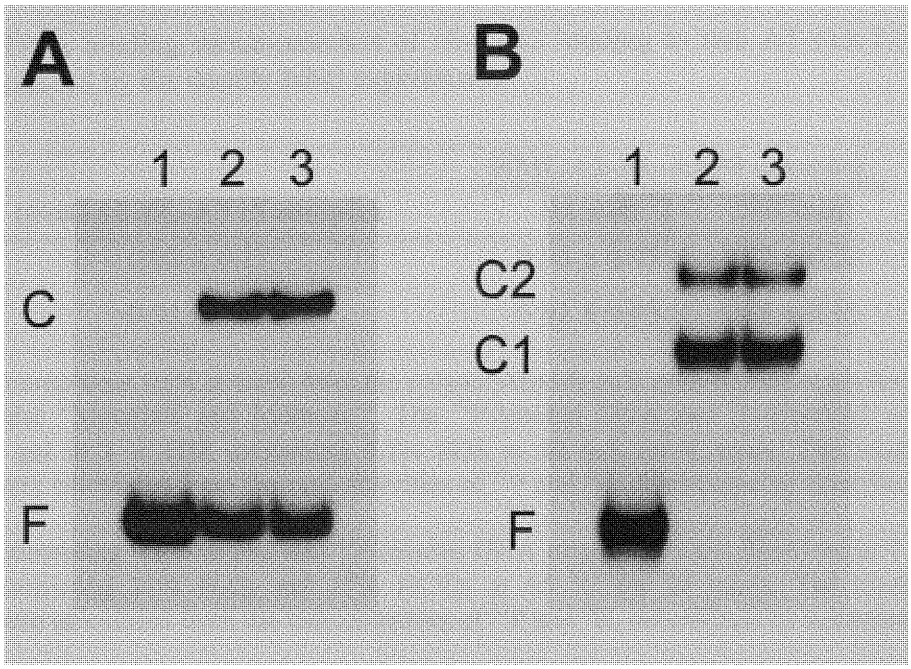


Figure 18. Determination of DNA-binding stoichiometries of the HMGA2-DNA complexes by double-label experiments. EMSA experiments using ^{32}P-labeled FL-AT15 (A) or FL-AT30 (B) and ^{3}H-HMGA2 were performed as detailed under "Materials and Methods." The autoradiograms of the ^{32}P-labeled DNA oligomers were shown and used to excise the DNA bands for scintillation counting. Lane 1 is the free DNA. Lanes 2 and 3 contained ^{32}P-labeled DNA oligomers and ^{3}H-HMGA2. F is the free DNA; C represents the HMGA2-FL-AT15 complex; C1 and C2 represent the first and second shifted bands of HMGA2-FL-AT30 complexes, respectively.

5.4.2 *In vitro* selection of DNA oligomers that bind to HMGA2 with high affinity

We next used systematic evolution of ligands by exponential enrichment (SELEX; (Robertson & Joyce, 1990; Tuerk & Gold, 1990; Ellington & Szostak, 1990; Cui et al., 1995)) experiments to determine the consensus DNA binding sequences for

Table 4. The DNA binding stiochiometries of HMGA2 binding to two AT-rich DNA oligomers determined by the double-label experiments

		HMGA2-FL-AT30 complex	
	HMGA2-FL-AT15 complex	1st shift	2nd shift
Method I [a]	0.92 ± 0.17	0.98 ± 0.05	1.97 ± 0.23
Method II [b]	0.93 ± 0.05	0.92 ± 0.05	1.88 ± 0.13

[a]HMGA2-DNA complexes were excised from the polyacrylamide gels after EMSA experiments as described under "Materials and Methods.". The ^{32}P-labeled DNA and ^{3}H-HMGA2 were eluted from the gel in 1 × TBE containing 0.1% SDS at 50 mA for 4 hours. The binding stoichiometris were calculated from the radioactivities of ^{32}P and ^{3}H, determined by scintillation counting.

[b]The gel slices containing ^{3}H-HMGA2 and ^{32}P-labeled DNA oligomers were dissolved in 1 ml of a solution containing 21% H_2O_2 and 17% $HClO_4$ at 65 °C for 24 hours. The binding stoichiometris were calculated from the radioactivities of ^{32}P and ^{3}H, determined by scintillation counting.

HMGA2 (Figure 19A). In these experiments, we utilized a library of DNA oligonucleotides containing a central randomized region of 15 bp, flanked by two 23 bp regions with defined sequences in each end (total 61 bp; Figure 19A). 15 bp randomized sequences were chosen because each HMGA2 molecule binds to 15 AT base pairs (Cui et al., 2005). For the first round of the SELEX experiments, 9.6×10^{-13} mol of DNA oligonucleotides (~5.8×10^{11} molecules) were used to ensure a completely random population for selection (approximately 500 copies of each of the 4^{15} possible sequences on average). This library of oligonucleotides was mixed with increasing amounts of HMGA2, incubated at 22 °C for 60 min, and loaded on a 12% polyacrylamide gel to separate free and bound DNA. FL-AT15 was used as a control to define the position to which the HMGA2-DNA complex was migrated. The gel containing the HMGA2-DNA complexes was excised; DNA was eluted from the gel and amplified by PCR according

to methods as described under "Materials and Methods." The PCR products were purified and subjected to the next round of selection. After the 8th round, the concentration of NaCl was increased to 200 mM, and 30 μM (bp) of poly(dG-dC)$_2$ was added to the binding reactions to increase the binding specificity. The progress of the enrichment of DNA molecules recognized by HMGA2 was assessed by an EMSA experiment in which no HMGA2-DNA complex of the original DNA library and the DNA pool of the first round of selection was visible in the presence of 200 nM of HMGA2 (lanes 1 and 2, Figure 19B). In contrast, the DNA pools of rounds 3, 5, and 10 gave approximately 7.1%, 17.5%, and 35.2% of HMGA2-DNA complexes under the same conditions (lanes 3-5, Figure 19B).

After the 10th round of selection, the purified PCR products were cloned into Hind III-Kpn I sites of pUC18 and sequenced. We repeated this experiment twice and sequenced 102 individual clones in which 71 sequences are unique. Figures 20A and S4 show our sequencing results. To our surprise, all sequences have a common unique feature: the first five base pairs are AT-rich, the middle four to five base pairs are GC-rich, and the last five to six base pairs are AT-rich. These results suggest that our conditions for selection were optimal and that enrichment of the highest-affinity oligonucleutodes was successful. The 71 unique sequences were further analyzed by two multiple-sequence-alignment programs, ClustalX 1.81 (Thompson et al., 1997) and MEME 3.5.3 (Bailey et al., 2006), and two consensus sequences, 5'-ATATTCGCGAWWATT-3' and 5'-ATATTGCGCAWWATT-3' where W represents

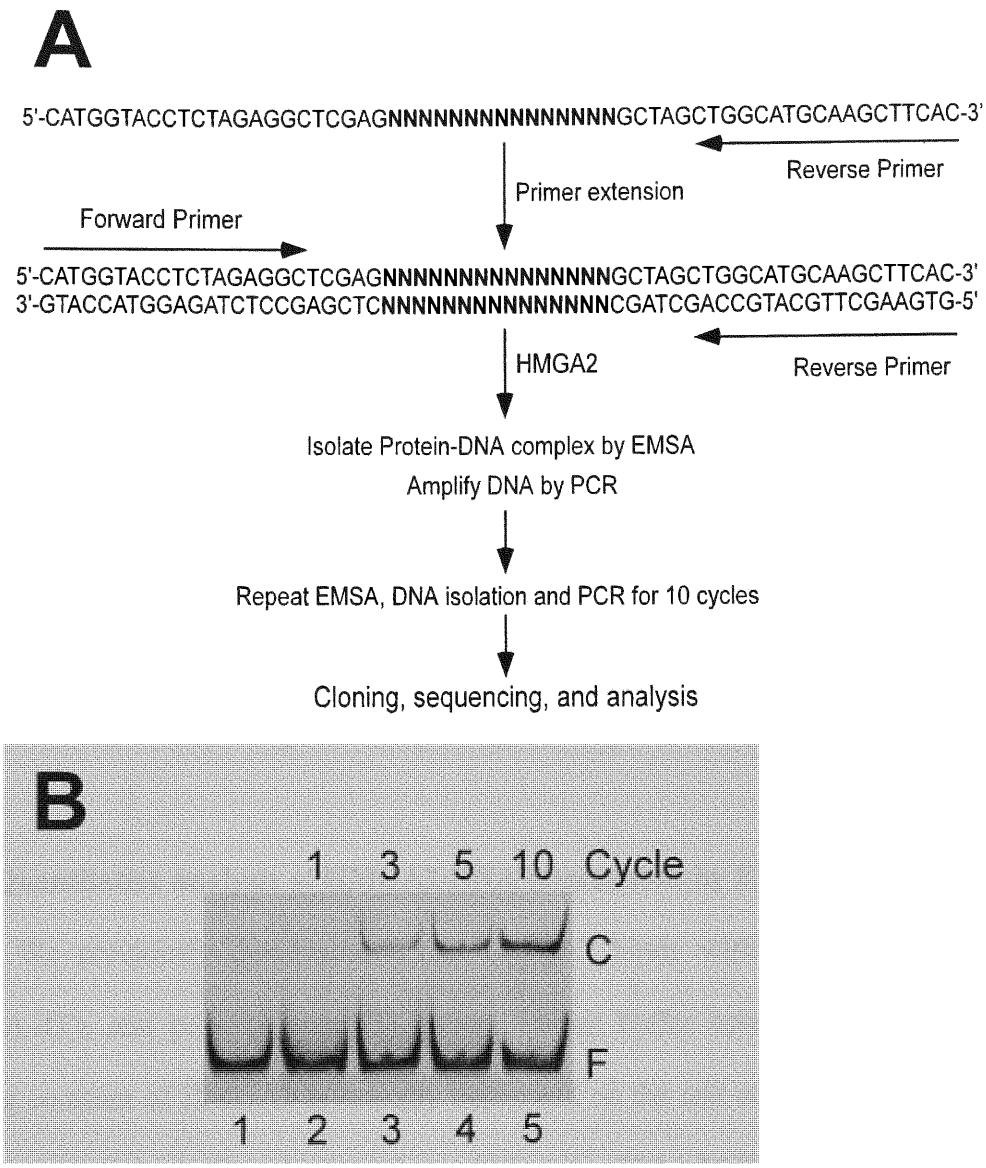


Figure 19. The SELEX experiments. (A) Strategy for SELEX analysis of HMGA2 DNA binding sites. (B) Representative EMSA analysis used to monitor the progress of enrichment for HMGA2 binding sites. About 80 nM of the DNA samples isolated after the first (lane 2), third (lane 3), fifth (lane 4), and tenth (lane 5) rounds of the selection were incubated with 200 nM of HMGA2 in an EMSA buffer containing 20 mM Tris-HCl (pH 8.0), 200 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 0.5 mM $MgCl_2$ and 5 % glycerol. EMSA experiments were performed as described under "Materials and Methods." Following electrophoresis, the PAGE gels were stained with SYBR Gold, destained, and photographed under UV light. Lane 1 is the free DNA library before the selection. The HMGA2-DNA complex and the free DNA were designated as C and F, respectively.

A or T, were identified. These two consensus sequences are identical to those obtained from manual alignment of the SELEX sequences (Table 5). Another striking feature of these sequences is that, although each clone is unique, certain sequences appeared multiple times. For example, a 5 bp AT-rich sequence, 5'-ATATT-3', occurred 58 times; two GC-rich sequences, 5'-CGCG-3' and 5'-GCGC-3', appeared 22 and 13 times, respectively (Figures 20A and S4). This finding further suggests that we have identified HMGA2 binding sites with high affinity. We also sequenced 20 clones of the original library to confirm that the starting library was generated randomly. These sequences are shown in Figure 20B. As expected, they look random and none of these sequences contains the consensus motifs, such as 5'-ATATT-3' and 5'-CGCG-3'. Our SELEX results are summarized in Figure 20C as a sequence logo.

5.4.3 Quantitative analysis of HMGA2-DNA interactions by EMSA

We previously demonstrated that HMGA2 binds with very high affinity to poly(dA-dT)$_2$ and poly(dA)poly(dT) under low-salt buffer conditions, and the DNA-binding constant of HMGA2 is strongly dependent on the salt concentration (Cui et al., 2005). At physiologically relevant salt conditions, i.e. 200 mM NaCl, it binds to AT-rich DNA with a binding constant of 10^6 to 10^7 M^{-1} (Cui et al., 2005). In the present study, we employed quantitative EMSA to measure the DNA binding constants of HMGA2 to different DNA oligomers in a buffer containing 200 mM NaCl. We also included 1.5 μM (bp) of poly(dG-dC)$_2$ (about 25-fold over 1 nM of ^{32}P-labeled oligonucleotides) as a competitor to reduce non-specific binding. In this case, the DNA-binding constant should reflect specific recognition of different DNA sequences by HMGA2. Figure 21A shows a

A

```
ATATT CGCGG ATATT
ATATT CGCGT TAATT
ATATT CGCGC TAATT
ATATT CGCGG TAAAT
ATATT CGCGC AATTT
ATATT CGCGT TAAAT
ATATT CGCGT AATAT
ATATT CGCGA TAATT
ATATT CGCGG TAATT
ATATT CGCGA ATATT
ATATT GCGCG TAATT
ATATT GCGCA TTATT
ATATT GCGCC ATATT
ATATT GCGCG ATATT
ATATT GCGCA ATATT
ATATT GCGCA TAATT
ATATT GCGCG ATTAT
AAATA CGCGA TAATT
AAAAT CGCGA TTAAT
TTAAT CGCGA TTATT
```

B

```
GCTGT TGCGT GTAGC
TGCTG CCCTG TCGGT
AGCTA GGCCG TGAGC
GGGAG CCTCG TACCG
CAGTA TCTGT TTGTG
CGGTG GGCGT GTGCC
GTCGT TGCCT TTGGT
CATGG CTTTG TCCCG
CGCAT ACGAT AGTTT
CGGCT GGAAA TGGAA
GCTGC CTTTG CAATT
TTCGT GGAGT GGTCT
TCCTG GTCGG CGTTT
GCGTT TTGTA CCTAG
TTTGT TGCGC GCCAT
GTTGT GTATC CTATT
CAGGA TGGGC GGTTC
TGTTG TGCGG GGCCG
CTAAG TCTTG TGTGC
TGCCG TGCGT GTGTG
```

C

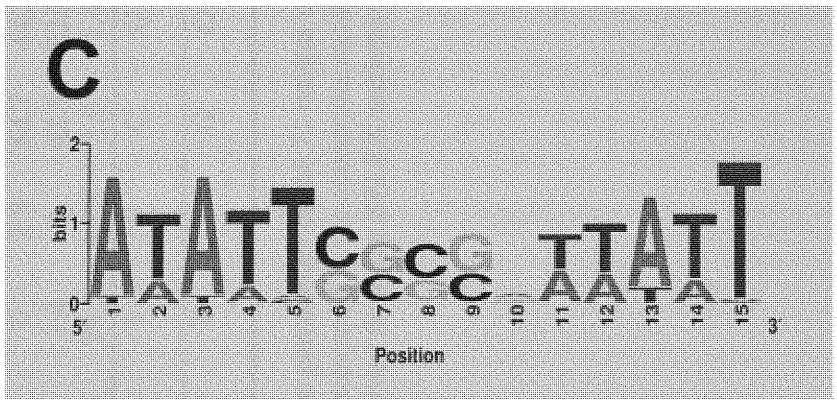


Figure 20. Sequence analysis of the SELEX experiments for HMGA2. (A) HMGA2 binding sequences identified after 10 rounds of the SELEX experiments. 20 sequences are shown here. The other 51 sequences are shown in Figure S4. (B) DNA sequences from the randomized oligonucleotide library before selection. (C) Sequence logo of the 71 SELEX sequences shown in Figures 20A and S4. Sequence conservation, measured in bits of information (2 bits is the highest), is illustrated by the height of stacking of the four letters for each position in the binding sites. The relative heights are proportional to their frequencies shown in the 71 SELEX sequences. The sequence logo is generated by WebLogo (available at www.bio.cam.ac.uk/cgi-bin/seqlogo/logo.cgi).

Table 5. Occurrence of nucleotides (%) in the consensus sequences obtained from the SELEX experiments

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	**96.2**	23.6	**95.3**	19.3	2.8	1.4	1.4	0	0.9	**42.9**	**49.1**	**36.8**	**83.5**	28.8	0.5
C	0	0	0	0	2.8	**54.7**	**44.3**	**61.3**	**40.1**	15.1	0	0	0	0	1.4
G	0	0	0	1.4	0	**42.9**	**52.8**	**35.9**	**59.0**	25.5	1.4	0	0	0	0
T	3.8	**76.4**	4.7	**79.2**	**94.3**	1.0	1.4	2.8	0	16.5	**49.5**	**63.2**	16.5	**71.2**	**98.1**
Consensus sequence 1[b]	**A**	**T**	**A**	**T**	**T**	**C**	**G**	**C**	**G**	**A**	**W**	**W**	**A**	**T**	**T**
Consensus sequence 2[b]	**A**	**T**	**A**	**T**	**T**	**G**	**C**	**G**	**C**	**A**	**W**	**W**	**A**	**T**	**T**

[a]Numbers in this table represent the percentage of the 71 unique clones from the SELEX experiments with the indicated bases at that position. Bold numbers correspond to the consensus nucleotides at each position.
[b]Two consensus sequences are derived from the SELEX analysis where W = A or T.

typical EMSA experiment in which HMGA2 was titrated into 1 nM of ^{32}P-labeled FL-SELEX1, a consensus sequence obtained from our SELEX experiments. The binding constants were obtained by fitting the binding data to equation 1 as described under "Materials and Methods." Our binding curves and binding constants are summarized in Figure 21B and Table 6, respectively. As expected, HMGA2 binds to FL-SELEX1 with the highest affinity. Intriguingly, when the middle GC-rich sequence was mutated to an AT-rich sequence (either 5'-TATA-3' or 5'-AAAA-3'), the DNA binding affinity was reduced 5-fold. These results indicate that the middle GC-rich sequence is required for high-affinity binding. Our results also demonstrated that one 5 or 6 bp AT-rich sequence is not sufficient for high affinity binding of HMGA2 to DNA. Mutation of either AT-rich sequence of FL-SELEX1 to a non-AT-rich sequence significantly decreases the apparent DNA binding constant. These results suggest that all three segments in FL-SELEX1 are critical to high affinity binding. Consistent with our previous results (Cui et al., 2005), HMGA2 also tightly binds to 15 bp A-track and alternate AT sequence although the binding affinity is three-fold lower than that of FL-SELEX1; it does not bind to GC-rich sequences.

5.4.4 Biological implication

The identification of the consensus sequences for HMGA2 is an important step towards characterization of its biological functions *in vivo*. As demonstrated previously, HMGA2 is a transcriptional factor involved in adipocytic proliferation and differentiation during embryogenesis (Zhou et al., 1995; Kools et al., 1995; Li et al., 2006). However, the mechanism by which HMGA2 regulates fat cell proliferation and

differentiation is still unknown. Knowing the DNA binding specificity of HMGA2, it is possible to locate its binding sites within the genome and map its regulatory network in the cell. Furthermore, the discovery of the preferred binding sites of HMGA2 may provide essential information for designing new anti-cancer or anti-obesity drugs. It has been shown that HMGA2 is an oncoprotein associated with many benign and malignant tumors (Young & Narita, 2007); it is also involved in obesity (Anand et al., 2000). These studies suggest that HMGA2 is a potential target for treatment of cancers and obesity. We showed here that HMGA2 strongly and specifically binds to the minor groove of the two AT sites in the SELEX sequences, suggesting that the minor groove binding agents are good inhibitors of HMGA2-DNA interactions. Indeed, we have recently demonstrated that netropsin, a well-known minor groove binder, competitively inhibited HMGA2 binding to the SELEX sequences (Miao et al, 2007). It is possible to design new minor groove binders to alleviate tumor development and progression through inhibition of the important HMGA2-DNA interactions.

5.5 Conclusion

Our results, presented in this article, represent a systematic and unbiased analysis of HMGA2 DNA binding sites under physiologically relevant conditions (200 mM NaCl and 30 μM (bp) of poly(dG-dC)$_2$). Using a SELEX procedure, we have demonstrated that HMGA2 preferentially binds to a type of 15 bp AT-rich DNA sequence: the first 5 bp are AT-rich, the middle 4 to 5 bp are GC-rich, and the last 5 to 6 bp are AT-rich. Alignment of 71 unique SELEX sequences revealed two consensus sequences for HMGA2: 5'-ATATTCGCGAWWATT-3' or 5'-ATATTGCGCAWWATT-3' where W

= A or T. Our quantitative EMSA assays showed that the three segments in the consensus sequences (two AT-rich segments and one GC-rich segment) are required for high affinity binding; mutations of these sequences significantly reduced the DNA-binding affinity of HMGA2. These results indicate that HMGA2 does not randomly recognize any AT-rich sequences. In contrast, it binds to specific AT-rich DNA sequences. We also showed that HMGA2 binds to AT-rich sites as a monomer under physiologically relevant salt conditions.

Table 6. DNA binding constants for HMGA2 binding to different DNA oligomers

DNA oligomers	Sequence (top strand)[a]	K_{app} (M^{-1}) [b]
FL-SELEX1	5′-ATATTCGCGATTATT-3′	$7.45 \pm 0.77 \times 10^6$
FL-304	5′-TATATATATATATAT-3′	$2.16 \pm 0.14 \times 10^6$
FL-AT15	5′-AAAAAAAAAAAAAAA-3′	$2.19 \pm 0.24 \times 10^6$
FL-308	5′-ATATTTATATTTATT-3′	$1.44 \pm 0.27 \times 10^6$
FL-306	5′-ATATTAAAAATTATT-3′	$1.46 \pm 0.40 \times 10^6$
FL-315	5′-ATATTCGCGACTGTC-3′	$<<1.0 \times 10^5$
FL-317	5′-GTGTCCGCGATTATT-3′	$<<1.0 \times 10^5$
FL-GC	5′-GGGGGGGGGGGGGGG-3′	0

[a]All DNA oligomers are 59 bp double strand DNA oligonucleotides with the following sequence,
5'-CATGGTACCTTCAGAGGCTCGAGN_{15}GCTGACTGGCATGCAAGCTG-3', where N_{15} represents the 15 bp DNA sequence listed in the table. Only top strands are shown.
[b]The apparent DNA-binding constants of HMGA2 were measured as decribed under "Materials and Methods." The values are the average of at least three independent determinations.

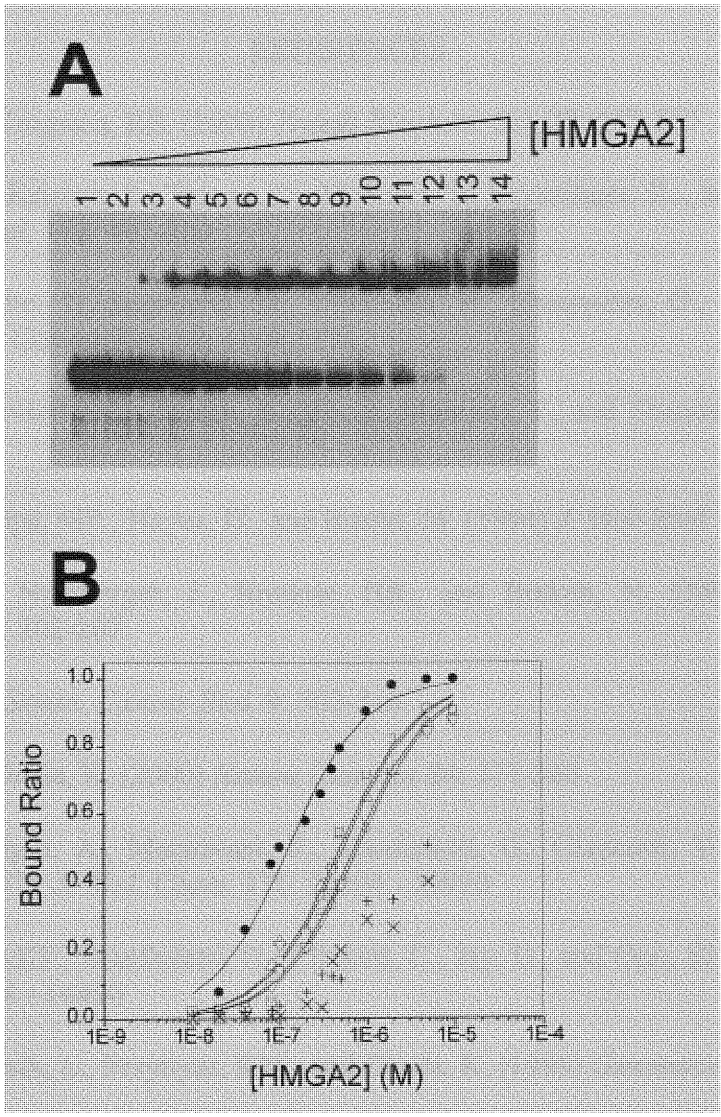


Figure 21. Quantitative EMSA experiments to determine the DNA binding constants of HMGA2 binding to different DNA oligomers. (A) Binding of HMGA2 to the DNA oligomer FL-SELEX1. 1 nM of ^{32}P-labeled FL-SELEX1 was incubated with increasing concentrations of HMGA2 in 50 µl of 1 × EMSA binding buffer containing 20 mM Tris-HCl (pH 8.0), 200 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 0.5 mM $MgCl_2$, 5 % glycerol and 1.5 µM (bp) poly(dG-dC)$_2$. EMSA experiments were performed as described under "Materials and Methods." The autoradiogram of the ^{32}P-labeled FL-SELEX1 was shown. The radioactivities were quantified with a PhosphorImager. Lane 1 is the free FL-SELEX1. In addition to FL-SELEX1, lanes 2 to 14 also contain 10, 20, 40, 80, 100, 200, 300, 400, 500, 1000, 2000, 5000, and 10000 nM of HMGA2, respectively. (B) Quantification analysis of the binding data from the EMSA experiments. The bound ratio of DNA was plotted against the protein concentration. The curves are generated from fitting the data to Equation 1 as described under "Materials and Methods." Symbols: ●, FL-SELEX1; △, FL 300; □, FL 304; ○, FL 306; ▽, FL 308; ×, FL 315; +, FL317.

REFERENCES

Aalfsa, JD., and Kingston, RE. (2000) *Trends Biochem. Sci.* **25**: 548-555

Abdulkadir, SA., Krishna, S., Thanos, D., Maniatis, T., Strominger, JL., and Ono, SJ. (1995) *J. Exp. Med.* **182**: 487-500

Abe, N., Watanabe, T., Sugiyama, M., Uchimura, H., Chiappetta, G., Fusco, A., and Atomi, Y. (1999) *Cancer Res.* **59**: 1169-1174

Agalioti, T., Lomvardas, S., Parekh, B., Yie, J., Maniatis, T., and Thanos D. (2000) *Cell* **103**: 667-678

Alfonso, PJ., Crippa, MP., Hayes, JJ., and Bustin, M. (1994) *J. Mol. Biol.* **236**: 189-198

Allemandou, F., Nussberger, J., Brunner, HR., and Brakch, N. (2003) *J. Biomed. Biotechnol.* **2003**: 202-207

Aman, P. (1999) *Semin. Cancer Biol.* **9**: 303-318

Anand, A., and Chada, K. (2000) *Nature Genetics* **24**: 377-380

Arlotta, P., Rustighi, A., Mantovani, F., Manfioletti, G., Giancotti, V., Tell, G., and Damante, G. (1997) *J. Biol. Chem.* **272**: 29904 - 29910

Arlotta, P., Tai, AKF., Manfioletti, G., Clifford, C., Jay, G., and Ono, SJ. (2000) *J. Biol. Chem.* **275**: 14394 - 14400

Ashar, HR., Cherath, L., Przybysz, KM., and Chada, K. (1996) *Genomics* **31**: 207-214

Asher, HR., Fejzo, MS., Tkachenko, A., Zhou, X., Fletcher, JA., Weremowicz, S., Morton, CC., and Chada, K. (1995) *Cell* **82**: 57-65

Ayoubi, TAY., Jansen, E., Meulemans, SMP., and Van de Ven, WJM. (1999) *Oncogene* **18**: 5076-5087

Bailey, TL., and Gribskov, M. (1998) *Bioinformatics* **14**: 48-54

Bailey, TL., Williams, N., Misleh, C., and Li, WW. (2006) *Nucleic Acids Res.* **34**: W369-W373

Baldassarre, G., Battista, S., Belletti, B., Thakur, S., Pentimalli, F., Trapasso, F., Fedele, M., Pierantoni, G., Croce, CM., and Fusco, A. (2003) *Mol. Cell. Biol.* **23**: 2225-2238

Baldassarre, G., Fedele, M., Battista, S., Vecchione, A., Klein-Szanto, AJP., Santoro, M., Waldmann, TA., Azimi, N., Croce, CM., and Fusco, A. (2001) *Proc. Natl. Acad. Sci. U. S. A* **98**: 7970-7975

Ballin, JD., Shkel, IA., and Record, MT. (2004) *Nucleic Acids Res.* **32**: 3271-3281

Banks, GC., Li, Y., and Reeves, R. (2000) *Biochemistry* **39**: 8333-8346

Bannister, AJ., and Kouzarides, T. (1996) *Nature* **384**, 641-643

Battista, S., Fidanza, V., Fedele, M., Klein-Szanto, AJP., Outwater, E., Brunner, H., Santoro, M., Croce, CM., and Fusco, A. (1999) *Cancer Res.* **59**: 4793-4797

Beaujean, N., Bouniol-Baly, C., Monod, C, Kissa, K., Jullien, D., and Aulner, N. (2000) *Dev. Biol.* **221**: 337-354

Berlingieri, MT., Manfioletti, G., Santoro, M., Bandiera, A., Visconti, R., Giancotti, V., and Fusco, A. (1995) *Mol. Cell. Biol.* **15**: 1545-1553

Berlingieri, MT., Pierantoni, GM., Giancotti, V., Santoro, M., and Fusco, A. (2002) *Oncogene* **21**: 2971-2980

Bhugra, B., Smolarek, TA., Lynch, RA., Meloni, AM., Sandberg, AA., Deaven, L., and Menon, AG. (1998) *Cancer Lett.* **126**: 119-126

Bianchi, ME., and Agresti, A. (2005) *Curr. Opin. Genet. Dev.* **15**: 496-506

Boger, DL., Fink, BE., Brunette, SR., Tse, WC., and Hedrick, MP. (2001) *J. Am. Chem. Soc.* **123**: 5878-5891

Bol, S., Wanschura, S., Thode, B., Deichert, U., Van de Ven, WJM., Bartnitzke, S., and Bullerdi, J. (1996) *Cancer Genet. Cytogenet.* **90**: 88-90

Bonnefoy, E., Bandu, MT., and Doly, J. (1999) *Mol. Cell. Biol.* **19**: 2803-2816

Boring, CC., Squires, TS., and Tong, T. (1991) *Cancer Statistics* **41**: 19-36

Borrmann, L., Schwanbeck, R., Heyduk, T., Seebeck, B., Rogalla, P., Bullerdiek, J., and Wisniewski, JR. (2003) *Nucleic Acids Res.* **31**: 6841-6851

Borrmann, L., Wilkening, S., and Bullerdiek, J. (2001) *Oncogene* **20**: 4537-4541

Breslauer, KJ., Remeta, DP., Chou, WY., Ferrante, R., Curry, J., Zaunczkowski, D., Snyder, JG., and Marky, LA. (1987) *Proc. Natl. Acad. Sci. U. S. A* **84**: 8922-8926

Breslauer, KJ, and Sturtevant, JM. (1977) *Biophys. Chem.* **7**: 205-209

Bruhn, SL., Pil, PM., Essigmann, JM., Housman, DE., and Lippard, SJ. (1992) *Proc. Natl. Acad. Sci. U. S. A* **89**: 2307-2311

Bussemakers, MJG., van de Ven, WJM., Debruyne, FMJ., and Schalken, JA. (1991) *Cancer Res.* **51**: 606-611

Bustin, M. (1999) *Mol. Cell. Biol.* **19**: 5237-5246

Bustin, M. (2001) *Trends Biochem. Sci.* **26**: 152-153

Bustin, M., and Reeves, R. (1996) *Prog. Nucleic Acid Res. Mol. Biol.* **54**: 35-100

Cantor, CR., and Schimmel, PR. (1980) *Biophysical Chemistry*. Freeman, WH., NY

Carey, J. (1988) *Proc. Natl. Acad. Sci. U. S. A* **85**: 975-979

Catez, F., Yang, H., Tracey, KJ., Reeves, R., Misteli, T., and Bustin, M. (2004) *Mol. Cell. Biol.* **24**: 4321-4328

Chalikian, TV., Plum, GE., Sarvazyan, AP., and Breslaver, KJ. (1994) *Biochemistry* **33**: 8629-8640

Chan, HM., and La Thangue, NB. (2001) *J. Cell Sci.* **114**: 2363-2373

Chang, ZG., Yang, LY., Wang, W., Peng, JX., Huang, GW., Tao, YM., and Ding, X. (2005) *Dig. Dis. Sci.* **50**: 1764-1770

Chau, KY., Arlotta, P., Patel, UA., Crane-Robinson, C., Manfioletti, G., and Ono, SJ. (1999) *FEBS Lett.* **457**: 429-436

Chau, KY., Patel, UA., Lee, K., Lam, HY., and Crane-Robinson, C. (1995) *Nucleic Acids Res.* **23**: 4262–4266

Chellappan, SP., Hiebert, S., Mudryj, M., Horowitz, JM., and Nevins, JR. (1991) *Cell* **65**: 1053-1061

Chiappetta, G., Bandiera, A., Berlingieri, MT., Visconti, R., Manfioletti, G., Battista, S., Martinez-Tello, FJ., Santoro, M., Giancotti, V., and Fusco, A. (1995) *Oncogene* **10**: 1307-1314

Chiappetta, G., Botti, G., Monaco, M., Pasquinelli, R., Pentimalli, F., Bonito, MD., D'Aiuto, G., Fedele, M., Iuliano, R., Palmieri, EA., Pierantoni, GM., Giancotti, V., and Fusco, A. (2004) *Clin. Cancer Res.* **10**: 7637-7644

Chin, MT., Pellacani, A., Wang, H., Lin, SSJ., Jain, MK., Perrella, MA., and Lee, M. (1998) *J. Biol. Chem.* **273**: 9755-9760

Christy, RJ., Kaestner, KH., Geiman, DE., and Lane, MD. (1991) *Proc. Natl. Acad. Sci. U. S. A* **88**: 2593-2597

Christy, RJ., Yang, VW., Ntambi, JM., Geiman, DE., Landschulz, WH., Friedman, AD., Nakabeppu, Y., Kelly, TJ., and Lane, MD. (1989) *Genes Dev.* **3**: 1323-1335

Cramer, SF., and Patel, A. (1990) *Am. J. Clin. Pathol.* **94**: 435-438

Crothers, DM. (1971) *Biopolymers* **10**: 2147-2160

Cui, T., Joynt, S., Morillo, V., Baez, M., Hua, Z., Wang, X., and Leng, F. (2007) *Protein & Peptide Lett.* **14**: 87-91

Cui, T., Wei, S., Brew, K., and Leng, F. (2005) *J. Mol. Biol.* **352**: 629-645

Cui,Y., Wang, Q., Stormo, GD., and Calvo, JM. (1995) *J. Bacteriol.* **177**: 4872-4880

Das, D., Peterson, RC., and Scovell, WM. (2004) *Mol. Endocrinol.* **18**: 2616-2632

Ding, HF., Bustin, M., and Hansen, U. (1997) *Mol. Cell. Biol.* **17**: 5843-5855

Distel, RJ., Robinson, GS., and Spiegelman, BM. (1992) *J. Biol. Chem.* **267**: 5937-5941

Dragan, AI., Liggins, JR., Crane-Robinson, C., and Privalov, PL. (2003) *J. Mol. Biol.* **327**: 393-411

Du, W., and Maniatis, T. (1992) *Proc. Natl. Acad. Sci. U. S. A* **89**: 2150-2154

Du, W., Thanos, D., and Maniatis, T. (1993) *Cell* **74**: 887-898

Ellington, AD., and Szostak, JW. (1990) *Nature* **346**: 818-822

Evans, A, Lennard, TWJ., Davies, BR. (2004) *J. Surg. Oncol.* **88**: 86-99

Evans, JN., Zajicek, J., Nissen, MS., Munske, G., Smith, V., and Reeves, R. (1995) *Int. J. Pept. Protein Res.* **45**: 554-560

Fajas, L., Paul, C., Vie, A., Estrach, S., Medema, R., Blanchard, JM., Sardet, C., and Vignais, ML. (2001) *Mol. Cell. Biol.* **21**: 2956-2966

Farmer, SR. (2005) *Int. J. Obes.* **29**: suppl. 1: S13-6

Fearon, ER., and Vogelstein, B. (1990) *Cell* **61**: 759-767

Fedele, M., Battista, S., Kenyon, L., Baldassarre, G., Fidanza, V., Klein-Szanto, AJP., Parlow, AF., Visone, R., Pierantoni, GM., Outwater, E., Santoro, M., Croce, CM., and Fusco, A. (2002) *Oncogene* **21**: 3190-3198

Fedele, M., Battista, S., Manfioletti, G., Croce, CM., Giancotti, V., and Fusco, A. (2001) *Carcinogenesis* **22**: 1583-1591

Fedele, M., Berlingieri, MT., Scala, S., Chiariotti, L., Viglietto, G., Rippel, V., Bullerdiek, J., Santoro, M., and Fusco, A. (1998) *Oncogene* **17**: 413-418

Fedele, M., Pierantoni, GM., Berlingieri, MT., Battista, S., Baldassarre, G., Munshi, N., Dentice, M., Thanos, D., Santoro, M., Viglietto, G., and Fusco, A. (2001) *Cancer Res.* **61**: 4583-4590

Fedele, M., Visone, R., Martino, ID., Troncone, G., Palmieri, D., Battista, S., Ciarmiello, A., Pallante, P., Arra, C., Melillo, RM., Helin, K., Croce, CM., and Fusco, A. (2006) *Cancer Cell* **9**: 459-471

Ferrari, S., Harley, VR., Pontiggia, A., Goodfellow, PN., Lovell-Badge, R., and Bianchi, ME. (1992) *EMBO J.* **11**: 4497-4506

Field, SJ., Tsai, FY., Kuo, F., Zubiaga, AM., Kaelin, WG., Livingston, DM., Orkin, SH., and Greenberg, ME. (1996) *Cell* **85**: 549-561

Flohr, AM., Rogalla, P., Bonk, U., Puettmann, B., Buerger, H., Gohla, G., Packeisen, J., Wosniok, W., Loeschke, S., and Bullerdiek, J. (2003) *Histol. Histopathol.* **18**: 999-1004

Foti, D., Iuliano, R., Chiefari, E., and Brunetti, A. (2003) *Mol. Cell. Biol.* **23**: 2720-2732

Frank, O., Schwanbeck, R., and Wisniewski, JR. (1998) *J. Biol. Chem.* **273**: 20015-20020

Friedmann, M., Holth, LT., Zoghbi, HY., and Reeves, R. (1993) *Nucleic Acids Res.* **21**: 4259-4267

Gattas, GJF., Quade, BJ., Nowak, RA., Morton, CC. (1999) *Genes Chromosomes, and Cancer* **25**: 316-322

Gehring, WJ., Affolter, M., and Burglin, T. (1994) *Annu. Rev. Biochem.* **63**: 487-526

Geierstanger, BH., Volkman, BF., Kremer, W., Wemmer, DE. (1994) *Biochemistry* **33**:5347-5355

Geierstanger, BH., and Wemmer, DE. (1995) *Annu. Rev. Biophys. Biomol. Struct.* **24**: 463-493

Geurts, JMW., Schoenmakers, EFPM., Roijer, E., Astrom, AK., Stenman, G., and van de Ven, WJM. (1998) *Oncogene* **16**: 865-872

Geurts, JMW., Schoenmakers, EFPM., Roijer, E., Stenman, G., and Van de Ven, WJM. (1997) *Cancer Res.* **57**: 13-17

Geurts, JMW., Schoenmakers, EFPM., and Van de Ven, WJM. (1997) *Cancer Genet. Cytogenet.* **95**: 198-205

Ghersa, P., Huijsduijnen, RH., Whelan, J., and DeLamarter, JF. (1992) *J. Biol. Chem.* **267**: 19226-19232

Giancotti, V., Bandiera, A., Buratti, E., Fusco, A., Marzari, R., Coles, B., and Goodwin, GH. (1991) *Eur. J. Biochem.* **198**: 211-216

Giancotti, V., Berlingieri, MT., DiFiore, PP., Fusco, A., Vecchio, G., and Crane-Robinson, C. (1985) *Cancer Res.* **45**: 6051-6057

Giancotti, V., Buratti, E., Perissin, L., Zorzet, S., Balmain, A., Portella, G., Fusco, A., and Goodwin, GH. (1989) *Exper. Cell Res.* **184**: 538-545

Giancotti, V., Pani, B., D'Andrea, P., Berlingieri, MT., Fiore, PPD., Fusco, A., Vecchio, G., Philp, R., Crane-Robinson, C., Nicolas, RH., Wright, CA., and Goodwin, GH. (1987) *EMBO J.* **6**: 1981-1987

Gill, SC., and von Hippel, PH. (1989) *Anal. Biochem.* **182**: 319-326

Gonzalez, C., Moore, M., Ribeiro, S., Schmitz, U., Schroth, GP., Turin, L., and Bruice, TW. (2001) *Nucleic Acids Res.* **29**: e85

Goodwin, G. (1998) *Int. J. Biochem. & Cell Biol.* **30**: 761-766

Goodwin, GH., Cockerill, PN., Kellam, S., and Wright, CA. (1985) *Eur. J. Biochem.* **149**: 47-51

Green, MC. (1989) *Genetic Variants and Strains of the Laboratory Mouse.* Oxford University Press, pp 12-403

Grosschedl, R., Giese, K., and Pagel, J. (1994) *Trends in Genetics* **10**: 94-100
Haq, I. (2002) *Arch. Biochem. Biophys.* **403**: 1-15

Haq, I., Chowdhry, BZ., and Chaires, JB. (1997) *Eur. Biophys. J.* **26**: 419-426

Henderson, A., Bunce, M., Siddon, N., Reeves, R., and Tremethick, DJ. (2000) *J. Virol.* **74**: 10523-10534

Hennig, Y., Rogalla, P., Wanschura, S., Frey, G., Deichert, U., Bartnitzke, S., and Bullerdiek, J. (1997) *Cancer Genet. Cytogenet.* **96**: 129-133

Herrera, JE., and Chaires, JB. (1989) *Biochemistry* **28**: 1993-2000

Herrera, JE., Sakaguchi, K., Bergel, M., Trieschmann, L., Nakatani, Y., and Bustin, M. (1999) *Mol. Cell. Biol.* **19**: 3466-3473

Hess, JL. (1998) *Amer. J. Clin. Pathol.* **109**: 251-261

Hock, R., Furusawa, T., Ueda, T., and Bustin, M. (2007) *Trends Cell Biol.* **17**: 72-79

Holbrook, JA., Tsodikov, OV., Saecker, RM., and Record, MT. (2001) *J. Mol. Biol.* **310**: 379-401

Holmes, JK., and Solomon, MJ. (1996) *J. Biol. Chem.* **271**: 25240-25246

Huth, JR., Bewley, CA., Nissen, MS., Evans, JNS., Reeves, R., Gronenborn, AM., and Clore, GM. (1997) *Nature Stru. Biol.* **4**: 657-665

Hwang, CS., Mandrup, S., MacDougald, OA., Geiman, DE., and Lane, MD. (1996) *Proc. Natl. Acad. Sci. U. S. A* **93**: 873-877

Jen-Jacobson, L., Engler, LE., and Jacobson, LA. (2000) *Struct. Fold. Des.* **8**: 1015-1023

Jin, R., and Breslauer, KJ. (1988) *Proc. Natl. Acad. Sci. U. S. A* **85**: 8939-8942

John, S., Reeves, RB., Lin, JX., Child, R., Leiden, JM., Thompson, CB., and Leonard, WJ. (1995) *Mol. Cell. Biol.* **15**: 1786-1796

Johnson, KR., Lehn, DA., Elton, TS., Barr, PJ., and Reeves, R. (1988) *J. Biol. Chem.* **263**: 18338-18342

Johnson, KR., Lehn, DA., and Reeves, R. (1989) *Mol. Cell. Biol.* **9**: 2114-2123

Johnson, PR., and Hirsch, J. (1972) *J. Lipid Res.* **13**: 2-11

Kaufman, TC., Seeger, MA., and Olsen, G. (1990) *Adv. Genet.* **27**: 309-362

Kayser, K., Dunnwald, D., Kazmierczak, B., Bullerdiek, J., Kaltner, H., Zick, Y. (2003) *Pathol. Res. Pract.* **199**: 589-598

Kazmierczak, B., Hennig, Y., Wanschura, S., Rogalla, P., Bartnitzke, S., Van de Ven, W., and Bullerdiek, J. (1995) *Cancer Res.* **55**: 6038-6039

Kazmierczak, B., Rosigkeit, J., Wanschura, S., Meyer-Bolte, K., Van de Ven, WJM., Kayser, K. (1996) *Oncogene* **12**: 515-521

Kazmierczak, B., Wanschura, S., Rosigkeit, J., Meyer-Bolte, K., Uschinsky, K., Haupt, R., Schoenmakers, EFPM., Bartnitzke, S., Van de Ven, WJM., and Bullerdiek, J. (1995) *Cancer Res.* **55**: 2497-2499

Keshavarz, H, Hillis, SD., Kieke, BA., and Marchbanks. PA. (2002) *Hysterectomy Surveillance*-United State, 1994-1999. pp1-8. Centers for Disease Control and Prevention.

Klotzbucher, M., Wasserfall, A., and Fuhrmann, U. (1999) *Am. J. Pathol.* **155**: 1535-1542

Kools, PFJ., and Van de Ven, WJM. (1996) *Cancer Genet. Cytogenet.* **91**: 1-7

Kools, PF., Wanschura, S., Schoenmakers, EF., Geurts, JM., Mols, R., Kazmierczak, B., Bullerdiek, J., Van den, BH., and Van de Ven, WJ. (1995) *Cancer Genet. Cytogenet.* **79**: 1-7

Kopka, ML., Yoon, C., Goodsell, D., Pjura, P., and Dickerson, RE. (1985) *J. Mol. Biol.* **183**: 553-563

Kopka, ML., Yoon, C., Goodsell, D., Pjura, P., and Dickerson, RE (1985) *Proc. Natl. Acad. Sci. U. S. A* **82**: 1376-1380

Kottickal, LV., Sarada, B., Ashar, H., Chada, K, and Nagarajan, L (1998) *Biochem. Biophys. Res. Commun.* **242**: 452-456

Landsman, D., and Bustin, M. (1993) *Bioessays* **15**: 539-546

Lavery, R., and Pullman, B. (1981) *Nucleic Acids Res.* **9**: 3765-3777

Lavery, R., Pullman, B., and Corbin, S. (1981) *Nucleic Acids Res.* **9**: 6539-6552

Ledent, C., Marcotte, A., Dumont, JE., Vassart, G., and Parmentier, M. (1995) *Oncogene* **10**: 1789-1797

Ledermann, B. (2000) *Exp. Physiol.* **85**: 603-613

Lehn, DA., Elton, TS., Johnson, KR., and Reeves, R. (1988) *Biochem. Int.* **16**: 963-971

Leng, F., Priebe, W., and Chaires, JB. (1998) *Biochemistry* **37**: 1743-1753

Lewis, H., Kaszubska, W., DeLamarter, JF., and Whela, J. (1994) *Mol. Cell. Biol.* **14**: 5701-5709

Lewis, RT., Andreucci, A., and Nikolajczyk, BS. (2001) *J. Biol. Chem.* **276**: 9550-9557

Li, O., Vasudevan, D., Davey, CA., and Droge, P. (2006) *Genesis* **44**: 523-529

Lim, JH., West, KL., Rubinstein, Y., Bergel, M., Postnikov, YV., and Bustin, M. (2005) *EMBO J.* **24**: 3038-3048

Lohman, TM., and Overman, LB. (1985) *J. Biol. Chem.* **260**: 3594-3603

Lund, T., Holtlund, J., Fredriksen, M., and Laland, SG. (1983) *FEBS Lett.* **152**: 163-167

Magnaghi-Jaulin, L., Groisman, R., Naguibneva, I., Robin, P., Lorain, S., Le Villain, JP., Troalen, F., Trouche, D., and Harel-Bellan, A. (1998) *Nature* **391**: 601-605

Maher, J., and Nathans, D. (1996) *Proc. Natl. Acad. Sci. U. S. A* **93**: 6716-6720

Manfioletti, G., Giancotti, V., Bandiera, A., Buratti, E., Sautieere, P., Cary, P., Crane-Robinson, C., Coles, B., and Goodwin, GH. (1991) *Nucleic Acids Res.* **19**: 6793-6797

Manfioletti, G., Rustighi, A., Mantovani, F., Goodwin, GH., and Giancotti, V. (1995) *Gene* **167**: 249-253

Manning, GS. (1978) *Quart. Rev. Biophys.* **11**: 179-246

Mantovani, F., Covaceuszach, S., Rustighi, A., Sgarra, R., Heath, C., Goodwin, GH., and Manfioletti, G. (1998) *Nucleic Acids Res.* **26**: 1433-1439

Marky, LA., and Breslauer, KJ. (1987) *Proc. Natl. Acad. Sci. U. S. A* **84**: 4359-4363

McGhee, JD. (1976) *Biopolymers* **15**: 1345-1375

McLauchlan, J., Gaffney, D., Whitton, JL., and Clements, JB. (1985) *Nucleic Acids Res.* **13**: 1347-1368

Melillo, RM., Pierantoni, GM., Scala, S., Battista, S., Fedele, M., Stella, A., De Biasio, MC., Chiappetta, G., Fidanza, V., Condorelli, G., Santoro, M., Croce, CM., Viglietto, G., and Fusco, A. (2001) *Mol. Cell. Biol.* **21**: 2485-2495

Merika, M., Williams, AJ., Chen, G., Collins, T., and Thanos, D. (1998) *Mol. Cell* **1**: 277-287

Merscher, S., Marondel, I., Pedeutour, F., Gaudray, P., Kucherlapati, R., and Turc-Carel, C. (1997) *Genomics* **46**: 70-77

Meyer, B., Loeschke, S., Schultze, A., Weigel, T., Sandkamp, M., Goldmann, T., Vollmer, E., and Bullerdiek, J. (2007) *Mol. Carcinog.* Epub ahead of print

Micci, F., Panagopoulos, I., Bjerkehagen, B., and Heim, S. (2006) *Virchows Arch.* **448**: 838-842

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, PA., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, LM., Ding, W. (1994) *Science* **266**: 66-71

Misra, VK., Sharp, KA., Friedman, RA., and Honig, B. (1994) *J. Mol. Biol.* **238**: 245-263

Miyazawa, J., Mitoro, A., Kawashiri, S., Chada, KK., and Imai, K. (2004) *Cancer Res.* **64**: 2024-2029

Morrison, RF., and Farmer, SR. (2000) *J. Nutr.* **130**: 3116s-3121s

Muller, H., Bracken, AP., Vernell, R., Moroni, MC., Christians, F., Grassilli, E., Prosperini, E., Vigo, E., Oliner, JD., and Helin, K. (2001) *Genes Dev.* **15**: 267-285

Munshi, N., Merika, M., Yie, J., Senger, K., Chen, G., and Thanos, D. (1998) *Mol. Cell* **2**: 457-467

Neidle, S. (2001) *Nature Prod. Rep.* **18**: 291-309

Nilsson, M., Panagopoulos, I., Mertens, F., Mandahl, N. (2005) *Virchows Arch.* **447**: 855-858

Nissen, MS., Langan, TA., and Reeves, R. (1991) *J. Biol. Chem.* **266**: 19945-19952

Onate, SA., Prendergast, P., Wagner, JP., Nissen, M., Reeves, R., Pettijohn, DE., and Edwards, DP. (1994) *Mol. Cell. Biol.* **14**: 3376-3391

Palvimo, J., and Linnala-Kankkunen, A. (1989) *FEBS Lett.* **257**: 101-104
Patel, UA., Bandiera, A., Manfioletti, G., Giancotti, V., Chau, KY., and Cranerobinson, C. (1994) *Biochem. Biophys. Res. Commu.* **201**: 63-70

Paton, AE., Wilkinson-Singley, E., and Olins, DE. (1983) *J. Biol. Chem.* **258**: 13221-13229

Pelleymounter, MA., Cullen, MJ., Baker, MB., Hecht, R., Winters, D., Boone, T., and Collins, F. (1995) *Science* **269**: 540-543

Petit, MMR., Mols, R., Schoenmakers, EFPM., Mandahl, N., and Van de Ven, WJM. (1996) *Genomics* **36**: 118-129

Petit, MMR., Schoenmakers, EFPM., Huysmans, C., Geurts, JMW., Mandahl, N., and Van de Ven, WJM. (1999) *Genomics* **57**: 438-441

Piekielko, A., Drung, A., Rogalla, P., Schwanbeck, R., Heyduk, T., Gerharz, M., Bullerdiek, J., and Wisniewski, JR. (2001) *J. Biol. Chem.* **276**: 1984-1992

Pjura, PE., Grzeskowiak, K., and Dickerson, RE. (1987) *J. Mol. Biol.* **197**: 257-271

Pober, JS., and Cotran, RC. (1990) *Physiol. Rev.* **70**: 427-451

Postnikov, YV, Belova, GI., Lim, JH., and Bustin, M. (2006) *Biochemistry* **45**: 15092-15099

Postnikov, YV., Trieschmann, L., Rickers, A., and Bustin, M. (1995) *J. Mol. Biol.* **252**: 423-432

Pugh, BF., and Tjian, R. (1991) *Genes Dev.* **5**: 1935-1945

Read, CM., Cary, PD., Crane-Robinson, C., Driscoll, PC., and Norman, DG. (1993) *Nucleic Acids Res.* **21**: 3427 - 3436

Record, MT., Anderson, CF., and Lohman, TM. (1978) *Quart. Rev. Biophys.* **11**: 103-178

Reeves, R. (2000) *Environ. Health Perspect.* **108** Suppl 5: 803-809

Reeves, R. (2001) *Gene* **277**: 63-81

Reeves, R. (2003) *Biochem. Cell Biol.* **81**: 185-195

Reeves, R. (2004) *Methods Enzymol.* **375**: 297-322

Reeves, R., and Beckerbauer, LM. (2001) *Biochim. Biophys. Acta* **1519**: 13-29

Reeves, R., and Beckerbauer, LM. (2003) *Prog. Cell Cycle Res.* **5**: 279-286

Reeves, R., Langan, TA., and Nissen, MS. (1991) *Proc. Natl. Acad. Sci. U. S. A* **88**: 1671-1675

Reeves, R., Leonard, WJ., and Nissen, MS. (2000) *Mol. Cell. Biol.* **20**: 4666-4679

Reeves, R., and Nissen, MS. (1990) *J. Biol. Chem.* **265**: 8573-8582

Reeves, R., and Nissen, MS. (1993) *J. Biol. Chem.* **268**: 21137-21146

Reeves, R., and Nissen, MS. (1995) *Prog. Cell Cycle Res.* **1**: 339-349

Ribon, V., Johnson, JH., Camp, HS., and Saltiel, AR. (1998) *Proc. Natl. Acad. Sci. U. S. A* **95**: 14751-14756

Ribon, V., Printen, JA., Hoffman, NG., Kay, BK., and Saltiel, AR. (1998) *Mol. Cell. Biol.* **18**: 872-879

Robertson, DL., and Joyce, GF. (1990) *Nature* **344**:467-468

Rogalla, P., Drechsler, K., Schroder-Babo, W., Eberhardt, K., and Bullerdiek, J. (1998) *Anticancer Res.* **18**: 3327-3330

Rogalla, P., Kazmierczak, B., Meyer-Bolte, K., Tran, KH., Bullerdiek, J. (1998) *Genes, Chromosomes and Cancer* **22**: 100-104

Romine, LE., Wood, JR., Lamia, LA., Prendergast, P., Edwards, DP., and Nardulli, AM. (1998) *Mol. Endocrinol.* **12**: 664 - 674

Rommel, B., Rogalla, P., Jox, A., Kalle, CV., Kazmierczak, B., Wolf, J., and Bullerdiek, J. (1997) *Leuk. Lymphoma* **26**: 603-607

Rosenblatt, J., Gu, Y., and Morgan, DO. (1992) *Proc. Natl. Acad. Sci. U. S. A* **89**: 2824-2828

Rustighi, A., Mantovani, F., Fusco, A., Giancotti, V., and Manfiole, G. (1999) *Biochem. Biophys. Res. Comm.* **265**: 439-447

Sarhadi, VK., Wikman, H., Salmenkivi, K., Kuosma, E., Sioris, T., Salo, J., Karjalainen, A., Knuutila, S., Anttila, S. (2006) *J. Pathol.* **209**: 206-212

Scala, S., Portella, G., Fedele, M., Chiappetta, G., and Fusco, A. (2000) *Proc. Natl. Acad. Sci. U. S. A* **97**: 4256-4261

Scala, S., Portella, G., Vitagliano, D., Ledent, C., Chiappetta, G., Giancotti, V., Dumont, J., and Fusco, A. (2001) *Carcinogenesis* **22**: 251-256

Schneider, TD., and Stephens, RM. (1990) *Nucleic Acids Res.* **18**: 6097-6100

Schoenmakers, EFPM, Huysmans, C., and Van de Ven, WJM. (1999) *Cancer Res.* **59**: 19-23

Schoenmakers, EFPM., Wanschura, S., Mols, R., Bullerdiek, J., Van den Berghe, H., and Van de Ven, WJM. (1995) *Nature Genet.* **10**: 436-444

Schwanbeck, R., Gymnopoulos, M., Petry, I., Piekielko, A., Szewczuk, Z., Heyduk, T., Zechel, K., and Wisniewski, JR. (2001) *J. Biol. Chem.* **276**: 26012-26021

Schwanbeck, R., Manfioletti, G., and Wisniewski, JR. (2000) *J. Biol. Chem.* **275**: 1793-1801

Seville, LL., Shah, N., Westwell, AD., and Chan, WC. (2005) *Curr. Cancer Drug Targets* **5**: 159-170

Sgarra, R., Diana, F., Bellarosa, C., Dekleva, V., Rustighi, A., Toller, M., Manfioletti, G., and Giancotti, V. (2003) *Biochemistry* **42**: 3575-3585

Sgarra, R., Diana, F., Rustighi, A., Manfioletti, G., and Giancotti, V. (2003) *Cell Death Differ.* **10**: 386-389

Sgarra, R., Rustighi, A., Tessari, MA., Bernardo, JD., Altamura, S., Fusco, A., Manfioletti, G., and Giancotti, V. (2004) *FEBS Lett.* **574**: 1-8

Solomon, MJ., Strauss, F., and Varshavsky, A. (1986) *Proc. Natl. Acad. Sci. U. S. A* **83**: 1276-1280

Soriano, P., Montgomery, C., Geske, R., and Bradley, A. (1991) *Cell* **64**: 693-702

Spolar, RS., and Record, MT. (1994) *Science* **263**: 777-784

Szewczuk, Z., Petry, I., Schwanbeck, R., and Renner, U. (1999) *J. Biol. Chem.* **274**: 20116-20122

Tallini, G., and Dal Cin, P. (1999) *Advan. Anat. Pathol.* **6**: 237-246

Tamimi, Y., van der Poel, HG., Denyn, MM., Umbas, R., Karthaus, HFM., Debruyne, FMJ., and Schalken, JA. (1993) *Cancer Res.* **53**: 5512-5516

Tamimi, Y., van der Poel, HG., Karthaus, HF., Debruyne, FM., and Schalken, JA. (1996) *Br. J. Cancer* **74**: 573-578

Tessari, MA., Gostissa, M., Altamura, S., Sgarra, R., Rustighi, A., Salvagno, C., Caretti, G., Imbriano, C., Mantovani, R., Sal, GD., Giancotti, V., and Manfioletti, G. (2003) *Mol. Cell. Biol.* **23**: 9104-9116

Thanos, D., and Maniatis, T. (1992) *Cell* **71**: 777-789

Thomas, JO. (2001) *Biochem. Soc. Trans.* **29**: 395-401

Thompson, JD., Gibson, TJ., Plewniak, F., Jeanmougin, F., and Higgins, DG. (1997) *Nucleic Acids Res.* **25**: 4876-4882

Thompson, JD., Higgins, DG., and Gibson, TJ. (1994) *Nucleic Acids Res.* **22**: 4673-4680

Thompson, ME., Jensen, RA., Obermiller, PS., Page, DL., and Holt, JT. (1995) *Nat. Genet.* **9**: 444-450

Trieschmann, L., Postnikov, YV., Rickers, A., and Bustin, M. (1995) *Mol. Cell. Biol.* **15**: 6663-6669

Tuerk, C., and Gold, L. (1990) *Science* **249**: 505-510

Vallone, D., Battista, S., Pierantoni, GM., Fedele, M., Casalino, L., Santoro, M., Viglietto, G., Fusco, A., and Verde, P. (1997) *EMBO J.* **16**: 5310–5321

Vigo, E., Muller, H., Prosperini, E., Hateboer, G., Cartwright, P., Moroni, MC., and Helin, K. (1999) *Mol. Cell. Biol.* **19**: 6379-6395

Visvanathan, KV., and Goodbourn, SEY. (1989) *EMBO J.* **8**: 1129-1138

Walker, JM., Goodwin, GH., and Johns, EW. (1979) *FEBS Lett.* **100**: 394-398

Wang, DZ., Ray, P., and Boothby, M. (1995) *J. Biol. Chem.* **270**: 22924-22932

Wang, DZ., Zamorano, J., Keegan, AD., and Boothby, M. (1997) *J. Biol. Chem.* **272**: 25083-25090

Wanschura, S., Kazmierczak, B., Pohnke, Y., Meyer-Bolte, K., Bartnitzke, S., Van de Ven, WJM., and Bullerdiek, J. (1996) *Cancer Lett.* **102**: 17-21

Weir, HM., Kraulis, PJ., Hill, CS., Raine, AR., Laue, ED., and Thomas, JO. (1993) *EMBO J.* **12**: 1311-1319

Whitley, MZ., Thanos, D., Read, MA., Maniatis, T., and Collins, T. (1994) *Mol. Cell. Biol.* **14**: 6464-6475

Widlak, P., Pietrowska, M., and Lanuszewska, J. (2006) *Histochem. Cell Biol.* **125**: 119-126

Wolffe, AP. (1994) *Science* **264**: 1100-1101

Wolffe, AP. (1998) *Chromatin Structure and Function*, Academic Press, NY

Xiao, B., Spencer, J., Clements, A., Ali-Khan, N., Mittnacht, S., Broceno, C., Burghammer, M., Perrakis, A., Marmorstein, R., and Gamblin, SJ. (2003) *Proc. Natl. Acad. Sci. U. S. A* **100**: 2363-2368

Xiao, DM., Pak, JH., Wang, X., Sato, T., Huang, FL., Chen, HC., and Huang, KP. (2000) *J. Neurochem.* **74**: 392-399

Xu, N., Chen, CY., and Shyu, AB. (1997) *Mol. Cell. Biol.* **17**: 4611-4621

Yie, J., Merika, M., Munshi, N., Chen, G., and Thanos, D. (1999) *EMBO J.* **18**: 3074-3089

Young, AR., and Narita, M. (2007) *Genes Dev.* **21**: 1005-1009

Zhang, Y., Proenca, R., Maffei, M., Barone, M., Leopold, L., and Friedman, JM. (1994) *Nature* **372**: 425-431

Zhao, K., Kas, E., Gonzalez, E., and Laemmli, UK. (1993) *EMBO J.* **12**: 3237-3247

Zhou, X., Benson, KF., Ashar, HR., and Chada, K. (1995) *Nature* **376**: 771-774

Zhou, X., Benson, KF., Przybysz, K., Liu, J., Hou, Y., Cherath, L., and Chada, K. (1996) *Nucleic Acids Res.* **24**: 4071-4077

Zimmer, C., and Wahnert, U. (1986) *Prog. Biophys. Mol. Biol.* 47: 31-112

Appendix 1

A1. Supplemental Materials:

A1.1 Energetics of Binding the Mammalian High Mobility Group Protein HMGA2 to poly(dA-dT)$_2$ and poly(dA)poly(dT)

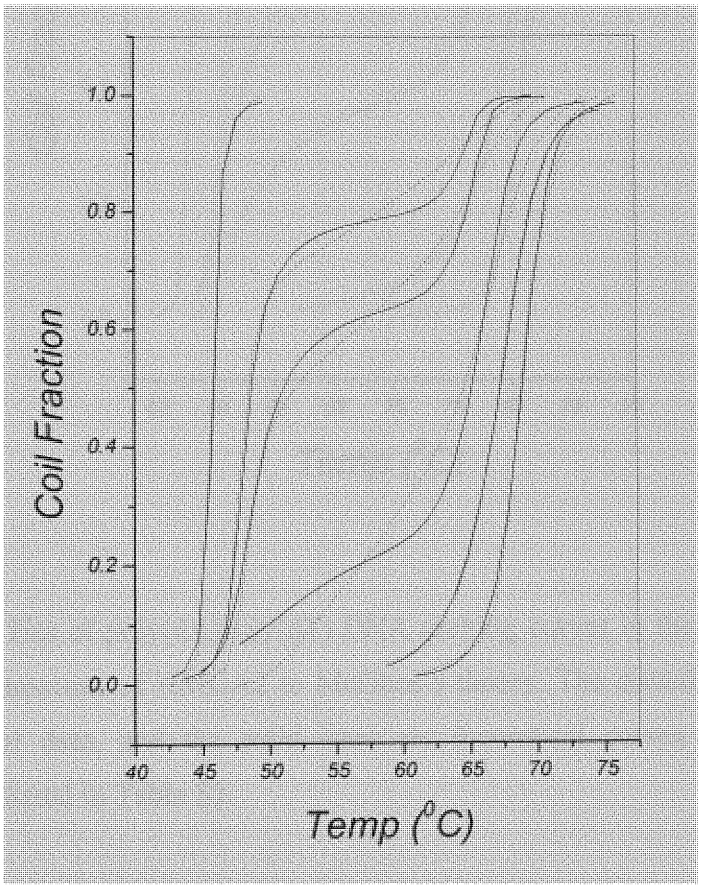


Figure S1. Fits of poly(dA-dT)$_2$ DNA melting curves in the presence of various amounts of HMGA2 to McGhee's theory. The dotted lines are experimental data, and the solid lines are theoretical curves with the parameters listed in Table S1.

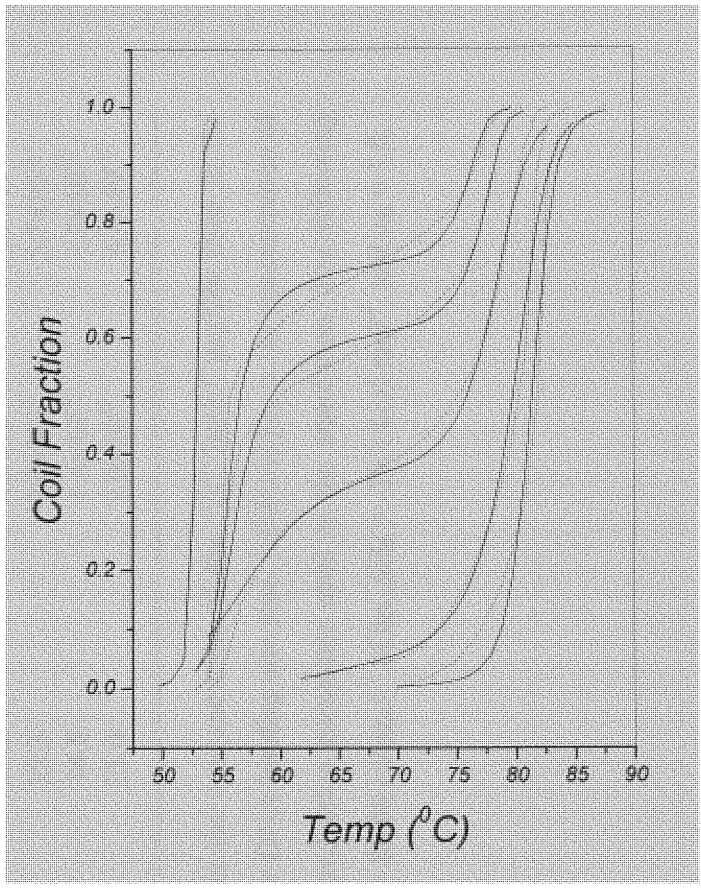


Figure S2. Fits of poly(dA)poly(dT) DNA melting curves in the presence of various amounts of HMGA2 to McGhee's theory. The dotted lines are experimental data, and the solid lines are theoretical curves with the parameters listed in Table S2.

Table S1. Binding parameters from melting of poly(dA-dT)$_2$ in the presence of HMGA2[a]

[DNA] (M bp)	[HMGA2] (M)	K (M^{-1})	n (bp)	$[HMGA2]_{sim}$ (M^{-1})	σ
2.0×10^{-5}	0	0	0	0	1.5×10^{-4}
2.0×10^{-5}	2.5×10^{-7}	2.5×10^{13}	15.2	2.4×10^{-7}	2.2×10^{-4}
2.0×10^{-5}	5.0×10^{-7}	1.0×10^{13}	14.2	4.5×10^{-7}	4.0×10^{-4}
2.0×10^{-5}	1.0×10^{-6}	2.1×10^{13}	15.3	9.0×10^{-7}	5.7×10^{-3}
2.0×10^{-5}	1.5×10^{-6}	2.6×10^{13}	14.8	1.4×10^{-6}	1.2×10^{-2}
2.0×10^{-5}	2.0×10^{-6}	5.3×10^{13}	15.0	2.0×10^{-6}	6.4×10^{-3}

[a][DNA] is the concentration of poly(dA-dT)$_2$. [HMGA2] is the concentration of the "AT hook" peptide used in the experiments. K is the DNA binding constant at 25^{0}C. n is the DNA binding site size. $[HMGA2]_{sim}$ is the concentration of the "AT hook" peptide used in the simulation. σ is the nucleation parameter used in the simulation. The following parameters are constrained and used in the simulation: ΔH_m = 8,400 cal mol^{-1} of bp, ΔH_b = -29.1 kcal mol^{-1}, and ω_h = 1.0.

Table S2. Binding parameters from melting of poly(dA)poly(dT) in the presence of HMGA2[a]

[DNA] (M bp)	[HMGA2] (M)	K (M^{-1})	n (bp)	$[HMGA2]_{sim}$ (M^{-1})	σ
2.0×10^{-5}	0	0	0	0	2.5×10^{-5}
2.0×10^{-5}	2.5×10^{-7}	1.3×10^{12}	15.9	3.1×10^{-7}	2.5×10^{-4}
2.0×10^{-5}	5.0×10^{-7}	9.0×10^{11}	15.0	4.7×10^{-7}	4.0×10^{-4}
2.0×10^{-5}	1.0×10^{-6}	5.9×10^{11}	14.6	7.7×10^{-7}	4.4×10^{-3}
2.0×10^{-5}	1.5×10^{-6}	1.5×10^{12}	14.8	1.2×10^{-6}	4.0×10^{-3}
2.0×10^{-5}	2.0×10^{-6}	2.2×10^{12}	15.1	1.9×10^{-6}	2.8×10^{-3}

[a][DNA] is the concentration of poly(dA)poly(dT). [HMGA2] is the concentration of the "AT hook" peptide used in the experiments. K is the DNA binding constant at 25°C. n is the DNA binding site size. $[HMGA2]_{sim}$ is the concentration of the "AT hook" peptide used in the simulation. σ is the nucleation parameter used in the simulation. The following parameters are constrained and used in the simulation: ΔH_m = 8,900 cal mol^{-1} (bp), ΔH_b = +0.6 kcal mol^{-1}, and ω_h = 1.0

A1.2 Specific Recognition of AT-Rich DNA Sequences by the Mammalian High mobility Group Protein AT-hook 2: A SELEX study

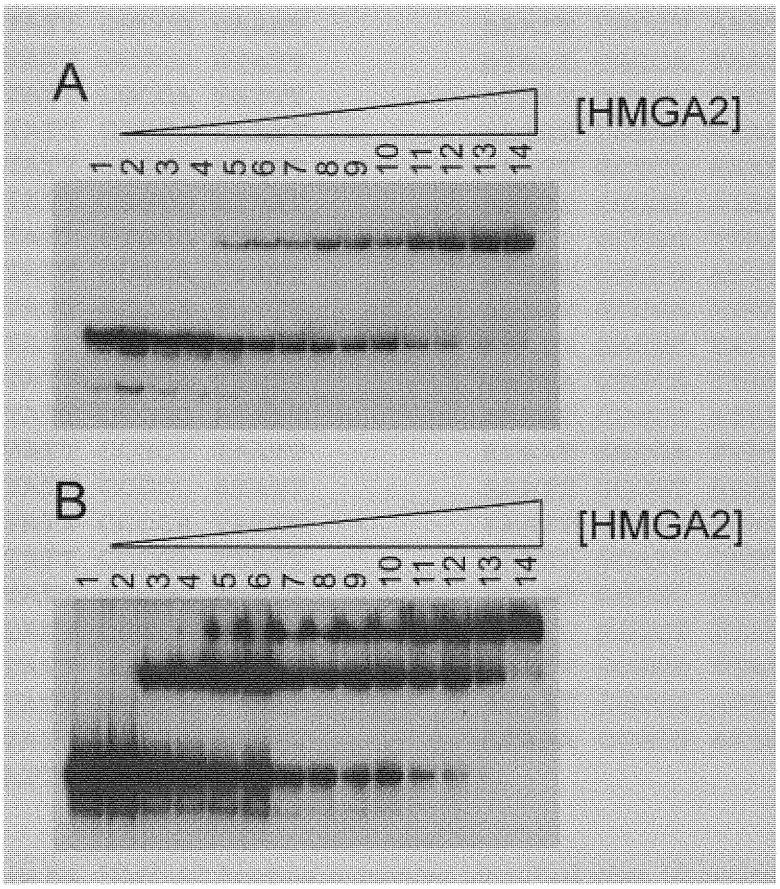


Figure S3. EMSA experiments of HMGA2 binding to FL-AT15 and FL-AT30. 1 nM of ^{32}P-labeled FL-AT15 (A) or FL-AT30 (B) was incubated with increasing concentrations of HMGA2 in 50 µl of 1 × EMSA binding buffer containing 20 mM Tris-HCl (pH 8.0), 200 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 0.5 mM $MgCl_2$, 5 % glycerol and 1.5 µM (bp) poly(dG-dC)$_2$. EMSA experiments were performed as described under "Materials and Methods." The autoradiograms of the ^{32}P-labeled oligonucelotides were shown. The radioactivities were quantified with a PhosphorImager. Lane 1 is the free DNA oligonucleotides FL-AT15 or FL-AT30. In addition to DNA oligonucleotides, lanes 2 to 14 also contain 10, 20, 40, 80, 100, 200, 300, 400, 500, 1000, 2000, 5000, and 10000 nM of HMGA2, respectively.

ATATT CGGCA ATATT
ATATT GCCGA TTATT
ATATT CGCCA ATATT
ATATT GCCGA TAATT
ATATT CGCCG GAAAT
ATATT GCCCC ATATT
ATATT CGCCA TATTT
ATATT GGCGG AAATT
ATATT GGCGG AAATT
ATATT CGGCG ATAAT
ATATT GTGGA ATATC
ATATT CCGCA TTATT
ATATT GCCGA TATTT
ATATT GGCGA TAAAT
ATATT GCGGA TTATT
ATATT CCCGG ATATT
ATATT CCGCT TAATT
ATATT CCGCC ATATT
ATATT CCCCA TATTT
ATATT CCCGA TTATT
ATATT GGCGC ATTTT
ATATT CGCCG TTAAT
ATATT CGCCG ATATT
ATATT CCGCA TAAAT
ATATT GCCGA ATATT
TATTT GCGCA TTAAT

ATAAT CGCGT TAATT
AAATC GCGGA ATAAT
TATTT CGCGT ATTTT
AATTT GCGCA TTATT
AAATT CGCGT AAATT
ATAAT CGCCA TAATT
ATAAT CGGCG TTATT
AATGT CGTGA ATATT
AAATT GCTCG AAAAT
AAATT CGGGA ATAAT
AAAAT GCCGG TTAAT
ATAAT GGCCG ATTTT
ATAAT GCCGA ATATT
AAATT GACGC TAAAT
TATTA CGGCA TTATT
AAAAT CCGCC ATATT
ATAAT GGCGA TTATT
AAATT GCCGT AAATT
AAATT GCGGT ATATT
TAATC GCATT TATTC
ATAAT GGCGC ATAAT
AAATC GGGGT TTATT
AAATT GCGGT TTATT
ATAAT CGCCG ATAAT
ATAAA CCGCG ATTAT

Figure S4. HMGA2 binding sequences identified after 10 cycles of the SELEX experiments. 51 sequences are shown here. See text for details.

Appendix 2

A2. Study the DNA-binding specificity of the AT-hook motif by competition dialysis assay

A2.1 Introduction

In the NMR structure, an individual AT-hook motif of HMGA proteins has been shown to specifically bind to 5 bp AT rich sequence in the DNA minor groove. However, the binding specificity has not been systemically studied. Here, we show how the sequence specificity of AT-hook binding to DNA was explored using competition dialysis assay. Our hypothesis is that if some DNA bind to the AT-hook motif with higher affinities, concentrations of the AT-hook peptide in these DNA samples after dialysis will be higher than those in DNA samples with lower affinities.

A2.2 Materials and Methods

A2.2.1 AT DNA duplexes

Total 16 AT DNA duplexes were used in study of the binding specificity of the AT-hook peptide. Each duplex was prepared by annealing two complementary single-stranded oligonucleotides, which contains a sequence of 5'-GGGGGC$(A/T)_5$CGGGGG-3'. The sequences of the AT region on the top strands were listed in Table S3. Theoretically, all 32 AT sequences with a length of 5 base pairs have been included in the AT region on both top and bottom strands of the 16 AT DNA duplexes. Concentrations of the oligonucleotide were determined using their molar extinction coefficient at 260 nm. In preparation of the duplex, equal amount of each oligonucleotide

was mixed in 1 × BPE plus 34 mM Na^+, heated at 95°C for 15 min, and gradually cooled down to room temperature.

Table S3. The sequences of the 5 bp AT region in the DNA used for studying the binding specificity of the AT-hook peptide

Sequence number	Sequences (5' to 3', top strand)	Sequence number	Sequences (5' to 3', top strand)
1	AAAAA		ATTAA
2	AAAAT	10	TTAAA
3	AAATA	11	ATTTA
4	AATAA	12	AATAT
5	ATAAA	13	ATAAT
6	TAAAA	14	TAATA
7	AAATT	15	TATAA
8	AATTA	16	TATAT

A2.2.2 AT-hook peptide (ATHP)

A peptide, NH_2-Lys-Arg-Pro-Arg-Gly-Arg-Pro-Arg-Lys-Trp-COOH, containing the same amino acid sequence as the third DNA-binding domain of the HMGA proteins was synthesized from Advanced ChemTech and used without further purification.

A2.2.3 Nucleic acids

Clostridium perfringens (CP), *Micrococcus lysodeikticus* (ML), and calf thymus (CT) DNA were purchased from Sigma. The DNA was dissolved in 1× BPE plus 34 mM Na^+ and sonicated on ice 10 times at 200 W each time with 10s intervals. The sonicated sample was further treated by phenol extraction, followed by dialyzing the sample in 1× BPE plus 34 mM Na^+ overnight. The DNA concentration after dialysis was then determined using their molar extinction coefficient at the maximal UV absorbance.

Poly(dA), poly(dT), poly(dA-dT)$_2$, poly(dA)poly(dT), poly(dG-dC)$_2$, and poly(dI-dC)$_2$ were obtained from Amersham Biosciences and used without further purification. The poly(A)poly(U) were purchased from Sigma. The poly(A)poly(dT) DNA-RNA hybrid was prepared by mixing poly(A) with poly(dT) in a 1:1 molar ratio, heating to 95°C for 15 min, and slowly cooling down to room temperature. All DNA polymers were dissolved in 1× BEP plus 34 mM Na^+, and RNA samples were dissolved in the same buffer made by DEPC water. The purity and conformation of all purchased or self-prepared nucleic acids were routinely checked by their UV spectra and melting profile.

A2.2.4 Competition dialysis assay

The competition dialysis assay was schematized as in Figure S5. In the experiment, 0.5 ml of 75 μM (bp) DNA was placed in either a dialysis cassette of 7000 MWCO (Pierce) or a regular dialysis tubing of 8000 MWCO (Spectrum Laboratories Inc.). Couple of DNA with different conformations and sequences was then equilibrated with 200 ml of 1 μM AT-hook peptide solution in 1 × BPE plus 34 mM Na^+ for 7 days. After dialysis, ATHP concentration in each sample was determined using its intrinsic fluorescence from the tryptophan residue. Alternatively, the sample was mixed with 50 μl of fluorescamine solution (0.5 mg/ml in acetone) under vortexing, and the fluorescence intensity from the fluorescamine derivative was used for calculating the ATHP concentration. In the first method for measuring the ATHP concentration, the excitation and emission wavelength was set to 295 nm and 360 nm, respectively. The slit is 2 nm for excitation and 1.8 nm for emission. The ATHP concentration in the sample

was calculated using a standard curve generated from solutions containing known amount of ATHP under the same instrumental conditions. Similarly, in the second method using the fluorescamine derivative, the excitation and emission are 390 nm and 465 nm, respectively.

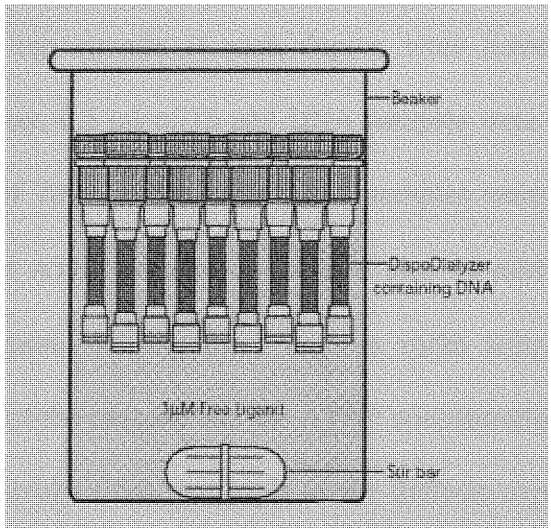


Figure S5. The competition dialysis experiment. Couple of DNA of distinct conformations and sequences were kept in individual dialysis units, respectively; and dialyzed against 200 ml of 1 μM AT-hook peptide in 1 × BPE plus 34 mM Na^+. The system was equilibrated at room temperature with stirring for 7 days. After equilibrium, the sample was taken out of the dialysis unit, and the amount of the DNA-bounded AT-hook peptide was determined.

A2.3 Results

A2.3.1 Comparison of the binding affinities of the ATHP to 32 AT sequences

From NMR study, it is known that each AT-hook motif of HMGA proteins bind to 5 bp of AT sequence in the DNA minor groove. In order to further reveal the sequence specificity of this binding, the binding affinities of the ATHP to 32 AT sequences were

evaluated using competition dialysis assay. In the experiment, the 32 AT sequences were designed in both top and bottom strands of 16 duplexes, which are generated by annealing two complementary oligonucleotides. The full sequence of each oligonucleotide has been described in Materials and Methods, and the sequences of the 5 bp AT region on the top strands of each duplex were listed in Table S3. In the competition dialysis assay, 0.5 ml of 15 μM DNA samples were injected into a dialysis cassette with MWCO of 7000, and total 16 DNA samples were subsequently dialyzed against 200 ml of 1 μM ATHP solution in 1× BPE plus 34 mM Na^+. The dialysis was performed under continuous stirring for 7 days at room temperature. Samples after dialysis were taken out of the cassette and the ATHP concentrations in the samples were determined. The final results from two trials of the dialysis experiments are shown in Figure S6. The blanked bars represent the data from the first trial, while the bars with stripes represent the result of the second trial. Obviously, the binding affinities of the ATHP to the 16 AT DNA determined in the two trials are not consistent. Besides, it was found that approximate 90% of the AT-hook peptide was lost during dialysis without known reason. A similar competition dialysis experiment was also performed using regular dialysis tubing of MWCO 8000. Instead of determining the ATHP concentration by its intrinsic fluorescence property, the sample in individual dialysis unit was mixed with 50 μl of fluorescamine solution (0.5 mg/ml in acetone), and the fluorescence intensity of the derivative produced from the reaction was measured at 390 nm excitation and 465 nm emission. The determined ATHP concentrations in 16 DNA samples were shown in Figure S7. Compared to the results obtained from the two trials described

above, the peptide concentrations determined here using the fluorescamine derivative are much higher. However, the binding specificity derived here does not match to any result obtained using the dialysis cassette. Due to the inconsistent results from different trials of the competition dialysis assays, the sequence specificity of the AT-hook peptide binding to DNA cannot be concluded.

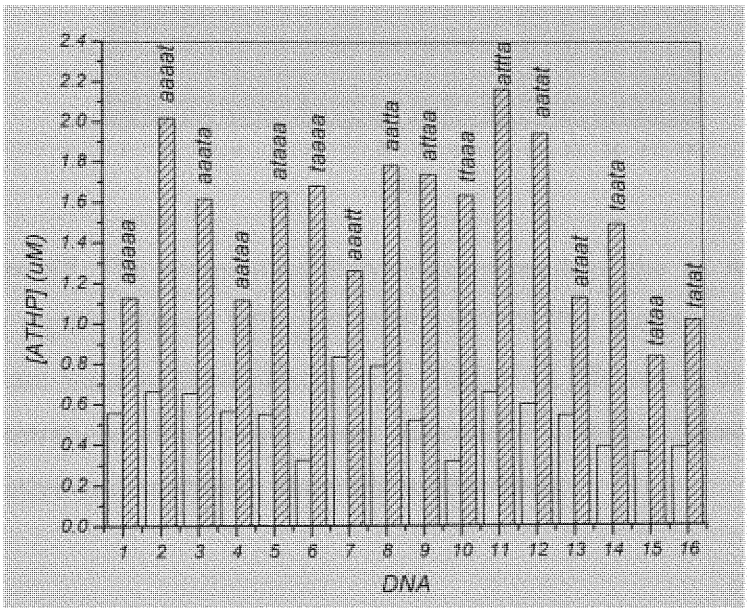


Figure S6. The competition dialysis assay result of the 16 AT DNA using dialysis cassette of MWCO 7000. The blanked and striped bars are data from the first and second experiment, respectively. For each experiment, 0.5 ml of 15 μM AT DNA was placed in an individual dialysis cassette of MWCO 7000, and total 16 samples were equilibrated with 200 ml of 1 μM ATHP solution in 1 × BPE plus 34 mM Na^+ for 7 days. After equilibrium, the ATHP concentration in each cassette was determined as described in Materials and Methods.

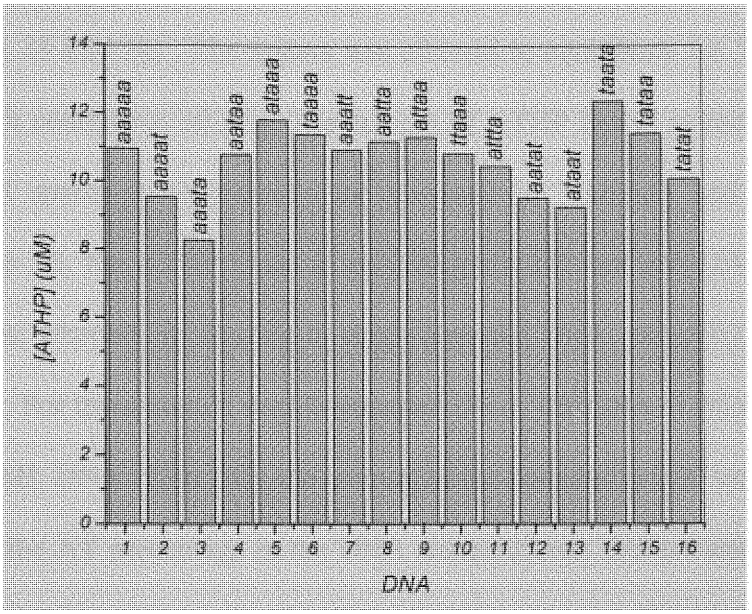


Figure S7. The competition dialysis assay result of the 16 AT DNA using regular dialysis tubing of 8000 MWCO. The dialysis conditions used are same as those described in Figure A2. However, the ATHP concentrations after dialysis were determined from the fluorescamine derivative as described in Materials and Methods.

A2.3.2 Binding affinities of the ATHP to 11 different nucleic acids

The binding affinities of the ATHP to 11 different nucleic acids were compared using the competition dialysis assay. Physical prosperities of these nucleic acids are listed in Table S4. These 11 nucleic acids were chosen based on their distinct conformations and nucleotide compositions. Qualities of the nucleic acids were checked by their UV spectra and the T_m obtained in melting studies. In the dialysis experiment, 0.5 ml of 75 μM (bp) each nucleic acid was injected into a dialysis cassette with MWCO of 7000. The 11 samples were then placed in a beaker containing 200 ml of 1 μM ATHP

in 1 × BPE plus 34 mM Na^+. The system was allowed to equilibrate for 7 days at room temperature with continuous stirring. The amount of ATHP bound to each nucleic acid was determined spectrophotometrically as described in Materials and Methods. As shown in Figure S8, the ATHP binds to poly(dA)poly(dT) and poly(dA-dT)$_2$ with the highest affinities; whereas, its binding to single-stranded DNA, poly(dA) and poly(dT), was pretty poor. For three natural double-stranded DNA, CP DNA (31% GC), CT DNA (42% GC) and ML DNA (72% GC), the amount of the bounded ATHP were found to be proportional to the AT percentage in the DNA compositions; this is consistent with the fact that the AT-hook peptide prefers bind to AT rich sequences. The binding specificity of the ATHP to different nucleic acids revealed here is very similar to that of some minor groove-binding drugs, such as distamycin A, netropsin and DAPI.

Table S4. The physical properties of the 11 nucleic acids used in the competition dialysis assay

Conformation	Nucleic acids	λ (nm)	ε (l M^{-1} cm^{-1})
Single-stranded DNA	poly(dA)	257	8600
	poly(dT)	264	8520
Natural double-stranded DNA	*C. perfringens* DNA (31% GC)	260	12476
	Calf thymus DNA (42% GC)	260	12824
	M. lysodeikticus DNA (72% GC)	260	13846
Synthesized double-stranded DNA	poly(dA-dT)$_2$	262	13200
	poly(dA)poly(dT)	260	12000
	poly(dG-dC)$_2$	254	16800
	poly(dI-dC)2	251	13800
Double-stranded RNA	poly(A)poly(U)	260	14280
DNA-RNA hybrid	poly(A)poly(dT)	260	12460

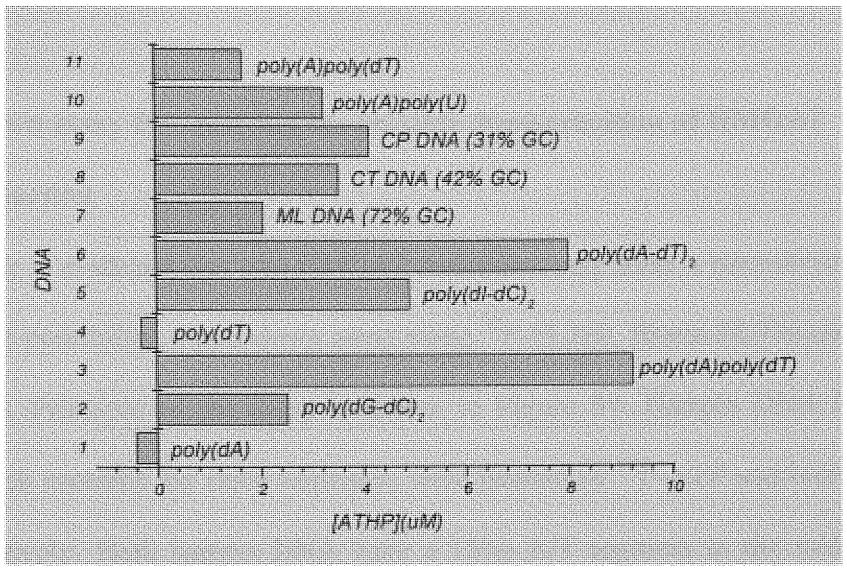


Figure S8. The competition dialysis assay result of the 11 nucleic acids. 0.5 ml of 75 μM each nucleic acid was placed in a dialysis cassette with MWCO 7000 and dialyzed against 200 ml of 1 μM ATHP solution in 1× BPE plus 34 mM Na^+ for 7 days at room temperature. The bounded ATHP in each dialysis cassette after equilibrium was measured and plotted here.

Appendix 3

A3. Pull-down assay for verifying the possible dimmerization of HMGA2

A3.1 Introduction

There is evidence that two HMGA2 molecules may form a homodimer. In order to verify this possibility, a pull-down assay using ^{32}P-FL260, Biotin-FL336 and HMGA2 was applied. Principle of the assay is shown in Figure S9. Hairpin DNA FL260 and FL336, which contains the same sequence, were labeled by ^{32}P-ATP and biotin at the 5' end, respectively. Both hairpin DNA have a 15bp A/T segment in the middle, which specifically binds to HMGA2. Our hypothesis is that if the HMGA2 can form a homodimer that binds to both biotin-FL 336 and ^{32}P-FL260 at the same time, the ^{32}P-FL260 will be detected after co-precipitation with biotin-FL336 during the pull-down assay.

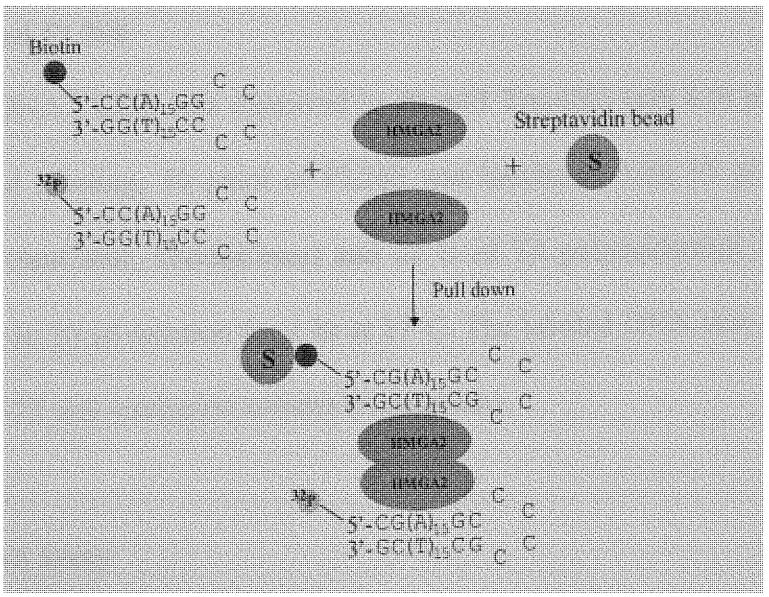


Figure S9. A carton representing the pull-down assay, in which the ^{32}P-FL260 will co-precipitate with biotin-FL336 if two HMGA2 proteins can form a homodimer that binds simultaneously to both ^{32}P-FL260 and biotin-FL336.

A3.2 Materials and Methods

A3.2.1 Materials

DNA: FL260: 5'-CCAAAAAAAAAAAAAAAGCCCCCGCTTTTTTTTTTTTTTTGG-3'; and FL360:Biotin-5'-CCAAAAAAAAAAAAAAAGCCCCCGCTTTTTTTTTTTTTTTGG-3', were purchased from MWG-Biotech Inc. and used without further purification. DNA concentrations were determined using calculated extinction coefficient at 260 nm.

A3.2.2 HMGA2 and ^{3}H-HMGA2 protein

HMGA2 and ^{3}H-HMGA2 protein was purified from *E. coli* BLR/DE3 cells as described previously. Stock solutions of the protein were prepared once and used for all pull-down assays.

A3.2.3 ^{32}P-labeled FL260

Hairpin DNA FL260 was dissolved in 20 mM Tris-HCl (pH 8.0) plus 100 mM NaCl. For 5'-end labeling, 10 pmol of FL260 was incubated with 15 μl of ^{32}P-ATP (3000Ci/mmol, 10mCi/ml, PerkinElmer) and 20 U of T4 polynucleotide kinase (New England Biolab) in 20 μl reaction for 1 hr at 37°C. The labeled DNA was subsequently purified using G-25 size-exclusion column (Amersham).

A3.2.4 Pull-down assay

The streptavidin agarose beads (Cat. 69203-3) were obtained from Novagen. For each sample in the pull-down assay, 50 μl of the streptavidin agarose beads in 50 % v/v

suspension containing 100 mM sodium phosphate pH 7.5 and 0.02 % sodium azide, was spun at 13,000 rpm for 5 minutes, followed by washing the pellet twice with 50 μl of the binding buffer. For the pull-down assay, 100 μl of the binding reaction was equilibrated at room temperature for 30 min and added to the prepared streptavidin agarose beads after that. The mixture was kept at room temperature for 30 min before spinning at 13, 000 rpm for 5 min. The pellet was washed twice with 50 μl of the binding buffer, and resuspended in 30 μl of denaturing buffer containing 10 mM EDTA (pH 8.0), 95 % (v/v) formamide, and 0.1 % SDS. Next, the sample was heated in a 65°C water bath for 10 min and spun immediately at 13, 000 rpm for 3 min. 20 μl of the supernant was mixed with 2.3 μl of the 10 × DNA loading buffer (50 % glycerol, 0.4 % bromphenol blue, and 0.4 % xylene cyanol FF) and 0.5 μl of 5 % SDS. The sample was then heated at 95°C for 10 min, cooled on ice, and loaded to a 15 % urea denaturing polyacrylamide gel. The electrophoresis was performed at 300 V in 0.5 × TBE for 4 hours. The gel was subsequently dried and exposed to X-film.

A3.3 Results

A3.3.1 Effect of HMGA2 concentration on the pull-down assay

As more frequent contact between two HMGA2 molecules are expected as a result of increased HMGA2 concentration, the influence of HMGA2 concentration on the self association of HMGA2 monomers were studied. In a pull-down assay, increasing amount of HMGA2 from 0 to 2 μM was incubated with 200 nM ^{32}P-FL260 and 200 nM biotin-FL336 in 100 μl binding buffer containing 20 mM Tris-HCl (pH 8.0),

20 mM NaCl, 0.5 mM EDTA, 1 mM DTT, and 0.5 mM $MgCl_2$. After 30 min equilibrium at room temperature, the binding reaction was mixed with 50 µl pre-equilibrated streptavidin agarose beads, and the pull-down assay was conducted as described in the Materials and Methods. As revealed in Figure S10, the amount of co-precipitated ^{32}P-FL260 increased in proportional to the HMGA2 concentration, which may indicate more stable complexes formed by HMGA2 homodimer, ^{32}P-FL260 and biotin-FL336.

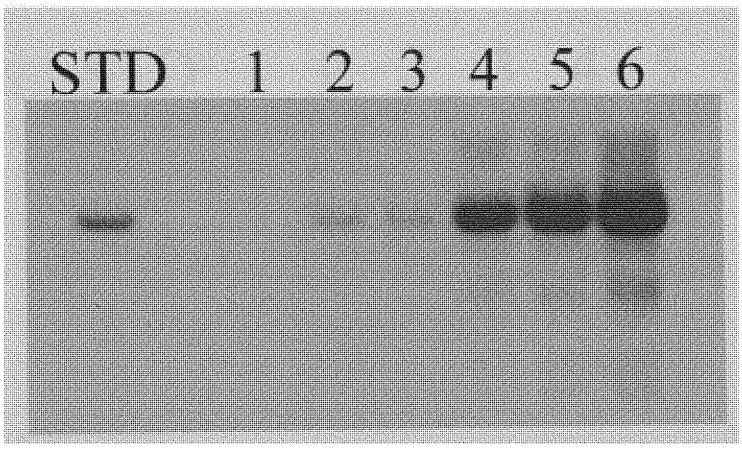


Figure S10. The pull-down assay using 200 nM ^{32}P-FL260 and 200 nM biotin-FL336 with 0, 200, 400, 800, 1200, and 2000 nM HMGA2 from lane 1 to 6 in 1 × binding buffer containing 20 mM Tris-HCl (pH 8.0), 20 mM NaCl, 0.5 mM EDTA, 1 mM DTT, and 0.5 mM $MgCl_2$. STD is the ^{32}P-FL260 used as the molecular weight standard.

A3.3.2 Salt-dependence of the pull-down assay

As salt concentration may affect the stability of DNA-HMGA2 complex and possible homodimer of HMGA2, a pull-down assay using increasing amount of NaCl was performed. The result was shown in Figure S11. 200 nM ^{32}P-FL-260 and 200 nM biotin-FL336 were incubated with 800 nM HMGA2 in binding buffer containing different salt concentrations. Unexpected, the amount of the co-precipitated ^{32}P-FL260 did not change as the salt concentration was increased from 20 mM to 200 mM.

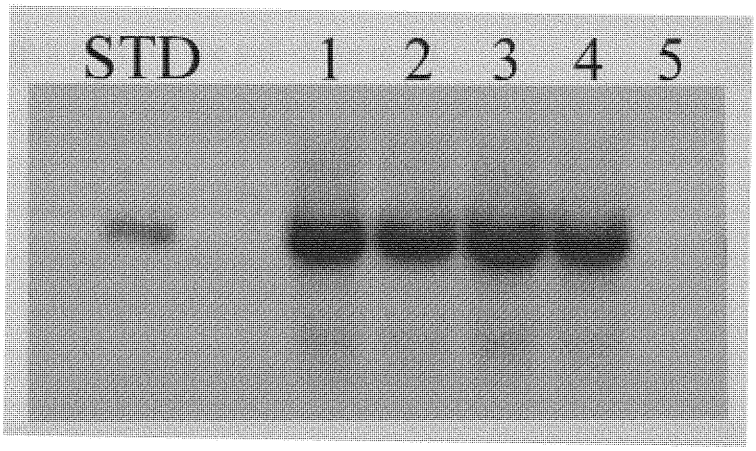


Figure S11. The pull-down assay using 200 nM ^{32}P-FL260, 200 nM biotin-FL336 and 800 nM HMGA2 in 1 × binding buffer of 20 mM Tris-HCl (pH 8.0), 0.5 mM EDTA, 1 mM DTT, 0.5 mM $MgCl_2$ with 20, 50, 100, 200, and 400 mM NaCl, respectively, from lane 1 to 5. STD is the ^{32}P-FL260 used as the molecular weight standard.

A3.3.3 Effect of poly(dG-dC)$_2$ on the pull-down assay

Effect of poly(dG-dC)$_2$ on the pull-down assay was evaluated by introducing increasing amount of poly(dG-dC)$_2$ as non-competing DNA in 1 × binding buffer

containing 20 mM Tris-HCl (pH 8.0), 20 mM NaCl, 0.5 mM EDTA, 1 mM DTT, and 0.5 mM $MgCl_2$. In the binding reactions, the concentration of ^{32}P-FL260, biotin-FL336, and HMGA2 were fixed at 200, 200 and 800 nM, respectively. Stock poly(dG-dC)$_2$ was added to the binding reaction so that the poly(dG-dC)$_2$ final concentration (bp) is 1-5 × of the total A/T base pair concentration including that from ^{32}P-FL260 and biotin-FL336. As shown in the lane 1 to 3 in Figure S12, addition of 1 × to 2 × of poly(dG-dC)$_2$ in the binding reaction obviously reduced the amount of the pull-down ^{32}P-FL260; However, further increase of the poly(dG-dC)$_2$ concentration from 3 × to 5 × as shown in lane 4 to 6 has no significant effect on the pull-down result.

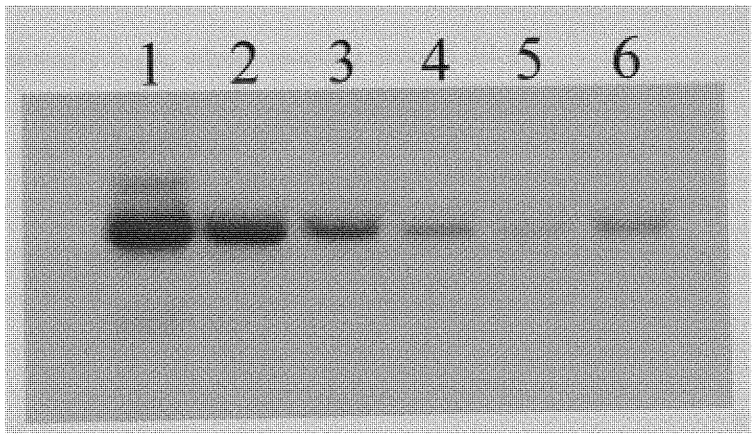


Figure S12. The pull-down assay using 200 nM ^{32}P-FL260, 200 nM biotin-FL336 and 800 nM HMGA2 in 1 × binding buffer containing 20 mM Tris-HCl (pH 8.0), 20 mM NaCl, 0.5 mM EDTA, 1 mM DTT, and 0.5 mM $MgCl_2$. The poly(dG-dC)$_2$ concentrations are 0, 6, 12, 18, 24 and 30 μM (bp), respectively, from lane 1 to 6.

A3.3.4 The molar ratio of HMGA2 to FL260 in the pull-down complex

To find the molar ratio of HMGA2 to FL260 in the pull-down complex, 0.8 μM ^{3}H-HMGA2 (specificity=375 cpm/pmol) was incubated with 0.2 μM ^{32}P-FL260 (specificity=4000 cpm/pmol) and 0.2 μM biotin-FL336 in 100 μl binding reaction containing 20 mM Tris-HCl (pH 8.0), 20 mM NaCl, 0.5 mM EDTA, 1 mM DTT, and 0.5 mM $MgCl_2$. The pull-down assay was performed as described above and the final pellet that contains the protein-DNA complex was washed once with 50 μl binding buffer and resuspended in 100 μl binding buffer. The suspension was directly transferred to 15 ml Aquesol 2 liquid scintillation counter cocktail, and the ^{32}P and ^{3}H activities were counted on a Beckman LS5000TD LSC. Surprisingly, the calculated molar ratio of HMGA2 to FL260 was abnormally high, which reached around 30:1.

VITA

TENGJIAO CUI

October 28, 1972	Born, Sichuan, China
1990-1994	Bachelor, Biochemical Engineering Tianjin Science & Technology University, China
1994-1997	Master, Biochemical Engineering Tianjin Science & Technology University, China

PUBLICATIONS AND PRESENTATIONS

Energetics of binding the Mammalian High Mobility Group Protein HMGA2 to poly(dA-dT)$_2$ and poly(dA)poly(dT). Tengjiao Cui, Shuo Wei, Keith Brew and Fenfei Leng, *J. Mol. Biol.* (2005) **352**:629-645

Large Scale Preparation of the Mammalian High Mobility Group Protein A2 for Biophysical Studies. Tengjiao Cui, Suzanne Joynt, Victor Morillo, Maria Baez, Zhichun Hua, Xiaotang Wang and Fenfei Leng, *Protein and Peptide Letters* (2007) **14**: 87-91

Biophysical Society 48th Annual Meeting, February 14-18, 2004, Baltimore, MD
Specific Binding of High Mobility Group Protein HMGA2 to Minor Groove of DNA: Calorimetric and UV Melting Studies.
Tengjiao Cui, Shuo Wei, Keith Brew, and Fenfei Leng

Biophysical Society 51th Annual Meeting, March 3-7, 2007, Baltimore, MD
The Mammalian high mobility group protein AT-hook 2 (HMGA2) is a sequence-specific DNA-binding protein.
Tengjiao Cui and Fenfei Leng